Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1	9	0	0	0	0	Well-written with (mostly) consistent style. Main issues are generally well-explained with supporting examples. [Graham Feingold, United States of America]	Noted. Thank you!
9-2	9	0	0	0	0	ESM are the current state of the art models, but the chapter often refers to AOGM and ESM. Actually, the difference is not clearly specified and it is somewhat confusing for the reader. In the future version it must be clearly stated which models are considered as AOGCM or ESM (e.g. by improving Table 9.1) [Eric Martin, France]	Accepted. Thank you!
9-3	9	0	0			Many aspects of this chapter are well written wth numerous relevant references. However, there are some aspects that are not that good and my first comments are about overall structure. [Gunilla Svensson, Sweden]	Noted. Thank you!
9-4	9	0	0			Reanalysis fields are often used in climate model evaluation. Depending on what parameter that is used, they can either be very close to the observations or they might be a mere model product with the same faults as the climate models likely have. I think a discussion on this, referring back to the observational chapter should be included. [Gunilla Svensson, Sweden]	Accepted. Will present more consistent links to observational chapters
9-5	9	0	0			I have only had time to go through the evaluation chapter in depth but I have had a briefer look at the others and I note that there is a section called Model evaluation in Chapter 6 on "Evaluation of Carbon Cycle Models". That, in combination of the extensive text on this subject in Chapter 9 is simply too much. [Gunilla Svensson, Sweden]	Accepted.There is a clear separation of labour between Chapters 6 and 9, in that Chapter 9 only evaluates the carbon cycle in coupled models. Will ensure this is emphasised clearly enough.
9-6	9	0	0			In the begining of the chapter, the models to be evaluated are listed as AOGCMs, ESMs, EMICs and RCMs. In chapter 9.4 the majority of evaluation concerns AOGCMs, I think but it is not clear because this text is not written based on the different chategories. Also, an additional model type - CCM has gotten its own subsection - that is not well-balanced. EMICs does not seem to be dicussed much at al. [Gunilla Svensson, Sweden]	Accepted. Will ensure better balance and appropriate labelling.
9-7	9	0	0			My strongest concern with the chapter is that so little of the evaluation text consider the atmospheric circulation, the backbone of the climate model. To have a special subsection on ozone and not discussing the general circulation (Hadley circulation, ITZC, storm tracks, storm intensities, jet streams, mean sea level pressure etc) does not give a balanced and comprehensive summary of the evaluations that are necessary to increase the credability for the use of the models for future climate predictions. [Gunilla Svensson, Sweden]	Accepted. Assessment of atmospheric circulation will be expanded considerably.
9-8	9	0	0			I realize that much of the science assessed in this report must be based on papers published on CMIP3 analysis. However, I really hope that a lot of effort will be made to update figures with more recent CMIP5 results. There is not much expectation that the CMIP5 in many cases will show much improvement but it adds value to show that CMIP3 and CMIP5 show about the same. [Gunilla Svensson, Sweden]	Noted. For SOD, many more CMIP5 results will be available. We will pay particular attention to the similarities and difference between CMIP3 and CMIP5.
9-9	9	0	0			Detailed comments: [Gunilla Svensson, Sweden]	Noted.
9-10	9	0				Throughout the entire report there is general agreement that natural variability could dominate the anthropogenically forced response in the near-term and that in mid-latitudes the NAO/AO/NAM related variability is a key for understanding and quantifying this uncertainty. In my opinion an explicit figure on the expected changes (near and long term) in NAO/AO/NAM variability derived from the CMIP5 models used would be extremely helpful for the reader. The same could be argued for ENSO variability. Both figures could be placed either in chapter 9, 11 or 14. [Christof Appenzeller, Switzerland]	Noted, but this is for Chapters 11, 12, or 14.
9-11	9	0				Some evaluation of tropical atmospheric circulations (Hadley, Walker) is needed as much time is devoted to these in the projections chapters [Julie Arblaster, Australia]	Accepted. Assessment of atmospheric circulation will be expanded considerably.
9-12	9	0				Generally, this is a very good and helpful chapter. However, I think some key statements concerning the limits and interpretation of models should be made more prominent and should be taken into account more consistently throughout this chapter. I explain below. [Gregor Betz, Germany]	Accepted.
9-13	9	0				This chapter claims to measure the reliability of the models (as said in the section 9.8). However, there is little of quantitativity in this chapter. Although it provides a ver useful discussion of different strengths and weaknesses of several models, there is no clear "quantitativeness" in the analysis, only qualitative (as for comparing models and commenting their flaws). A much more clear definition of "quatitativeness" must be used thorugh the chapter, in a similar way to the "probability" or "likelihood" used in the other chapters. [Juan	Rejected. Through the explicit display of biases and use of metrics we are being quantitative throughout the chapter. Moreover, we use the calibrated uncertainty language as stipulated. As stated in the uncertainty guide, the use of probabilistic language

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						A. Blanco, Canada]	sometimes is simply not appropriate.
9-14	9	0				The organization of this chapter is not satisfactory. The sections 9.7 and 9.8 should be the first ones, to clearly discuss at the beginning the important of model evaluation and the main results of them. Later, each of the sections can desribe these issues in more detail. [Juan A. Blanco, Canada]	Rejected. The organisation of the Chapter is clearly spelled out in the beginning: 9.7 and 9.8 as handovers should be in the end. However, the Executive Summary needs to be and will be expanded.
9-15	9	0				No discussion of tipping points and little discussion of thresholds in chapter 9 - I think that brief discussion would usefully contextualise climate models following on from the discussion of these in chapter 1, 5 and 12. I could suggest a sentence or two if you wish. [Mark Charlesworth, United Kingdom of Great Britain & Northern Ireland]	Rejected. There is no known observation of a tipping point or threshold having been crossed during the instrumental period, and hence there is no possibility to evaluate a model on its ability to cross such a threshold.
9-16	9	0				PMIP2 is used repeatedly but never defined. OBGC is defined more than once and the definitions are not consistent (10/30, 39/1). [James Christian, Canada]	Accepted. Will check for consistency in defining acronyms.
9-17	9	0				Some of the summary paragraphs are shopping lists of skill assessments in IPCC-speak. This is very hard to follow and could perhaps be broken up with a bit more physical detail. I assume this is IPCC-approved language, but I find it hard to envision what "medium evidence of mixed skill" means, and to the uninitiated it sounds a lot like "skill not significantly different from zero". [James Christian, Canada]	Taken into account. Use of calibrated uncertainty language is stipulated by IPCC procedures, but we will strive to streamline and clarify.
9-18	9	0				I avoid the phrase "diurnal cycle” (which appears about 30 times, see e.g. section 9.5). Technically diurnal is the opposite of nocturnal, so diurnal is only one phase of the diel cycle. So it’s OK to talk about e.g. diurnal temperatures in a climate model, but “diurnal cycle” is sort of like referring to the annual cycle as the “summer cycle”. I am aware that this usage is well established and my chances of changing it at this point are not very good. [James Christian, Canada]	Rejected. This is not the place to change the common usage of a term.
9-19	9	0				I am not sure whether there is a recent literature study on this, but it would be nice to have some information about in how far the latest generation of models is able to reproduce the climate changes associated with the 11-year solar cycle, at the very least in terms of global temperature, where several recent studies appear to confirm a temperature difference of about 0.1 K between maxima and minima. [Georg Feulner, Potsdam]	Accepted. Will check for literature.
9-20	9	0				In order to compare climate models to data, recent statistical thinking looks at observations as draws from a distribution (namely the climate). Since climate is not directly observed, it needs to be estimated using data that are observed with error. Thinking of the model output as well as climate described with error requires setting up a hierarchical model. There has been several examples of this in the literature (e.g. Milliff et al, QJRMS 2011, Li et al, J.Am.Stat.Ass. 2010). Using a more direct approach, two-sample methods allows the comparison of two distributions (that from the model output and that from the data), e.g. Orskaug et al., Tellus A, 2011. The development of singledimensional index numbers is an oversimplification of a multidimensional situation, and tends to lose information (see e.g. http://www2.meteo.uni-bonn.de/mitarbeiter/olga/Orals/Wednesday_1/David_Stephenson.pdf). [Peter Guttorp, USA]	Noted. Such a discussion is already included in the chapter.
9-21	9	0				Some sections give short (one or two sentence) introductions on the considered processes and their importance, but some not. I would appreciate that a short introduction is given for all sections. [Farahnaz Khosrawi, Sweden]	Accepted. Will strive for more consistency in style across subsections by relating explicitly to AR4 assessment.
9-22	9	0				There are too many abbreviations used. Please check if they are all really necessary or if some could be omitted. [Farahnaz Khosrawi, Sweden]	Accepted. Will check, but in a technical document like this the use of abbreviations is necessary.
9-23	9	0				Refering to my previous comment, it could be worth to add a list of abbreviations at the end of the chapter. [Farahnaz Khosrawi, Sweden]	Will consider.
9-24	9	0				There are many quantitative measures for quality assessment of models and their results. Nevertheless, there is also always a cerrtain amount of expert judgement involved. Chapter 1 correctly states that experts tend to be overconfident. This is particularly relevant for this chapter, which should be more explicit about the subjective elements of confidence and fidelity assessment. The usefulness of this chapter would be strengthened if uncertainty assessments were based on broader, well documented expert panels, including sceptical scientists from the wider community of natural scientists. [Gerbrand KOMEN, Netherlands]	Noted. However, the Chapter team is what it is; we bring in CAs as needed; and our task is the assess the knowledge that exists, not to create new knowledge by experimenting with new expert panels.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-25	9	0				I am missing a thorough discussion of limitations and problems of validation data sets at a central place of the chapter. Often reanalysis data sets are used or gridded data sets, and all of them have their limitations. E.g., Hofstra, New and My Sweeney (2010), Clim. Dynam., 35: 841-858; Kysely and Plavcova (2010) J. Geophys. Res., D23118; and Maraun, Osborn and Rust (2012), Clim Dynam., DOI 10.1007/s00382-011-1176-0 all find serious problems of the E-OBS data set, which is routinely used for validation of regional climate models over Europe. [Douglas Maraun, Germany]	Taken into account. But assessing the quality of datasets is the purview of the observational chapters; SOD will make more explicit links.
9-26	9	0				I would suggest not to use the term validation metric in case indices such as mean, standard variation, percentiles etc are referred to. A metric is a distance measure with clearly defined properties, e.g., it is always equal to or larger than zero. This is definitely not the case for, e.g., daily minimum temperatures. Please use the term indices instead. I have found at least one example (see comment above), please check the text accordingly. [Douglas Maraun, Germany]	Rejected. "Metric" has become a standard term i the field, although it does not strictly correspond to the mathematical use of the term..
9-27	9	0				There is almost no reference made to atmosphere-only runs or the AMIP component of CMIP5, which can be crucial to evaluation of the models' atmosphere processes. Once of the main advances in CMIP5 compared with CMIP3 is the range of model configurations available for use in understanding the models' climate projections. This does not come across in this chapter. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. To the extent that new analyses of CMIP5 AMIP-type runs are available in the literature, we will assess.
9-28	9	0				There is little reference made to the high-top models, other than a brief mention in section 9.1.3.3.7, even though the point is made there that comparison of high-top and low-top models can be used to highlight the role of the stratosphere. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. To the extent that new analyses of high/low top models are available in the literature, we will assess.
9-29	9	0				My commendations to the authors on the enormous amount of work done to compile this chapter. I have two overall comments. First, based on the fields that I have seen from the CMIP5 analysis, there may be some places where the discussion regarding model improvement is slightly optimistic. No doubt this will be addressed in the revisions as more results are available to the authors. Second, it is natural for the authors to write and reference what they know best (which often includes their own work). I have tried to supply some additional references where I could think of them, but likely there are many more that have been omitted. [J. David Neelin, United States]	Noted. Thank you!
9-30	9	0				This chapter evaluates the performance of AOGCM models to be used for the future projections for the global as well as regional climate in the upcoming IPCC AR5 report. In general most of the model could capture the mean state and varaibility of different modes of earths ocean,atmosphere and land in different time scale with resonable accuracy. However most of the sections the results from CMIP3 have been discussed. It would be nice to give some results from CMIP5 particularly in MJO and Indian Ocean Dipole mode simulations since there have been considerable improvements in the model physics and resolutions as compared to CMIP3 . The performance of the model in terms of sea level variability and its long term change have not been shown. It is one of the most important parameter in the global climate change projections. IPCC is giving climate change projection since its first report published in 1990. Its now two decades over since its first report published. During these last two decade we have high quality observations from different platforms over global ocean, land,ice and atmosphere. Can it be possible to include one figure showing actually how the climate change projection by IPCC performed ? May be for instance the time series of global sea level rise projected by past IPCC and actual observation. I know it may be out of the scope at this stage but this will give better acceptability that how IPCC models are improved in recent time. [HASIBUR RAHAMAN, India]	Noted. As per agreement between Chapters 9 and 13, sea-level change will be assessed in Chapter 13, including model evaluation. Chapter 9 deals with heat-content changes. Comparison of projections assessed in the successive IPCC reports is done by Chapter 1.
9-31	9	0				The executive summary of this chapter does not provide likelihood and confidence assessments using the calibrated IPCC language. This makes it difficult for the reader to assess the reliability of the provided assessments. [Sonia Seneviratne, Switzerland]	Accepted. A clear weakness of the FOD, which will be improved upon in the SOD.
9-32	9	0				In summary, I like the way that process-based metrics are valued in this chapter. However, I would like to see more discussion of issues around a) the use of single metrics v multivariate basket of metrics and b) the importance of accounting for structural uncertainty in our metrics. For a) the chapter needs to bring out the dangers of focussing on one killer metric – supposing a model does well at one metric used to constrain sea-ice extent but poorly at a metric designed to constrain climate sensitivity – the climate system is highly interdependent in very complex ways and we cannot ignore this and so we need to consider a variety of	Taken into account. Text in section 9.8 revised somewhat.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						metrics. If we don't we are in danger of providing over-confident but conflicting assessments of how different climate variables will be affected in future. That a model might be good at one metric and not at another arises because of structural errors and so it is important to account for this in a multivariate metric. Then, if we do this well, we will be better protected from conflicting projections from later CMIP exercises where structural errors will be different. [David Sexton, UK]	
9-33	9	0				Simple climate models (SCMs) have been discussed and used in SAR, TAR and AR4, but this chapter (evaluation of climate models) ignores this class of models. SCMs are still frequently used in both climate science and policy research (e.g. Roe and Baker, 2007, Nature; Meinshausen et al., 2009, Nature, 10.1038/nature08017; Rogelj et al., 2010, Nature, 10.1038/4641126a; Tanaka et al., 2009, Climatic Change, doi:10.1007/s10584-009-9566-6; Johansson, 2011, Climatic Change, doi:10.1007/s10584-010-9969-4; Armour and Roe, 2011, Geophysical Research Letters, doi:10.1029/2010GL045850). Clarifications are required as to why SCMs are not discussed in this chapter. It must be acknowledged (probably at the beginning of this chapter) that SCMs are actively used in climate research but are not evaluated in this chapter. [Katsumasa Tanaka, Switzerland]	Noted. While we agree they are still used in the climate research community, this does not require that the chapter evaluate them. Their purpose is well known and recognized in the community.
9-34	9	0				We miss a clear connection to Annex 1: Atlas of Global and Regional Climate Projections. It seems important to highlight to what extent Chapter 9 can inform regarding the foundation of the WGI Atlas. [Thomas Stocker/ WGI TSU, Switzerland]	Accepted. We have included several pointers to the Atlas and have been in regular contact to ensure consistency of regions, reference data, etc. Figure 9.40 was constructed specifically at the request of the Atlas team to support their assessment.
9-35	9	0				These portrait diagrams (Figs 9.9, 9.38) require further explanation to be given in the captions - it's necessary to guide the reader through the interpretation of these figures step-by-step. It is also not clear to us why the mean and median-model are all biased negative compared to the data (first two columns). [Thomas Stocker/ WGI TSU, Switzerland]	Accepted - caption and text have been expanded to clarify
9-36	9	0				The Executive Summary, and the chapter in general, uses many qualitative expressions, e.g., 'reasonably well', 'mixed skill'. Please provide quantitative statements wherever possible. Currently many of the quantitative results from the tables and figures are not highlighted in the chapter text and summary statements. [Thomas Stocker/ WGI TSU, Switzerland]	Accepted. A clear weakness of the FOD, which will be improved upon in the SOD.
9-37	9	0				Avoid the term 'IPCC models' [Thomas Stocker/ WGI TSU, Switzerland]	Accepted.
9-38	9	0				Is it possible to provide a comparison of CMIP3 vs CMIP5 model biases for key parameters, e.g. in the form of Taylor diagrams? [Thomas Stocker/ WGI TSU, Switzerland]	Accepted. Comparison of CMIP3 vs. CMIP5 is provided wherever possible.
9-39	9	0				Please describe how multimodel results are combined, put on a common grid, and presented in, e.g., maps (incl. grid information etc). Please check and ensure consistency of approach across chapters, especially for Chapters 9, 11, 12, 14 and, of course, Annex I: Atlas [Thomas Stocker/ WGI TSU, Switzerland]	Accepted.
9-40	9	0				We recommend the consistent use of clear introductory and concluding paragraphs for the various sub-sections of the chapter, so that it is clear what the new findings coming from your assessment are relative to the previous assessments. The relevance of your assessment findings for the climate change projections assessed in Chapters 11 and 12 should be provided in a concise form. We think that this will also improve the focus of your sections. [Thomas Stocker/ WGI TSU, Switzerland]	Accepted.
9-41	9	0				I'm sorry I have run out of time to review Chapter 9 as thoroughly as I'd hoped, and I only have a couple of generic comments. First, I very much welcome the presence of CMIP3 data in many of the figures and discussion. I realise this may partly be just a consequence of limited availability of CMIP5 data, but I would urge the authors to make full use of CMIP3 information in later drafts. This is a big opportunity for AR5 as it's the first time we have the chance for quite thorough comparison with earlier generation (but nevertheless quite mature) models. Consistency across model generations (or not) seems to me to be a useful tool in understanding the robustness of results. [Richard Wood, UK]	Accepted. Comparison of CMIP3 vs. CMIP5 is done wherever possible.
9-42	9	0				I'd also encourage a dialogue with Chapter 12 to develop some explicit statement on the basis for inference in model projections. I think it is very important that AR5 makes an explicit statement about how it is using the	Accepted. Section 9.8.3 was designed for exactly this purpose in mind. Will see what we can do concerning

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						various sources of information to make inferences about future climate. E.g. what is the status of the CMIP5 models vs CMIP3? PPEs vs MMEs? If each conclusion is based on an ad hoc expert judgement melding the various sources of information, that needs to be stated. Some conclusions at present seem to be simply arrived at by reading off numbers from the CMIP ensemble, whereas some bring in a wider range of information. There may be good reason for this, but in any case it would help the reader if the basis for inference could somehow be made clear in each section - for example through some kind of shorthand or footnotes. A common approach with Ch 12 would really strengthen the report. [Richard Wood, UK]	procedure.
9-43	9	0				References: Arakawa, A., J.-H. Jung, and C.-M. Wu, 2011: Toward unification of the multiscale modeling of the atmosphere, Atmos. Chem. Phys., 11, doi:10.5194/acp-11-3731-2011, 3731–3742. Collier, J. C., and G. J. Zhang, 2009: Aerosol direct forcing of the summer Indian monsoon as simulated by the NCAR CAM3. Clim. Dyn., 32,,313-332, DOI 10.1007/s00382-008-0464-9. Lau K. M., and K. M. Kim, 2006: Observational relationships between aerosol and Asian monsoon rainfall, and circulation. Geophys Res Lett 33(L21810). doi:10.1029/2006GL027546 Liu, P., B. Wang, K. R. Sperber, T. Li, and G. A. Meehl, 2005: MJO in the NCAR CAM2 with the Tiedtke Convective Scheme. J. Climate, 18, 3007-3020. Mapes, B.E., and R. B. Neale, 2011: Parameterizing Convective Organization to Escape the Entrainment Dilemma, J. Adv. Model. Earth Syst., 3, M06004, doi:10.1029/2011MS000042. Ramanathan, V., C. Chung, D. Kim, T. Bettge, L. Buja, J. T. Kiehl, W. M. Washington, Q. Fu, D. R. Sikka, and M. Wild, 2005: Atmospheric Brown Clouds: Impacts on South Asian Climate and Hydrological Cycle. PNAS, Vol. 102, No. 15, 5326-5333. Song, X. and G. J. Zhang, 2009: Convection parameterization, tropical Pacific double ITCZ, and upper ocean biases in the NCAR CCSM3. Part I: Climatology and atmospheric feedback. J. Climate, 22, 4299-4315. Waliser, D. E., W. K. Lau, J. H. Kim, 1999: The Influence of Coupled Sea Surface Temperatures on the Madden Julian Oscillation: A Model Perturbation Experiment. J. Atmos. Sci., 56, 333-358. Wu, X., L. Deng, X. Song, G. Vettoretti, W. R. Peltier, and G. J. Zhang, 2007: Impact of a modified convective scheme on the Madden-Julian Oscillation and El Nino-Southern Oscillation in a coupled climate model. Geophys. Res. Lett., 34, doi:10.1029/2007GL030637. Zhang, G. J., and X. Song, 2010: Convection parameterization, tropical Pacific double ITCZ, and upper ocean biases in the NCAR CCSM3. Part II: Coupled Feedback and the Role of Ocean Heat Transport. J. Climate, 23, 800-812. [Guang Zhang, United States of America]	Noted, but unable to discern the action required here.
9-44	9	1	1	1	1	Unlike in the past, some modeling groups tuned their model to approximately fit the historical estimates of radiative forcing changes - pre-industrial to present day. Where is this discussed in AR5 WG1 Report? How does this explicit tuning of the model over the historical period impact the assessment of the models' simulation of the historical period? One can no longer say that the successful simulation of the historical period is evidence of the models being reliable estimators of the future changes. This discussion could be put near the climate sensitivity section in this chapter. [Ronald Stouffer, USA]	Accepted. Will include discussion when we discuss tuning.
9-45	9	1	1	1		Evaluation of Climate Models [Medani Bhandari, Nepal]	Noted.
9-46	9	1	1			This is a very good first order draft! Nevertheless, some sections and text blocks are incoherent and sometimes do not match well with others and therefore appear as "foreign substances". To my understanding the chapter should focus on the evaluation of climate models (i.e. AOGCMs) which are the basement for AR5. Details about the evaluation/strategies of evaluation of other models (e.g. Earth-System Models, Chemistry-Climate-Models) should be discussed only, if results derived from such models are explicitely discussed in	Noted. Thank you!

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						other chapters of AR5 too and provide fundamental contributions to the conclusions of AR5. A detailed discussion of model systems in this chapter, which are not really used in other chapters should not be covered in Chapter 9. Concrete examples are given below. [Martin Dameris, Germany]	
9-47	9	1		1		Some rationale should be provided for the selection of baseline averaging periods for model evaluation. The brief mention on page 5, line 21 is insufficient. Given the general preference among climate scientists for at least 30-year means, and the availability of observations for >30 years, why use 21-year means in Figures 9.3-9.5? What is the period of record used for the comparison in Fig. 9.6 and why was it selected? Why use a 20-year mean in 9.8? a 14-year mean in Fig. 9.16? etc. I know it is impossible to use a single consistent time period for all comparisons, but where 30-year periods are possible they should be used. [Philip Mote, USA]	Accepted. We will make clearer reference to the observational chapters in SOD; many choices of reference period lengths are dictated by data availability.
9-48	9	1		50		I reviewed Chapter 9 and this chapter looks fine as well. I have the above two suggestions for the section 9.5.2.2.1 [Anthony Lupo, USA]	Noted, thank you!
9-49	9	1		78		Overall this is a very good start at providing a comprehensive evaluation. Thanks to all of you for your effort in putting it together. [Duane Waliser, USA]	Noted, thank you!
9-50	9	1		171		A considerable fraction of this chapter depends on updates based on CMIP5 experimental results that were unavailable or far from complete at the time of the FOD. Updates that include CMIP5 data are essential for most of this chapter. [James Kinter, United States of America]	Noted.
9-51	9	1				This is a nice chapter. Section 9.1 does a good job at laying out the progress made in Earth System modelling. However this does not quite translate into the following sections where the evaluation of ESM components are somewhat limited to the C and S cycles (even though AOD are shown in the S cycle sub-section). [Olivier Boucher, France]	Accepted. As more CMIP5 ESM results became available, these were incorporated
9-52	9	1				Note that the EC-Earth model results for CMIP5 have become available and published in peer-review literature. If not on the ESG server the data can be found on climexp.knmi.nl, together with the other CMIP5 runs. The references are 14. Hazeleger, W. et al, 2010. "EC-Earth: seamless earth system prediction in action." Bull. American Met. Soc., 91, 1351-1356. and 1. Hazeleger W., et al., 2011: "EC-Earth V2.2: description and validation of a new seamless Earth system prediction model." Clim Dyn. in press [Sybren Drijfhout, Netherlands]	Noted.
9-53	9	1				Overall I found this chapter well-written and clear, with lots of interesting material. [Nathan Gillett, Canada]	Noted, thank you!
9-54	9	1				In terms of the overall AR5, I think the primary purpose of this chapter is to inform the use of models for projections, and to a lesser extent for attribution. This is dealt with explicitly in section 9.8. But I think the chapter as a whole might be re-focused somewhat with this aim more in mind. Summary statements might focus more clearly on which aspects of model performance could be trusted in the context of projections and which not. The chapter reviews the performance of a range of different models and model types, and the models considered vary depending on the variable discussed. Since the majority of projections in the report will be based on the CMIP5 ensemble, it might be helpful to more clearly distinguish between the performance of the CMIP5 models at simulating each aspect and the performance of other models or model classes. This could also be carried through into some of the summary statements. Secondly, models are never referred to by name in the text of the chapter (although they are named in tables and figures), even when one specific model is described (phrases like 'a particular ESM' are used). This applies both to sections describing model formulation and to those describing model results. But information on processes included in each model, and which models do better and which do worse would be helpful to the reader in interpreting their projections of future change. While this information could in some cases be gleaned from figures, or by examing the references cited, it would be more straightforward and helpful to name models where they are discussed in the text. [Nathan Gillett, Canada]	Taken into account. However, while it is clear that Chapters 11 and 12 will be mostly based on CMIP5 results, the depth of analysis of CMIP3 is so vastly greater, and quite possibly the difference between CMIP3 and CMIP5 perfomance sometimes quite small, that we absolutely must include comprehensive assessments of CMIP3 results. Will consider explicitly mentioning individual models in text.
9-55	9	1				In several places, summaries in the chapter list evidence and agreement separately for each variable, even in cases of 'mixed results' (e.g. 9.5.3.7). This makes the text hard to read, even though I know this is following the good practice guidance note. Perhaps the evidence and agreement information could be included in a table, and confidence could be reported in the text. This would allow statements to be grouped according to whether they were 'high confidence', 'medium confidence' or 'low confidence', which would make the text	Accepted, will implement this strategy.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						easier to read and interpret. [Nathan Gillett, Canada]	
9-56	9	1				For chaptere 10, it would be really useful to be able to refer to a power spectrum that compares observations with models - I like the NH comparison for the last millennium, but can't see any 'real' data in it - I think that would be really helpful to be able to relate to from ch10 - proxy based reconstructions may be very uncertain to use but might be worth adding some, and the instrumental data. Also, it would be great if a comparison of precipitation interannual or trend variability between models and data could be shown for latitude bands. My impression from a very rushed check is that chapters 9 and 10 could mesh even better, if chapter 9 showed a few more things relevant for chapter 10 this would help chapter 10 a lot to bring the detection and attribution results, that are, after all, usually based on variability from models, into context. [Gabi Hegerl, UK]	Accepted, see Figure 9.33.
9-57	9	1				it would be good if a general phrase is defined which is used to summarise AOGCM and ESM like CM (climate models) [Frank Kreienkamp, Germany]	Rejected. We use the terms that are currently in use in the literature.
9-58	9	1				quite often the authors are writing 'Most models ...' or 'A growing number of models..' please give a detailed information eg. In table; please do not use useless phrases [Frank Kreienkamp, Germany]	Accepted, will be more specific.
9-59	9	1				I commend the authors heartily for the herculean effort required to analyze and present CMIP5 model results. For figures based solely on CMIP3 models, the figure should either represent a substantial advancement in knowledge of the climate system, or it should be updated with CMIP5 models. Presenting evaluation of CMIP3 models has less value with the passage of time. I have not done a comparison with AR4, but some figures (e.g. Fig 9.9) are familiar and do not carry the note "placeholder for second order draft" so it appears there is no intention to update them. If so they should be deleted. [Philip Mote, USA]	Accepted. Where we used a figure from AR4, this was purely as a placeholder (although not always stated). SOD will not contain any figure from AR4.
9-60	9	1				The figures in Supplement to chapter 8 of AR4 were very valuable diagnostics of model performance, especially difference plots model-obs. I would hope that similar plots will be given for AR5. Examples Figs 9-3 - 9-6. The figures in the suppl to chapter 8 showed astonishing differences in annual precipitation (60 cm yr-1) that cast would seem to cast doubt on modeled precip to be of any utility, say, in using the models to represent primary productivity, agriculture, etc. If similar differences persist in the AR5 runs, there should be some discussion of the point, especially regarding coupling to ecosystem models. [Stephen E Schwartz, USA]	Noted. Revised figures show intermodel spread explicitly. Have not included supplementary figures so as to keep chapter self-contained.
9-61	9	1				For all line graphs, throughout the chapter, please provide online tables of the data. For the individual models, use color or otherwise identify the individual models. [Stephen E Schwartz, USA]	Accepted concerning model identification where feasible. Rejected concerning table provision. It is beyond the scope of this assessment to provide access to processed model output that is already publicly available through the CMIP5 archive.
9-62	9	1				The whole chapter consists of a mix of CMIP3 and CMIP5 results. I understand that this is unavoidable for this draft, and will most probably be unavoidable for the final report as well. However, much care has to be put into making a clear separation. For instance, sec. 9.4.1.1 is about CMIP5, while 9.4.1.1.2 (which is probably wrongly numbered and should read 9.4.1.2 – there is no 9.4.1.1.1) is about CMIP3. Although correctly stated, this difference can easily be overlooked. [Andreas Sterl, Netherlands]	Taken into account. However, the depth of analysis of CMIP3 is so vastly greater, and quite possibly the difference between CMIP3 and CMIP5 perfomance sometimes quite small, that we absolutely must include comprehensive assessments of CMIP3 results. Will make difference clear.
9-63	9	1				overall, chapter 9 is in very well written and the intro discussion on how to test climate models is well done, as is the wide range of observational comparisons to more than just mean climatologies. [Bruce Wielicki, USA]	Noted. Thank you!
9-64	9	3	1	4	10	I suggest model-observation comparison studies should be more emphasized in AR5. Global observations of vertical cloud and humidity profiles, which were not avaliable during the AR4 era, are opening a new window for AR5 and should be included. [Jonathan Jiang, United States of America]	Noted - Chapter 9 extensively assesses model-data comparisons, inlcuding cloud and humidity profiles.
9-65	9	3	1	73	15	The authors are commended for a well written, concise and readable chapter. Obviously, much of the content awaits the results and analysis of the CMIP5 experiments, as much of the current discussion is based on CMIP3. The reviewer hopes the quality of the text and content is maintained in future drafts. [David Bader, USA]	Noted - thank you!
9-66	9	3	3	4	20	Evaluaton does everything except what is essential; provide evidence that they are capable of forecasting future climate. Without this they are useless. Evaluation is ultimately based on the opinions of those paid to	Rejected - Chapter 9 does include in Section 9.8 a connection between model performance of

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						provide the models and should be ignored. [VINCENT GRAY, NEW ZEALAND]	simulations of the past and the use of models in D&A and projections. The assessment is based on a vast body of published literature.
9-67	9	3	3			Despite this aim of simulating future climate, initially the chapter seems devoted merely to assessing the simulation of current climate, in all its detail. Commendably, it does much more, including consideration of trends, past climates and even sensitivity to GHG. It would help if more of the material and arguments in later sections (especially 9.8) were evident in the summary and introduction. [Ian Watterson, Australia]	Taken into account - The SOD Executive Summary contains much more of the detailed results of this Chapter.
9-68	9	3	4	3	6	Confidence in using climate models is based on careful evaluation of model performance, making use of increasingly comprehensive observationally-based data sets, well-designed model intercomparison and model-observation comparison activities. [Hui Su, United States of America]	Noted.
9-69	9	3	6	3	7	Here you give the "definition" for the content of the chapter, i.e. that it provides an assessment of climate model evaluation! OK! But in the following detailed descriptions of the evaluation of other models are also given which, to my understanding, are not the basis for AR5, e.g. Chemistry-Climate Models. As said before, this chapter should mainly cover the evaluation of those models which are used for the detailed discussions in other chapters and provide the basis for general conclusions of AR5. [Martin Dameris, Germany]	Taken into account - For the SOD, careful attention has been paid to the appropriate balance of models to be assessed here. Some models, though not used explicitly in later chapters, provide insight into model performace that is important to convey.
9-70	9	3	6	3	7	The construct "…provides an assessment of climate model evaluation…" is awkward and should be rewritten. It does draw attention, however, to the fact that the chapter does two things: It evaluates the models, and it also evaluates the methods used for model evaluation. This is good and should be called out explicitly in both the introduction and conclusions sections of the chapter. [David Randall, USA]	Accepted - SOD makes it clear that it is both the models and the evaluation methods that are being assessed.
9-71	9	3	8	3	8	Somewhere in this chapter, explaination of why these models are considered should be given. [Jonathan Jiang, United States of America]	Accepted - SOD Executive Summary includes a pointer to the use of the models assessed here.
9-72	9	3	8	3	8	While many people have a clear view of the distinction between AOGCM and ESM, I'm not sure that there is a systematic way of assigning a given model to one group or the other: is this a systematic or qualitative categorisation? [Martin Juckes, UK]	Noted - Bullet point on p. 9-3 l. 13/14 gives this definition.
9-73	9	3	8	3	20	A 5th category, energy balance models, is also referred to in the text; atmosphere only models (AGCMs) are also part of CMIP5 and are likely to be referred to in the literature at least. [Martin Juckes, UK]	Taken into account - AGCMs are defined in revised text.
9-74	9	3	10	3	20	Should Atmospheric General Circulation Models (AGCMs) also be useful for model diagnostic, especially in evaluating model performance in simulating the past and current climate? [Hui Su, United States of America]	Taken into account - some AGCM results are assessed.
9-75	9	3	13	3	13	climatically important this implies a priory knowledge. In practice, it is necessary to test the importance of "other factors" such as biogeochemical cycles on climate. Hence I suggest rephrasing and harmonization with the other three bullets. [Elisa Manzini, Germany]	Taken into account - text revised
9-76	9	3	13	3	14	the phrase "ESM" meaning an earth-system GCM is becoming widespread and is used throughout the FOD. But I think it is misleading, and misses the key point that these models are still GCMs at their core – a better phrase is needed which emphasises the important difference between Earth System GCMs and, say EMICs (which are also "earth system models"). If AR5 uses the phrase ESM then this will become set in stone. Now is an opportunity to come up with a better one. One suggestion would be "ES-GCM" (analogous to A-GCM, O-GCM, AO-GCM etc... for Atmos-, ocean-, or coupled atmos-ocean-GCMs) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Rejected - ESM for Earth-System GCM is now in widespread used, as witnessed by the naming convention in many CMIP5 models. AR5 is to assess the research and is thus tied to what the research community has decided to use.
9-77	9	3	22	3	23	".. uncertainty is understood and quantified": say something about need for stability of data on time scales of interest. [Martin Juckes, UK]	Noted - stability is subsumed under the characteristics listed here.
9-78	9	3	23	3	23	whose uncertainty is understood and quantified --> whose uncertainty should be (or must be) understood and quantified [Claudio Cassardo, Italy]	Taken into account. Text revised
9-79	9	3	23	3	23	These observational data have been described in earlier chapters. ← It will be useful to list what observational data are included in this chapter, especially satellite datasets, which have not been used much in this chapter. [Hui Su, United States of America]	Taken into account - a Table now provides details of observational data sets used.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-80	9	3	24	3	24	We should mention here that many of these new 'metrics' are observational based, especially from the collocated A-Train satellite observations. [Jonathan Jiang, United States of America]	Taken into account - metrics are defined in text as being based on observations. It would, however, be too narrow to single out the A-Train system here.
9-81	9	3	25	3	25	It would be of interest here, to summarize, by means of a short bullet list, the advancements in the area of model 'metrics'. The area of model metrics has already an history (for instance, Taylor diagrams existed before AR5), so to quantify the advancement would be very useful. [Elisa Manzini, Germany]	Taken into account - advances made by using model performance metrics is spelled out more explicitly in SOD Executive Summary.
9-82	9	3	30	3	30	Satellite simulators? It is not clear what is done. Is it like e.g. taking ouput according to the satellites measured coordinates? [Farahnaz Khosrawi, Sweden]	Noted - please see main text of chapter.
9-83	9	3	30	3	30	satellite simulators: are the pro and cons of this approach assessed? [Elisa Manzini, Germany]	Taken into account - please see main text of chapters.
9-84	9	3	30	3	33	The 'satellite simulator' method is new, but the uncertainty of current simulator -mostly 1-D forward model calculation - is quite large, and thus it may not be better than the standard comparision using the satellite retrieval. Also, satellite observations are also used in evaluating model simulated parameters other than clouds, e.g. water vapor, sea surface temperature, tropause heright, etc. [Jonathan Jiang, United States of America]	Taken into account - please see main text of chapters.
9-85	9	3	30	3	33	Another advance since the AR4 is the extensive use of `satellite simulators' in climate models. This involves on-line calculations which provide output more directly comparable to remote sensing observations from satellites. This approach is particularly valuable in evaluating the representation of clouds and cloud processes in climate models. In addition to clouds, standard satellite observations should also be used to evaluate many other modeled parameters such as water vapor, and tropopause height, sea surface temperature, rainfall, etc. Also, unlike the standard satellite retrievals, error-bars or uncertainties of the 'satellite simulators' have largely not been specified in this and previous chapters. [Hui Su, United States of America]	Taken into account - please see main text of chapters.
9-86	9	3	30			The terms "satellite simulators" and the variant used later ("instrument simulators") reflect the scientific community's language but are inaccurate. I have seen the term "observation proxy" used but this isn't much better, nor is the term "forward operator" as used in the data assimilation community. What is needed is a concise way to express the process of turning model variables into synthetic observations. [Robert Pincus, USA]	Noted - AR5 is to assess the research and is thus tied to nomenclature that the research community has decided to use. Description has been revised somewhat in main body of chapter.
9-87	9	3	31			What does 'on-line' mean here? I guess this means that the calculations are carried out while the model is running rather than at a post-proecessing stage. But this could be clarified. [Nathan Gillett, Canada]	Taken into account - text removed from ES.
9-88	9	3	31			Using such a technical term as on-line I think should be avoided in the executive summary. The simulators can and are run off-line on model level data also. [Gunilla Svensson, Sweden]	Taken into account - text removed from ES.
9-89	9	3	31			same caution as above comment; suggest changing "which provide" to "which can provide under some circumstances" [Duane Waliser, USA]	Taken into account - text removed from ES.
9-90	9	3	32	3	33	Why is that valuable for cloud and cloud processes? [Farahnaz Khosrawi, Sweden]	Taken into account - please see main text of chapters.
9-91	9	3	35	3	42	As written now, still quite vague. Possibly, an quite interesting addition, would be a critically assessment of the limitation of the "ensemble of opportunity", now at its phase 5 (eg, not a novelty). [Elisa Manzini, Germany]	Taken into account - text revised, and more discussion in main body of text.
9-92	9	3	37	3	40	The sentence is quite long and it would be wise to split it into two sentences (first sentence should then stop after available, l39 and then continue with "However, …". [Farahnaz Khosrawi, Sweden]	Taken into account - text revised.
9-93	9	3	37	3	40	Multi-model ensembles allow to take model uncertainty into account, this is important for prediction and projections when no observations are available at all. The current statement is a little narrow in scope. [Annarita Mariotti, U.S.A.]	Taken into account - text clarified.
9-94	9	3	37	3	40	The multi-model ensemble allows for some assessment of uncertainty in climate model capabilities in cases where suitable observations are not available, but more importantly it allows one to begin investigating the connection between particular model errors/biases and particular characteristics or process parameterisations in a model. ← It should be mentioned here or somewhere in this executive summary that one of improvements	Taken into account - text revised.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						since the AR4 is the use of global observations especially those from the satellite for assessment of uncertainty in climate model simulations (e.g. Jiang et al. 2012, "Evaluation of Cloud and Water Vapor Simulations in IPCC AR5 Climate Models Using NASA 'A-Train' Satellite Observations," J. Geophys. Res. http://mls.jpl.nasa.gov/library/JiangEtAl_JGR_IPCC.pdf). [Hui Su, United States of America]	
9-95	9	3	39	3	39	"importantly it allows one to begin investigating the connection…..": It is not clear what is meant. [Farahnaz Khosrawi, Sweden]	Taken into account - will aim at clarifying sentence, but this piece seems very clear.
9-96	9	3	44	4	10	In AR4 it was concluded that aerosol and cloud forcing have the largest uncertainty. AR5 should summarize the improvements made in this area. Also, recently study suggest that the model uncertainty in simulating the moisture and clouds is the largest in the upper troposphere above 300 hPa levels. This should be included here. [Jonathan Jiang, United States of America]	Rejected/Taken into account. Forcing uncertainty is comprehensivly assessed in Chapters 7, 8, and 10. Simulations of upper-level moisture and clouds is contained in Ch09 already.
9-97	9	3	44	4	45	"... ability of climate models to simulate historical climate, its change, and its variability" is expressed confusingly. Perhaps "ability of climate models to simulate past climate, including changes over time and variability,..." [Robert Pincus, USA]	Noted, but original wording seemed preferable to proposed alternative.
9-98	9	3	44			Biases in cloud ice water path are considerably smaller between CMIP3 and CMIP5 based on annual mean Taylor metrics (Li et al. 2012). This is an update to Waliser et al. (2009) that is cited later. A more complete comment (e.g. abstract) with the citation is given in a later comment should you decide to highlight this aspect of cloud improvement. [Duane Waliser, USA]	Noted - however, the Executive Summary must strike the right balance between brevity and comprehensiveness.
9-99	9	3	45	3	45	"change" How is it meant here? future climate change? If so, how can it be assessed its improvement? Please, be precise. [Elisa Manzini, Germany]	Noted -- included heading to make clear that all refer to 'historical'.
9-100	9	3	48	4	2	I understand that these bullets are based on very preliminary results. For the following draft, please be quantitative. [Elisa Manzini, Germany]	Taken into account - SOD is much more specific, using calibrated uncertainty language.
9-101	9	3	48	4	2	In addition, there are 4 classes of models considered (see lines 10-19), hence the question is: to which class of models is this list referring? [Elisa Manzini, Germany]	Taken into account - SOD is much more specific, using calibrated uncertainty language.
9-102	9	3	48	4	2	None of the executive summary of this chapter uses the calibrated IPCC language (likelihood and confidence assessments). Such assessments would be particularly crucial for these 5 bullet points. [Sonia Seneviratne, Switzerland]	Taken into account - SOD is much more specific, using calibrated uncertainty language.
9-103	9	3	48			A fuller evaluation of the degree to which model biases have changed since the AR4 depends on the analysis of the CMIP5 data. It should be noted that this statement may not apply to all models or it may not apply to the same degree for all models. [James Kinter, United States of America]	Taken into account - SOD contains more on CMIP3/CMIP5 comparison.
9-104	9	3	50	3	50	"…is well simulated…" --> "..are well simulated…" [Hai Lin, Canada]	Editorial - revised
9-105	9	3	50			I think the statement on "the diurnal cycle of surface…" is to strong, it is not demonstrated in the chapter (see further comments) [Gunilla Svensson, Sweden]	Taken into account - text revised and consistent with body of chapter.
9-106	9	3	53			The Executive Summary claims that the NAO is better represented in models than suggested in the AR4; however, the text (cf pp. 46-47) does not bear out this conclusion - all the cited work describes shortcomings in model simulations of the NAO. [James Kinter, United States of America]	Taken into account - text revised and consistent with body of chapter.
9-107	9	3	56			What about carbon uptake by the land? The models include this too, but perhaps this isn't mentioned because it is less realistic. This could be clarified. [Nathan Gillett, Canada]	Taken into account - revised version includes bullet on carbon cycle.
9-108	9	3		4		even with those few optimistic words between line 11 and 20 on page 4 --- a reader gets a overall negative impression about the ability of climate models throughout the summary; please add more hope --- or is there none? [Frank Kreienkamp, Germany]	Taken into account - text revised to provide more balanced presentation.
9-109	9	3				The executive summary should say a bit more on ESM and biogeochemical cycles. [Olivier Boucher, France]	Taken into account - revised version includes bullet on carbon cycle.
9-110	9	3				Executive summary "improved in many important respects". Insert a sentence regarding the caveats that many	Taken into account - text revised and consistent with

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						challenges persist especially at the regional scale in quantities like precipitation for specific problematic areas such as the South Pacific convergence zone, the southward extent of the Atlantic intertropical convergence zone, or the details of precipitation at the North American West Coast arrival of the Pacific storm tracks. [J. David Neelin, United States]	body of chapter.
9-111	9	4	1	4	2	This statement is not warranted, see Pielke Sr., R.A., and R.L. Wilby, 2011: Regional climate downscaling – what's the point? EOS. January 31 2012 pages 52-53 This article presents a list of reasons why multi-decadal regional climate forecasts are unable to provide skillful predictions to the impacts community. These can be summarized as: As a necessary condition for an accurate prediction, multidecadal global climate model simulations must include all first-order climate forcings and feedbacks. However, they do not. Current global multidecadal predictions are unable to skillfully simulate regional forcing by major atmospheric circulation features such as from El Niño and La Niña and the South Asian monsoon. While regional climate downscaling yields higher spatial resolution, the downscaling is strongly dependent on the lateral boundary conditions and the methods used to constrain the regional climate model variables to the coarser spatial scale information from the parent global models. Large-scale climate errors in the global models are retained and could even be amplified by the higher- spatial- resolution regional models. If the global multidecadal climate model predictions do not accurately predict large-scale circulation features, for instance, they cannot provide accurate lateral boundary conditions and interior nudging to regional climate models. Apart from variable grid approaches, regional models do not have the domain scale (or two-way interaction between the regional and global models) to improve predictions of the larger-scale atmospheric features. This means that if the regional model significantly alters the atmospheric and/or ocean circulations, there is no way for this information to affect larger scale circulation features that are being fed into the regional model through the lateral boundary conditions and nudging. For example, recent research indicates that terrestrial evaporation from the Eurasian continent contributes 80% of China's water resources. In this case, the regional model domain has to be large enough to include areas that are connected by soil moisture feedbacks. The lateral boundary conditions for input to regional downscaling require regional-scale information from a global forecast model. However the global model does not have this regional-scale information due to its limited spatial resolution. This is, however, a logical paradox because the regional model needs something that can be acquired only by a regional model (or regional observations). Therefore, the acquisition of lateral boundary conditions with the needed spatial resolution becomes logically impossible. Thus, even with the higher resolution analyses of terrain and land use in the regional domain, the errors and uncertainty from the larger model still persist, rendering the added simulated spatial details inaccurate. There is also an assumption that although global climate models cannot predict future climate change as an initial value problem, they can predict future climate statistics as a boundary value problem. However, for regional downscaling (and global) models to add value (beyond what is available to the impacts community via the historical, recent paleorecord and a worst case sequence of days), they must be able to skillfully predict changes in regional weather statistics in response to human climate forcings. This is a greater challenge than even skillfully simulating current weather statistics. [Marcel Crok, The Netherlands]	Rejected. This comment, arguing on the basis of the Pielke and Wilby Forum contribution to Eos, that is, an opinion piece, does not reflect the body of the published literature. The basis of the statement in the FOD Executive Summary is the assessment as laid out in Section 9.6, which addresses caveats and other issues. For further background, please also refer to the Section 11.10 of the Working Group I contribution to the AR4.
9-112	9	4	1	4	2	The value added by regional climate models is better understood when "perfect" (re-analyses) boundary conditions are used. The value added in projection/prediction mode has not been established. This is explained later on in the document but needs to be clarified here in the Executive Summary. [Annarita Mariotti, U.S.A.]	Taken into account - text revised and more detail provided in body of chapter
9-113	9	4	4	4	4	Rephrase "of course" to e.g. "Nevertheless" [Farahnaz Khosrawi, Sweden]	Accepted.
9-114	9	4	4	4	9	Here I suggest to bring up front, the persisting biases of the models. Also, given that there are 4 classes of models considered (see lines 10-19), persisting biases should be viewed within the model class. Eg, I would not expect an EMIC to reproduce the MJO. [Elisa Manzini, Germany]	Taken into account - text revised, but Executive Summary must be concise.
9-115	9	4	4	4	9	This paragraph fails to mention one additional issue with model performance, namely those cases where there is too little observational data to even evaluate the performance of the models. In such cases, the models could share common biases and thus uncertainty estimates based on multi-model range would fail to capture	Taken into account - Chapter 09 does discuss the value of model spread in assessing uncertainty and is clear that this can at best be a lower bound. Executive

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						the actual uncertainty of the projections. Mueller et al. (2011, GRL) provide such an example for evapotranspiration, where IPCC AR4 simulations present in some cases less range than observational estimates. An area where observational data is particularly scarce is for the evaluation of extremes (e.g. for soil moisture/droughts, runoff/floods, tropical cyclones, extratropical cyclones, thunderstorms, tornadoes). See IPCC SREX chapter 3, Section 3.2.1 for more details on this latter topic. Refs: Mueller, B. et al., 2011, Geophys. Res. Lett., VOL. 38, L06402, doi:10.1029/2010GL046230; IPCC SREX (2012), Chapter 3 (Seneviratne, Nicholls, et al. 2012). [Sonia Seneviratne, Switzerland]	Summary text revised.
9-116	9	4	4	4	9	After summarize the improvements since AR4 in the previous paragraph, this paragraph lists areas of model performance that remain to be improved. I suggest to add the following two items: (1) Cloud forcing was determined to have the largest uncertainty in previous ARs, this AR5 should state if there is any improvements in this regards. (2) Recent study has identified that the largest spread in model simulation of clouds and moisture resides in the upper troposphere, which should be included as an area remain to be improved in this AR5. [Hui Su, United States of America]	Taken into account - text revised.
9-117	9	4	4	4	20	The tone of the summary section overall is too negative in my view, and is also strangely structured. Models do a lot of things well, and this, in fairness, needs to be given more weighting. Leaving coments on what models are able to simulate well to the last few lines of the ES reads strangely (lines 17-20) -- discuss this earlier, before the errors. Highlighting the tropical Atlantic Ocean, the diurnal cycle of precipitation and the MJO, in particular, feel like cherry picking errors, and focusses on issues which are not self-evidently critical for getting large-scale climate features "correct enough" for confident projections or confident attribution. The summary on oceans (9.4.2.6) is much more positive on oceanic simulation overall, yet oceans are mentioned only once in the ES (apart from CO2 uptake), and that is a very negative coment. Furthermore, simply stating that "models have problems simulating... clouds" is too vague to be helpful, and could easily be misinterpreted. Models can simulate some aspects of clouds and associated radiation well, some not so well. The question (for this assessment) is whether these (and other) uncertainties are large enough to fundamentally undermine our confidence in model projections. The answer to that from the evidence in the chapter seems to me to be no, particularly in the light of multi-model ensembles and perturbed physics ensembles (and indeed multi generations of models) that allow sampling of this physical uncertainty, and tests for its impacts. Also bolstering our confidence in overall model sensitivity, for example, are fundamental physical arguments and understanding, simplified models, cloud resolving models, etc. [Robert Colman, Australia]	Accepted - SOD Executive Summary revised to provide more balance.
9-118	9	4	6			The uncertainties in preciptation run deeper than just diurnal cycle - the character (intenity/frequency) of preciptaion is not properly represented in models [Graeme Stephens, USA]	Taken into account - SOD is much more specific, using calibrated uncertainty language.
9-119	9	4	6			You could mention "temperature and salinity" (based on the Waliser et al. ref alluded to later) as they both impact density/circulatoin and the latter the global water cycle. [Duane Waliser, USA]	Taken into account - text revised substantially.
9-120	9	4	7	4	7	Julian is mis-spelled. [David Randall, USA]	Accepted - thank you!
9-121	9	4	7	4	7	add "the Inter-Tropical Convergence Zone" after "Madden-Julian Oscillation". Realistic simulation of ITCZ is still a major challenge. It should not be ignored when stating remaining model problems. [Guang Zhang, United States of America]	Taken into account - text revised.
9-122	9	4	7	4	9	The sentence "In some cases, model results are in general agreement with observations, but the observational uncertainty is sufficiently large as to render it impossible to make definitive statements about model quality." need to be improved or deleted. There is a previous chapter on observational data and you cannot doubt here on the quality of data to justify the reliability of model outputs. [Arona DIEDHIOU, France]	Taken into account - SOD Executive Summary will strive to be more specific. However, the statement from the FOD is correct - sometimes the observational knowledge simply isn't there.
9-123	9	4	7			Julien spelled wrong - Julian [Duane Waliser, USA]	Accepted - thank you!
9-124	9	4	11	4	11	Why is this? I would have thought the role for an IPCC assessment is not to focus on model weaknesses to provide motivation for model improvement. We have other forums for that. The role of an IPCC assessment is surely to take a balanced and clear eyed look at both model strengths and weakensses, with the aim of	Accepted - text revised.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						being able to answer the question: how reliable are they for use in climate change projections/attribution?. [Robert Colman, Australia]	
9-125	9	4	13	4	15	The claim that “contemporary climate and Earth system models are able to simulate modes of variability that arise spontaneously in the coupled system” is overly simple and general. This ability varies considerably among models and the mode of variability. [David Sauchyn, Canada]	Noted - text revised
9-126	9	4	16	4	16	"higher-order statistics such as correlations and teleconnections" I find "correlations" vague here. A bit more detail is warranted. [James Christian, Canada]	Noted - text revised
9-127	9	4	16	4	16	replace "correlations and teleconnections", as they are not examples of small scale features and higher-order statistics [Hai Lin, Canada]	Noted - text revised
9-128	9	4	16			This seems a little odd as a correlation or a teleconnection doesn't seem to be a particularly good example of a smaller scale feature or higher order statistic. [Duane Waliser, USA]	Noted - text revised
9-129	9	4	18	4	18	"Some fidelity" needs to be defined. The subjective aspects of fideltiy and confidence estimates must be made explicit. [Gerbrand KOMEN, Netherlands]	Noted - text revised
9-130	9	4	18	4	20	The last sentence on the change in climate sensitivity is not factual at all. It is suggested to compare the range of climate sensitivity as relevant for the AR4 with the updated range of climate sensitivity as reflecting the science of the AR5. There are clear statistical rules how to compare probability density functions of parameters and the language should reflect those. The value of climate sensitivity is quite üpolicy relevant therefore the authors should provide a very clear, scientifically sound language on any changes in the scientific knowlwedge. [Klaus Radunsky, Austria]	Noted - text revised and consistent with main body of text
9-131	9	4	18	4	20	When I read the excecutive summary first I though this sentence about the climate sensitivity did not belong here but since there is a section in the chapter describing how it is calculated I see why it is in the summary. However, I am not convinced that this is the right chapter for that information. [Gunilla Svensson, Sweden]	Rejected - as spelled out later in the Chapter, assessment of the spread of climate sensitivity, including the underlying physical mechanisms, is very much a matter of model evaluation.
9-132	9	4	19			My impression it hasn't changed since AR1 - if so, why not go back farther? [Duane Waliser, USA]	Noted - however focus here is on CMIP3 and CMIP5 generations of models
9-133	9	4	22			In the Executive Summary there are statements about climate model and Earth-System Model evaluation, but no statement about Chemistry-Climate Models. In the following sections of the chapter, Chemistry-Climate Models are mentioned several times, which is an imbalance. [Martin Dameris, Germany]	Noted -- focus in executive summary is on aspects relevant to model applications in later chapters. It cannot be comprehensive.
9-134	9	5	1	20	4	To my opinion, the first two sections of this chapter are much too long – or too short, depending on viewpoint: The two sections consist mainly of long lists of changes from AR4, without much explanations It is boring to read. To be interesting, explanations must be given, which would result in much more text. I doubt that this would be the aim of these sections. It suggest to substantially shorten these sections. Especially the 9.1.3.x.y sections are no necessary. [Andreas Sterl, Netherlands]	Taken into account -- text is being revised to focus on aspects needed to support later text, but it is felt necessary to provide the reader with aspects of model improvement since AR4.
9-135	9	5	5	5	7	Climate models are also very useful to understand better the internal cariability of the climate system in absence from transient external forcing. This is essential to attribute changes to external forcings and thus to separate "responses" from "noise". This aspect is missing here. [Marco Giorgetta, Germany]	Noted -- application of models to detection and attribution is explicit later in this section.
9-136	9	5	5	5	8	“primary tools … seasonal ...”: this is, I think, new since AR4 – something which should be emphasised and backed up by references. Statistical models remain an important tool. [Martin Juckes, UK]	Noted -- application of models to seasonal/interannual prediction is the focus of Chapter 11.
9-137	9	5	5	6	23	All the models are based on the "greenhouse" theory, which follows a real greenhouse in ignoring completely the real climate. Traditional meteorology has studied the climate for more than 200 years with scientific methods, observastions and techniques, and presents the current scientific view of the climate to all of us every night with the weather forecast. "Weather" is not just local it is global, and many media presentations emphasise its universality. The real climate is controlled by air pressure, air and ocean movemnets, wind directions and intensity, convection and evoporation/precipitation of water in all its forms, ocean oscillations, changes in the sun and in the earth's orbit, and volcanism. They all all play an intricate part. None of it could exist without the sun's radiation which warms the earrth by day warming that is psrtially lost by night..The	Rejected -- climate models are based on the same physical principle as global weather prediction models. They do not assume a static flat earth with permanent sunshine. Climate models and the physics that underly them have been described and reviewed extensively in the scientific literature. The definition of climate prediction vs projection is provided in the Glossary.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						replacement of this dynamic system, highly dependent on a knowledge of movement of fluids, has been replaced by an idealised, static, flat, unchanging earth with permanent sunshine, day and night, entirely dependent on "balanced"radiation exchanges.. It is completely unrealistic and is unsurpringly far less successful at forecasting future climate than traditional meteorolopgy, with all of its limitations. You even admit that they are incapable of forecasting by calling your model outputs "projections". [VINCENT GRAY, NEW ZEALAND]	
9-138	9	5	6	5	6	"making climate predictions" could appear misleading. Apart for predicting ENSO on seasonal scale, there exists to my knowledge no documented skill for useful climate prediction [Tor Eldevik, Norway]	Rejected -- various aspects of skill in climate predictions are assessed in Ch11
9-139	9	5	6	5	6	It is said "Climate models constitute the primary tools..for making climate predictions on seasonal to decadal time scales..". This statement neglects that now models are also used to represent and predict intra-seasonal variability. This seems very important especially with the great interest on extremes, typically occurring on shorter timescales. [Annarita Mariotti, U.S.A.]	Noted -- applications of models to climate prediction on various time scales is covered in Ch11
9-140	9	5	9	5	10	Which models are used in Chapters 10-12? This should be consistent with the subsequent descriptions of models and their evaluation in Chapter 9. [Martin Dameris, Germany]	Taken into account -- the text is being revised to make better links to the later chapters
9-141	9	5	10	5	10	"necessarily incomplete evaluation": In what sense? Concerning models considered or processes considered? [Farahnaz Khosrawi, Sweden]	Noted -- not all models can be evaluated in this chapter -- the limitations are discussed further in this paragraph.
9-142	9	5	10	5	10	"and so this is necessarily an incomplete evaluation" This sounds kind of obvious and useless. [Elisa Manzini, Germany]	Noted -- other review comments suggest that it was not obvious, so phrase will be retained.
9-143	9	5	10	5	16	why not including results from other CLIVAR? e.g. C20C (http://www.iges.org/c20c/) or CLIIPAS [ANNALISA CHERCHI, Italy]	Noted -- C20C experiments are included in CMIP3 and CMIP5 -- space limits preclude listing of every project.
9-144	9	5	13	5	16	what about project like ENSEMBLES or similar? [Frank Kreienkamp, Germany]	Noted -- some ENSEMBLES results are discussed in the section on downscaling -- list here in the introduction is not intended to be comprehensive
9-145	9	5	20	5	20	"discussed in Chapters 2 through 5" Chapter 6 also discusses observations relevant to ESMs that include carbon. [James Christian, Canada]	Editorial -- change 5 to 6. Thanks.
9-146	9	5	32	5	32	Section title: Wouldn't "model types" be more accurate than just "models"? [Farahnaz Khosrawi, Sweden]	Accepted - text revised.
9-147	9	5	32	5	43	Perhaps include reference to model families such as HadGEM2 (HadGEM2 Development Team, 2011; GMD 4, 723-757) which benefit from varying levels of complexity but a common physical framework. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Noted. This comment is relevant for addressing model independence as discussed briefly in Section 9.2.2.7 as well as evaluation of a fully-coupled model using comparisons of versions where components are swapped. However, we do not single this out as a specific model type here.
9-148	9	5	32			Section 9.1.2 Overview of Models to be Evaluated. Though mentioned in the first line of this section, simple energy balance models are not discussed in this chapter that I can see, even though reference is made to Chapter 9 for this purpose in Chapter 12 which uses in particular the MAGICC model. [Sarah Raper, United Kingdom of Great Britain & Northern Ireland]	Noted. We note that energy balance models are used for various purposes in the WG1 AR5 as suggested, however, these models are not evaluated in Chapter 9. We include certain EMICs with energy balance models in their evaluation. Thus, we did not include these in the current list in Section 9.1.2 because these are the models evaluated in the chapter.
9-149	9	5	34	5	34	Should 'However, a project differential warming' be 'projected'? [Mark Charlesworth, United Kingdom of Great Britain & Northern Ireland]	Editorial. This text does not appear in this location.
9-150	9	5	34		36	Text mentions simple energy balance models, but I find no mention of these. Particularly valuable recent work is	Noted. We note that energy balance models are used for various purposes in the WG1 AR5 as suggested,

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Held IM, Winton M, Takahashi K, Delworth T, Zeng F, Vallis GK (2010) Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing. J Climate 23:2418-2427. doi:10.1175/2009JCLI3466.1 Padilla, LE, Vallis GK, Rowley CW, (2011) Probabilistic Estimates of Transient Climate Sensitivity Subject to Uncertainty in Forcing and Natural Variability. J. Climate, 24: 5521–5537. doi: http://dx.doi.org/10.1175/2011JCLI3989.1 Schwartz S. E. (2012) Determination of Earth's transient and equilibrium climate sensitivities from observations over the twentieth century: Strong dependence on assumed forcing. Surveys Geophys. In press. http://www.ecd.bnl.gov/steve/pubs/ObsDetClimSensy.pdf [Stephen E Schwartz, USA]	however, these models are not evaluated in Chapter 9 given their long-term use in the IPCC process. We note that certain Earth-system Models of Intermediate Complexity (EMICs) use energy balance models as a component and are included in the evaluation of EMICs. Thus, we did not include these in the current list in Section 9.1.2 because these are the models evaluated in the chapter.
9-151	9	5	35	5	36	this is true since the beginning of climate science, the references used are too recent …. [ANNALISA CHERCHI, Italy]	Noted. However, these represent the latest discussions on this issue without requiring a complete account of the history.
9-152	9	5	40			"can be used" could be changed to "are used". Maybe something about what is lost should be said. [Gunilla Svensson, Sweden]	Accepted - text revised. Additional discussion on process representation to be included.
9-153	9	5	45	5	52	AOGCMs are still fundamental for studies aimed at the understanding of basi dynamical aspects in climate: this should be strenghtened despite the complexity of full ESMs [ANNALISA CHERCHI, Italy]	Noted. We agree that AOGCMs are still fundamental to understanding basic dynamical aspects of climate and consider 9.1.2.1 to reflect this withoutneed of further elaboration.
9-154	9	5	45			Atmosphere-only models are not mentioned here, but they have their place, particularly in light of systematic biases in sea surface temperature which are common in coupled models. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Noted. We agree that individual component models are used extensively in the research community but consider these to be a sub-class of the AOGCMs and do not separate them out here. The use of component models for model evaluation is discussed in Section 9.2.
9-155	9	5	51	5	51	There are other biogeochemical feedbacks which are important for climate which do not relate directly to the carbon cycle (eg methane, VOCs, ozone, maybe some aerosols) so I suggest removing "from the carbon cycle". [Olivier Boucher, France]	Taken into account - text revised. However, we have decided to delete the end of this sentence from "...in which biogeochemical feedbacks..." onwards. See response to comment 9-157.
9-156	9	5	51	5	52	"from the carbon cycle are less important than they are in century-scale projections" has this claim been demonstrated? [Elisa Manzini, Germany]	Accepted - Carbon cycle feedbacks are less important on seasonal to decadal timescales because carbon dioxide has a relatively long lifetime in the atmosphere. However, there is no single citation that demonstrates this, and we have have decided to delete the end of this sentence from "...in which biogeochemical feedbacks..." onwards.
9-157	9	6	2	6	2	ESM often include tropospheric ozone, and sometimes interactive methane too. [William Collins, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Text revised.
9-158	9	6	5	6	6	Why here this comment on computer requirement? Also AOGCMs, if for instance run at very high resolution, may be rather resource demanding and even more demanding of ESMs with highly parameterized biogeochemical cycles. [Elisa Manzini, Germany]	Noted. The computer efficiency of the model helps determine the intended use and guides the inclusion of the required components. This highlights the ability for alternative analyses using higher efficiency models where larger ensembles are required to sample uncerainty as discussed in later chapters.
9-159	9	6	19	6	23	Mention that RCMs are uncoupled. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Many RCMs are, indeed, uncoupled, and we make this more explicit. The

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							development of coupled RCMs was now addressed in Section 9.1.3.2.11 of the SOD.
9-160	9	6	27	6	27	section 9.1.3.1: What I find here missing is the fact that no clue is given, that the "mathematical expressions that best describe the system" is also related to the chosen / possible resolution of the numerical model in question, as well as the goal of the modeling. In other words, a characterization of what is "the system" at line 32 is very advisable. Possibly, the material of this section can also be best for a "box" on what is the physical & mathematical basis of a climate model, rather than in the main text of the assessment, given that these basics did not change between AR4 and AR5. [Elisa Manzini, Germany]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-161	9	6	27			The section 9.1.3.1 seems to be about decrasing uncertainty in model projections, but would benefit from including the aspects relating FAQ 1.1 Figure 1 (even if the figure itself is not particularly good) [Tor Eldevik, Norway]	Rejected - This section does not aim to talk about decreasing uncetainties. It merely aims at describing the model development process. The section has been merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning".
9-162	9	6	29	6	29	Almost all of the assumptions in the models violate basic physical principles. They assume the earth is flat,but it is round. They assume the sun shines all day and all night., It does not. They assume the earth is in equilibrium. It is not. They assume that you can use thermnodynamics equations. You cannot. They ignore the physical poribciples of the real climate, which include nthe importance of air pressure, air and ocean speed movement an circulation, convection and atmospheric heterogeneity, evaporatioin and condensation of water. If there are any basic physical principles it does comply with I defy you to strate them. [VINCENT GRAY, NEW ZEALAND]	Rejected - The assertions in this comment are incorrect. Modern GCMs do not assume the Earth is flat. The diurnal cycle of radiation is included. Air pressure, the circulation of the atmosphere and ocean are all part of the model's simulated variables. Convection, condensation and evaporation and heterogeneity are included through parametrizations as explained in this section. There are several recent textbooks highlighting the design of modern climate models, such asT. T. Warner, Numerical Weather and Climate Prediction, Cambridge University Press (January 17, 2011) and W. M. Washington, Introduction To Three-dimensional Climate Modeling, University Science Books; 2nd edition (May 16, 2005)
9-163	9	6	29	6	29	“well-known physical laws”: perhaps better than “fundamental” (AR4), but the well-known ones are not really the important ones as far as this report is concerned. As a sceptical reader, I would prefer that they were “independently verified” – that is verified independently of the climate modelling environment. [Martin Juckes, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-164	9	6	29	6	30	...(e.g., mass, energy and momentum conservation) ... [Marco Giorgetta, Germany]	Editorial. Agreed.
9-165	9	6	29	6	37	Should acknowledge that models also incorporate directly or indirectly a wealth of observational information, in the form of input into parametrisations. Models are not entirely theoretically based as is implied by point i), but can contain e.g. semi-empirical information providing constraints on parameters etc. [Robert Colman, Australia]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-166	9	6	29	6	37	This may be how EMICs are prepared, but for the complex ESMs you should either provide (or cite a source for) the mathematical expression of the complete set of physical laws or admit the process is less ideal. I believe that few, if any, modelling groups can support the overhead of keeping independent mathematical and coded representations of their ESM. [Martin Juckes, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-167	9	6	29	6	37	I believe that few, if any, modelling groups can support the overhead of keeping independent mathematical and coded representations of their ESM. So step (i) is applied on a component basis, (ii) may have more tuning than num. methods. [Martin Juckes, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-168	9	6	29	6	37	This paragraph misses an important point: A non-negligible portion of current climate models is not strictly based on physical laws. For instance, the representation of land surface processes requires the representation (and simplification) of biological processes. Similarly, chemical and biogeochemical processes represent a significant fraction of current climate models. [Sonia Seneviratne, Switzerland]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-169	9	6	30	6	30	"energy, momentum and mass conservation" [ANNALISA CHERCHI, Italy]	Accepted - text revised.
9-170	9	6	30			should it not be conservation of matter, energy and momentum? [Gunilla Svensson, Sweden]	Accepted - text revised.
9-171	9	6	32	6	33	The most basic mathematical problem facing the climate is in the turbulent motion of masses of air and water. The difficulty of solving this problem is the main barrier to long range forecasting of the climate. You think you have solved it by ignoring its very existence. [VINCENT GRAY, NEW ZEALAND]	Rejected - The turbulent motions on small scales are taken into account in all AOGCMs and ESMs through parametrizations. See Stensrud, D. J., 2007: Parameterization Schemes - Keys to understanding Numerical Weather Prediction Models. Cambridge University Press, Cambridge for examples on how this is done for the atmosphere.
9-172	9	6	32	6	37	Two principal steps are mentioned, but the chapter focuses entirely on the first (i.e., obtaining mathematical expressions) and unfairly neglects the second (i.e., numerical methods). I would like to see a paragraph discussing the evidence for the importance of numerical choices in climate simulations (e.g., Pfeffer 1992; Williamson & Olson 2003; Zhao & Zhong 2009). For example, Ren & Leslie (2011) have implemented a new time-stepping method (Williams 2009, 2011) into the SEGMENT-ice ice dynamics model, and have found that "This treatment has improved both spin-up and the conservation energetics of the physical processes." Therefore, the evidence is that model error is determined not only by the dynamics and physics, but also by the numerics. I would like to see all sources of model error covered fairly in the chapter. [References: Pfeffer, R., I. Navon, and X. Zou, 1992: A comparison of the impact of two time-differencing schemes on the NASA GLAS climate model. Mon. Wea. Rev., 120, 1381-1393; Ren, D., and L. M. Leslie, 2011: Three positive feedback mechanisms for ice-sheet melting in a warming climate. Journal of Glaciology, 57(206), 1057-1066; Williams, P. D., 2009: A proposed modification to the Robert-Asselin time filter. Mon. Wea. Rev., 137, 2538-2546; Williams, P. D., 2011: The RAW filter: An improvement to the Robert-Asselin filter in semi-implicit integrations. Mon. Wea. Rev., 139, 1996-2007; Williamson, D. L., and J. G. Olson, 2003: Dependence of aqua-planet simulations on time step. Quart. J. Roy. Meteor. Soc., 129, 2049-2064; Zhao, B., and Q. Zhong, 2009: The dynamical and climate tests of an atmospheric general circulation model using the second-order Adams-Bashforth method. Acta Meteor. Sin., 23, 738-749.] [Paul Williams, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-173	9	6	32			"…theoretical and observational…" (theory must usually be verified by experimentation and observation) [Richard Allan, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-174	9	6	33			Which are the derivations and the simplifications of the mathematical expressions that best describe the systems ? They should be indicated there. For example, is thermal conduction of air neglected compared to radiative transfer (see introductory comment) ? [François GERVAIS, France]	Noted. This description is of general nature and cannot list all the details of the simplifications made in climate models.The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-175	9	6	36	6	36	... of the discretized mathematical expressions ... [Marco Giorgetta, Germany]	Editorial. Agreed.
9-176	9	6	37	6	37	"latitude-longitude-height grid" pressure? [James Christian, Canada]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-177	9	6	43	6	43	"Model resolution" generally refers not just to the resolution in space, as implied here, but also to the resolution in time. I suggest changing this sentence to: "Numerical implementations allow for a choice of grid spacing and time step, often referred to as model resolution." [Paul Williams, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-178	9	6	44	6	44	"more accurate" does not mean with less errors, it should be emphasized [ANNALISA CHERCHI, Italy]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-179	9	6	44			"…more accurate models (although not necessarily more reliable simulations)…" [Richard Allan, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-180	9	6	45	6	45	... costs. An increase of the horizontal resolution by a factor of 2 typically increases the computational costs by a factor of 2^3. Currently ... [Marco Giorgetta, Germany]	Rejected - This is too much detail for this section.
9-181	9	6	45	6	47	As written this suggests that parameterisations are only needed because of limited model resolution. But of course this is only true of certain parameterisations, like convective parameterisations. Other parameterisations would be needed whatever the grid resolution. [Nathan Gillett, Canada]	Taken into account - The section has been merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-182	9	6	45	6	47	The language here needs to be improved. The statement that “certain processes are excluded” is correct, but excluded processes are, by definition, not represented. “Parameterised” processes are represented, though in an approximate form. The term “parameterisations” is widely used in the climate modelling community, but very unhelpful here. The point that needs to be explained is that some processes which can not be represented explicitly nevertheless need to be represented in some approximate form in order to avoid systematic errors in key budgets. A more descriptive phrase, such as empirical equations representing aggregated processes (e.g. transpiration of a range of plant types, or heat flux due to a unresolved turbulent eddies). [Martin Juckes, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning". The term parametrisation is in wide-spread use and has been retained here. However it's meaning has been extended to include not only unresolved processes but also those for which complexity must be reduced to be feasible (such as giogeochemical processes in vegetation).
9-183	9	6	45	6	47	"Excluded" is the wrong word. Parameterised processes arguably are not excluded, they are just included in a different way (i.e. not explicitly). I suggest changing the wording to: "certain processes are not explicitly included in the numerical solutions". [Paul Williams, UK]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning". The wording has been changed to "no longer explicitly represented".
9-184	9	6	46	6	46	Conceptual models are not always 'simple' -- far from it, e.g. in the case of modern clouds, convection or radiation parametrisations. This section does not do justice to this. A more accurate analogue might be one of coupled, complex, process models, some containing further layers of 'sub-models' within them. [Robert Colman, Australia]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-185	9	6	47	6	47	even "precipitation" should be included [ANNALISA CHERCHI, Italy]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-186	9	6	51			Apossible reference to make this point is A Unified Modeling Approach to Climate System Prediction Authors: Hurrell, James; Meehl, Gerald A.; Bader, David; Delworth, Thomas L.; Kirtman, Ben; Wielicki, Bruce Publication: Bulletin of the American Meteorological Society, vol. 90, issue 12, pp. 1819-1832 [David Bader, USA]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-187	9	6	53	6	53	Yes, aspects of the system are non-linear, and everything is non-linear eventually. But this statement 'highly non-linear' is so open ended it is not helpful. Many aspects of the climate system are in fact to a good approximation (and usefully) quasi linear under 'modest' perturbations (see AR4 fig. spm.5 for one example). This statement could be easily misconstrued and confused with chaotic initial value predictions etc. Reword to make this more precise, clarify here or elsewhere that it does not preclude model utility for a wide range of climate applications. [Robert Colman, Australia]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning". The sentence in question has been removed.
9-188	9	6	53	6	53	You emphasise the non-linearity in the climate system. Is there room to explore whether the climate system has preferred modes, and what these modes are? E.g, emergent constraints later in chapter. [Graham Feingold, United States of America]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning". Emergent behaviour is discussed further down in the Chapter.
9-189	9	6	53			Since the climate is highly non linear, the derivations and simplifications of the mathematical expressions referred to in comment 9 6 33 are of crucial importance. They should be recalled precisely somewhere in the AR5 report. [François GERVAIS, France]	Noted - This is beyond the purpose of this section. The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-190	9	6	54	6	56	This seems to confuse intiial condition uncertainty and model uncertainty. I would suggest that only initial condition ensembles are mentioned in bullett (iii), since this deals with dependence on initial conditions. Multi-model ensembles could be mentioned in bullett (i). [Nathan Gillett, Canada]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-191	9	6	56	6	56	CHAPTER 9: MODELS page 9.6, § 9.1.3.1, iii) line 56 non linear In order to describe the climate by a model in evolution, we have to explicit - a domain in R3 , comprising the general atmosphere (troposphere and stratosphere) , the land, the ocean, the soil, the biosphere, the cryosphere, with transmission properties, - the boundary conditions (including the solar irradiance, magnetic field, etc.) - all fluid mechanics and solid state mechanics, heat transfer, radiative transfer, changes of phase, etc.. where the state variables (as temperature, humidity, salinity , speed, energy, mass, flux, etc.) are function of a point M in the domain, of the time t and other variables as frequencies, etc. All these processes, separately are non linear functions of the point M, alone , and more ever because the matter in the domain of the phenomena is moving, at different scale of space and time. On the contrary of linear equations, there is no general methods to study or solve the nonlinear problem. We have just some collection of problems, some one stationary at the scale of time that interest us (i.e;, the data base of transition of level of vibration-rotation , in the GHG molecules) and others, problem of evolution with time (fluids dynamics of air in atmosphere or water in the ocean) [Robert DAUTRAY, France]	Noted - The section describes the various levels at which processes are treated in climate models, from the explicit numerical representation of known mathematical expressions to the parametrization of processes acting on unresolved scales. As such, it covers openly the issues highlighted by the reviewer. The section has been merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-192	9	7	10	7	10	... The next step in model development is the assembly, evaluation and calibration, also known as "tuning", ... (the calibration/tuning is needed after every assembly step. This should be understood by the reader, see also comment for line 14. [Marco Giorgetta, Germany]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-193	9	7	14	7	14	.. and the evaluation and calibration of that model ... [Marco Giorgetta, Germany]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning"
9-194	9	7	14	7	15	Again, 'why 'simple'? Indeed, what is meant by 'the need for simple process representations in all model components?' This undersells the sophistication in parametrisations in general -- (admittedly maybe not every single one!). [Robert Colman, Australia]	Taken into account - The section has been rewritten and merged with Box 9.1 into the new Box 9.1 called "Climate model development and tuning". The reference to "simple" has been removed.
9-195	9	7	21	8	23	Box 9.1: Who is the anticipated reader of this box? For a layman (e.g., policy maker) it is much too difficult and hard to follow. For instance, it does not become clear what tuning means and why it is necessary. The text should be rewritten by someone who has experience in writing for a lay public, e.g., a good journalist. [Andreas Sterl, Netherlands]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-196	9	7	23	8	21	In discussions of model tuning, it is important to also discuss whether and what observational constrains are used in the tuning practice. It will be more helpful to have list of parameters that commonly need adjustment or tuning in the models. [Hui Su, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-197	9	7	23			Box 9.1 It would be very helpful to actually explain that the model tuning of some parameters is done with a control (pre-industrial or present day forcings) simulation, rather than with a simulation driven with historically varying forcings. I am not aware of any modelling centre using a model development cycle that use historical simulations to tune their models - it would be far too computationally expensive. It is correct to mention (L52-L54) the difficulty to seperate knowledge about how previous models perform with a current model development, but is should be made clear that (unless evidence is available to say otherwise) that past long term changes in climate are not used to tune models in the development process. [Gareth S Jones, UK]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-198	9	7	25	7	26	The report states,"The global climate is an extremely complex system and there is no known set of equations that describes it completely." On this I agree. However, if you believe that the earth is a "complex system" why then is there no mention of complex systems theory (aka complexity theory), or of the way that the equations are formulated to take into account self-organizing and self-regulating aspects of the climate system? [Lee Klinger, USA]	Noted - Simple climate models make us of complex system theory building into them a-priori known self-organizing characteristics of the system. However, we have no complete knowledge of these characteristics. Consequently, AOGCMs and ESMs are built based on the principles described in Section 9.1.3.1 with complex behaviour an emergent feature of the models.
9-199	9	7	25	7	26	In fact one could likely postulate that no such sets of equations even exist, or that their number is not finite (if	Taken into account - The box has been rewritten and

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						they are to represent the climate system "completely"). Although the authors try to convey an important message with this sentence, I am not sure that it is useful as currently formulated. [Sonia Seneviratne, Switzerland]	merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-200	9	7	25	7	30	The division between fundamental and non-fundamental components might also be expressed as the division between what is explicitly resolved and what is abstractly (parametrically) represented. [Robert Pincus, USA]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-201	9	7	25	8	21	Honesty is the best policy. I really appreciate the straightforward discussion of this topic throughout this section; it is very well done. [Larry Thomason, United States of America]	Noted
9-202	9	7	26	7	29	Sentences should be changed as: "Climate models consists of fundamental component, i.e. the part of the system described by established theory, and a non-fundamental component, i.e. where important processes are included through the use of empirically derived (but physically plausible) relationship (see section 9.1.3). [ANNALISA CHERCHI, Italy]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-203	9	7	28	7	29	The use of empirically-derived but physically plausible relationships opens the way to question the results. What is the most important in Science ? To arrive to some results, even questionable, or to recognize than one is unable to reach a definite reliable prediction when uncertainties are too large ? [François GERVAIS, France]	Rejected - The comment does not address the text and is philosphical in nature.
9-204	9	7	34	7	35	The problem is not merely that "Many of these parameters are not well-constrained by theory or data", but also that deterministic parameterisation itself is not always appropriate. In these cases, "What is the correct parameter value?" is no longer the right question to be asking. I would like to see some discussion of the alternatives to deterministic parameterisation, including stochastic parameterisation. [Paul Williams, UK]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-205	9	7	41	7	42	"constitutes a skilful representation" note that at this point skill has not yet been defined [James Christian, Canada]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-206	9	7	45			A reference or two would help here [Gunilla Svensson, Sweden]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-207	9	7	48	7	48	The following can be added: ...Further it should be noted that the tuning process not necessarily leads to a unique and optimal set of parameters for a given model, but rather different combinations of tuning parameters can allow for equally plausible models for a chosen procedure, as demonstrated by Mauritsen et al. (2012). [Reference: Mauritsen, T. , B. Stevens, E. Roeckner, T. Crueger, M. Esch, M. Giorgetta, H. Haak, J. Jungclaus, D. Klocke, D. Matei, U. Mikolajewicz, D. Notz, R. Pincus, H. Schmidt, L. Tomassini, Tuning the climate of a global model, JAMES, 2012, submitted. [Marco Giorgetta, Germany]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-208	9	7	48			Box 9.1 I suggest adding sentence along the lines of "There have been recent efforts to develop systematic optimization methods, including multi-objective optimization that would quantify the trade-offs in adjusting internal model parameters within their error bounds to improve the match to observational criteria (e.g., Jackson et al. 2008, Neelin et al 2010, Covey et al 2012…) but these have not been applied to the current class of models." And/or refer forward to the discussion of perturbed parameter experiments page 9-16. [J. David Neelin, United States]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-209	9	7	50	7	57	the text needs work; not clear [Graham Feingold, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-210	9	7	50	8	21	Some centres (e.g. GFDL) use the observed 20th century global mean temperature evolution during their model development/tuning whilst others (e.g. Met Office) have a policy of not using 20th century trend information during their model development/tuning to avoid the risk of contaminating detection and attribution results. The authors may wish to consider a short discussion of these different philosophies and any implications for D&A. [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-211	9	7	51	7	51	Change 'deficiencies' to 'strengths and deficiencies' to reflect actual practise. [Robert Colman, Australia]	Editorial. Agreed.
9-212	9	7	53	7	53	change "potency" with "role" [ANNALISA CHERCHI, Italy]	Editorial
9-213	9	7	56	7	56	delete "it" before makes [Celeste Saulo, Argentina]	Editorial
9-214	9	7	57	7	57	any solution one liked': too vague, what is meant by this? Climate sensitivity? Representation of ENSO? If you mean a broad basket of climate aspects, including climate sensitivity change 'hard' to 'impossible'. Instead I suggest you drop 'solutions' (too much jargon for a FAQ anyway) and emphasise that detailed aspects of cliimate, particularly multiple different aspects (e.g. skill in simultaneously representing several modes of variability) are essentially impossible to tune for, but instead are in general emergent properties of the (usually complex) interaction of physical processes within the models (indeed providing tests for these processes). [Robert Colman, Australia]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-215	9	7	57	7	57	This sentence seems to be out of place. [Guang Zhang, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-216	9	7	57	8	2	To point out this concept, I suggest to reverse the sentence "every model that reasonably reproduces historical climate shows substantial warming as a result of increasing CO2" in "it is impossible to create a model able to reasonably reproduce historical climate that will not show substantial warming as a result of increasing CO2". [Claudio Cassardo, Italy]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-217	9	7	57			I'm not sure that I agree with this statement if it applies to any solution of one particular variable. I imagine that in many cases a model could be tuned to reproduce one particular scalar observed quantity, but because of the high dimensionality of the model, it could not be tuned to reproduce many observational fields at the same time. This might be clarified. [Nathan Gillett, Canada]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-218	9	7		7		Fig 9.1: as in the text atmosphere is used as an example this should be done also in the figure [ANNALISA CHERCHI, Italy]	Noted - Figure 9.1 has been removed.
9-219	9	8	1	8	2	This sentence seems to be out of place. [Guang Zhang, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-220	9	8	4	8	5	This is not entirely clear. Do you mean on needs all available data to run the model or to perform a model evaluation? It would be more adequate to write "…important for model evaluation and thus model development". [Farahnaz Khosrawi, Sweden]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-221	9	8	5	8	5	Emergent quantities is obscure and requires some explanation [Graham Feingold, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-222	9	8	13	8	14	You could add that tuning is necessary because there is an unavoidable gap between the equations of a model and "physical first principles." Over time the gap can be reduced, but it will never be eliminated. Similar gaps exist in many other fields, including astrophysics, biology, and engineering. [David Randall, USA]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-223	9	8	13	8	21	It should be mentioned at least in the summary here that, when applying the 'model tuning', differences between model output and observations can be considered significant only if they are not within uncertainties in the observed field etc. And, observationally based constrains on any parameter range must not be exceeded when doing any 'model tuning'. [Jonathan Jiang, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-224	9	8	14	8	14	climate simulations depend to a significant degree on the representation of poorly understood processes'. Way too open ended. What aspects? This is very misleading in this current form. Many important aspects important for climate and climate change are well modelled, and well understood. E.g. water vapour feedback which depends mainly on resolved circulation and the now reasonably well constrained and understood (small) changes to RH. Lapse rate changes which are well constrained by tropical adiabatic lapse rates. Atmospheric radiation, and its perturbations under forcing is very well understood in major aspects. [Robert	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Colman, Australia]	
9-225	9	8	15	8	15	"Leads to grades of the severity of tests that models can undergo" I think this could be written more clearly. [Graham Feingold, United States of America]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-226	9	8	15	8	15	The sentence seems to be garbled. "…it leads to grades of the severity…" [David Randall, USA]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-227	9	8	15			Confusing wording: "in that it leads to grades of severity of tests". Not sure what this means. [Bruce Wielicki, USA]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-228	9	8	18	8	19	"Models are not tuned to match a particular future climate; they are tuned to faithfully reproduce the past". I have particularly appreciated the previous introduction, but I personally do not like this sentence, especially the word "faithfully". One of the preferred arguments of the so-called climate change deniers is that the model tuning will make every model able to predict every behavior, or in other words that the model performance is strongly influenced by the tuning (so models are tuned in order to predict a temperature increment with increasing CO2). I think that, in these paragraph, it does not emerge clearly that this is not true. Tuning will refine a result, but generally does not change its general behavior. Thus, models "universally produce significant warming under increasing greenhouse gas concentrations" not because of the tuning, but because of their inner physical-chemical processes and/or their parameterizations. Tuning will just reduce the uncertainties. In other words, a model unable to simulate the increment of T for a given incrment of CO2 will never predict an increment of T under a particular tuning condition. [Claudio Cassardo, Italy]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-229	9	8	18	8	19	This sentence can be easily misunderstood. Models are not tuned to "faithfully reproduce the past". 1) if they were what is the point of this chapter? 2) models are tuned to try and replicate observed internal variability, but not past climate change. Should clarify that otherwise someone might get the impression that models are fully tuned to match past climate changes. [Gareth S Jones, UK]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-230	9	8	18			Models are "tuned to faithfully reproduce the past" does not seem to quite match the above text, which seems to suggest that models are deliberately tuned to represent current conditions (eg, TOA balance), and "tuning" to match the past (as in, the evolution of global temperature change) is more indirect... [Marcus Sarofim, USA]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-231	9	8	19	8	21	This is a very vague and unprecise sentence. What is "plausibly reproduce" and "significant warming"? It is also true that different models that all reproduce 20th century warming can give widely different climate sensitivities and therefore also different future climates. [Henning Rodhe, Sweden]	Taken into account - The box has been rewritten and merged with Section 9.1.3.1 into the new Box 9.1 called "Climate model development and tuning"
9-232	9	8	26	8	26	Figure 9.1 does not seem very effective. [Annarita Mariotti, U.S.A.]	Accepted - The Figure has been removed.
9-233	9	8	26			Figure 9.1: I know it is difficult graphically to include all the connections, but in this figure there is no direct connection between the carbon cycle and either the land or the biosphere. [James Christian, Canada]	Accepted - The Figure has been removed.
9-234	9	8	35	8	35	Replace "here" by "this assessment" or "this report". [Farahnaz Khosrawi, Sweden]	Editorial
9-235	9	8	38	8	38	remind that table 9.1 is incomplete, some information is missing for most of the models [ANNALISA CHERCHI, Italy]	Taken into account: table updated from CMIP5 questiononnaire
9-236	9	8	38			This comment refers to Table 9.1: It appears this table will assign a quantitative measure of “code independence” to each model. I am unaware of any studies that support such a measure. Without further independent support this estimate should be removed. It’s also not clear what value may be obtained from measures of code independence. [Robert Pincus, USA]	Taken into account: code independence removed from table and discussed intext
9-237	9	8	42	8	42	How is code independence defined? [Marco Giorgetta, Germany]	Taken into account: code independence removed from table and discussed intext
9-238	9	8	57	8	57	change to "cloud microphysical, aerosol processes, and their interactions" [Graham Feingold, United States of	Editorial. Agreed.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						America]	
9-239	9	8		9		section 9.1.3.2.1 is very weak: which are the real improvements? they should be listed clearly [ANNALISA CHERCHI, Italy]	Taken into account - Most of this section has been moved to Chapter 7.
9-240	9	9	1	9	14	Recent studies have identified that climate model simulations of upper tropospheric clouds and water vapor have the most spread (e.g. Jiang et al., 2012, http://mls.jpl.nasa.gov/library/JiangEtAl_JGR_IPCC.pdf). The model spread of upper tropospheric cloud water content is up to a factor of 500, compared to the model spread of lower tropospheric cloud water content on the order of 10. A major reason is likely to be the large spread in detrainment of ice and water vapor by deep convective clouds. This suggest parameterizations of deep convection remain highly uncertain in GCMs, and parameterized convective mass fluxes in the upper troposphere, which are major controls on detrainment, vary by factors of ten among GCMs (e.g. Folkins et al., 2006, J. Geophys. Res.,111, D23304, doi:10.1029/2006JD007325). [Jonathan Jiang, United States of America]	Taken into account - The Jiang et study is assessed in Section 9.4.1. in an extended cloud evaluation paragraph. The discussion of advances in cumulus parametrization has been moved to Chapter 7
9-241	9	9	2	9	14	You cannot possibly tackle convection because all your calculations are constrained by the absurd assumptions of the basic models.and your inability to deal with turbulent fluid flow. [VINCENT GRAY, NEW ZEALAND]	Rejected - Convection and turbulence in AOGCMs and ESMs are treated through parametrizations. While these treatments are not perfect, they have in numerous studies been shown to capturemany of the essential features of turbulent flows and convection in the atmosphere and ocean.
9-242	9	9	2	9	14	Suggest adding that "Recent research is also seeing the first use of kilometre-scale models in climate studies, which explicitly represent convection on the model grid without the need for a convective parameterisation scheme. These convection-permitting models have been shown to give a much more realistic representation of convection (Hohenegger et al., 2008; Wakazuki et al., 2008; Kendon et al., 2012)." [References: Hohenegger, C., P. Brockhaus, and C. Schar (2008) Towards climate simulations at cloud-resolving scales. Meteorol. Z., 17 (4), 383-394; Wakazuki, Y., M. Nakamura, S. Kanada, and C. Muroi (2008) Climatological reproducibility evaluation and future climate projection of extreme precipitation events in the Baiu Season using a high-resolution non-hydrostatic RCM in comparison with an AGCM. J. Meteorol. Soc. Jpn, 86 (6), 951-967; Kendon E. J., N. M. Roberts, C. A. Senior, and M. J. Roberts (2012) Realism of rainfall in a very high resolution regional climate model. Submitted to J. Climate] [Elizabeth Kendon, United Kingdom of Great Britain & Northern Ireland]	Taken into account -This discussion has been moved to Chapter 7.
9-243	9	9	2	9	30	The treatment of atmospheric convection remains one of the most critical areas in atmospheric models. While there have been no major developments in the basic approach to this problem, there have been important refinements in existing convection parameterisations. A long-standing weakness of convection in climate models has been the lack of sensitivity of the development of convective clouds to their environment… Improvements in representing the atmospheric boundary layer since the AR4 have focussed on basic boundary-layer processes, the representation of the stable boundary layer, and boundary layer clouds (Teixeira et al., 2008). Several global models have successfully adopted new approaches to the parameterization of shallow cumulus convection and moist boundary layer turbulence that acknowledge their close mutual coupling… In reading these paragraphs I recall that AR4 concluded that boundary layer clouds are related to the largest uncertainty in cloud forcing. Given the improvement made since AR4, it is important for this AR5 to review whether this case has changed. So far I have found any mention of it in this draft. [Hui Su, United States of America]	Taken into account -This discussion has been moved to Chapter 7.
9-244	9	9	4	9	9	Work on improving convection closure and trigger and on developing scale-aware convection parameterization should also be mentioned as it constitutes important research effort as well, for instance, the work by Wu et al. (2007), Arakawa and Jung (2010). [Guang Zhang, United States of America]	Taken into account -This discussion has been moved to Chapter 7.
9-245	9	9	6			After "Derbyshire 2004)." I suggest adding "There has been considerable advance in understanding and quantifying the observed processes involved, notably the impact of entrainment of lower free tropospheric air on the onset of deep convection (Holloway and Neelin 2009, Neelin et al. 2009). [J. David Neelin, United States]	Taken into account -This discussion has been moved to Chapter 7.
9-246	9	9	8	9	8	Connect the Derbyshere reference to the other references given before. [Farahnaz Khosrawi, Sweden]	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-247	9	9	8	9	8	remove ")" after "2008" [Hai Lin, Canada]	Editorial
9-248	9	9	8	9	14	The discussion of super parameterization could possibly be deleted. It has not been thoroughly tested as a replacement for cumulus parameterization. This method is still in a state of development and is similarin maturity to global convective scale modeling, which is mentioned much later in the chapter. The two should be discussed together. [David Bader, USA]	Taken into account -This discussion has been moved to Chapter 7.
9-249	9	9	9	9	9	"parameterizations" should be "parameterisations". [Masahiro Sugiyama, Japan]	Editorial
9-250	9	9	9	9	9	This section talks about super-parameterisation as a promising strategy. One related (but not necessarily directly relevant with IPCC) method is to use a global cloud-resolving model such as NICAM. It does show some improvement in simulating the MJO, for example (Miura et al. 2007, Science). I think it is desirable to show the possible future pathway for model improvement, citing NICAM studies (and other related ones). (Miura, H., M. Satoh, T. Nasuno, A.T. Noda, K. Oouchi, 2007: A Madden-Julian Oscillation event realistically simulated by a global cloud-resolving model. Science, 318, 1763-1765, doi:10.1126/science.1148443) [Masahiro Sugiyama, Japan]	Taken into account -This discussion has been moved to Chapter 7.
9-251	9	9	9	9	15	"In many climate models, cumulus parameterizations now calculate the typical verical velocity..." Please clarify that the models referred to treat only one updraft per column, whereas some models treat multiple subgrid clouds, each with its own vertical mass flux and transport gases and aerosols explicitly in each subgrid cloud and solve size- and composition-resolved cloud microphysics explicitly (Jacobson, M. Z., Development of mixed-phase clouds from multiple aerosol size distributions and the effect of the clouds on aerosol removal, J. Geophys. Res., 108 (D8), 4245, doi:10.1029/2002JD002691, 2003 as used in Jacobson, M. Z., Control of fossil-fuel particulate black carbon plus organic matter, possibly the most effective method of slowing global warming, J. Geophys. Res., 107, (D19), 4410, doi:10.1029/ 2001JD001376, 2002 and in Jacobson, M.Z., Short-term effects of controlling fossil-fuel soot, biofuel soot and gases, and methane on climate, Arctic ice, and air pollution health, J. Geophys. Res., 115, D14209, doi:10.1029/2009JD013795, 2010. [Mark Z. Jacobson, U.S.A.]	Taken into account -This discussion has been moved to Chapter 7.
9-252	9	9	11	9	14	There should be a statement whether or not superparameterizations are used in CMIP5 models listed in Table 9.1, and if any model does, please point it out. [Marco Giorgetta, Germany]	Taken into account -This discussion has been moved to Chapter 7.
9-253	9	9	12	9	12	"using so-called super-parameterisations" I think this term requires some explanation (see also 43/16). [James Christian, Canada]	Taken into account -This discussion has been moved to Chapter 7.
9-254	9	9	13		14	after promising results should be followed by “with improvement in the embedded CRM” and also cite Cheng and Xu (2011). [Kuan-Man Xu, USA]	Taken into account -This discussion has been moved to Chapter 7.
9-255	9	9	14	9	14	cross reference Chapter 7, if appropriate. [Graham Feingold, United States of America]	Taken into account -This discussion has been moved to Chapter 7.
9-256	9	9	14			I suggest adding "Stochastic parameterization of moist convection has been implemented in a number of models to address issues of missing subgrid scale variability (for review see Neelin et al. 2008)." [J. David Neelin, United States]	Taken into account -This discussion has been moved to Chapter 7.
9-257	9	9	16	9	24	what about clouds? it was one of the weakest component for AR4, what about now? a discussion on that should be given…. [ANNALISA CHERCHI, Italy]	Taken into account -This discussion has been moved to Chapter 7.
9-258	9	9	17	9	19	Though descriptions of the models in CMIP5 are still emerging, I believe relatively few models, and perhaps only one, have adopted assumed-PDF cloud schemes (the first of which was doi:10.1175/1520-0469(2002)059<1917:APPFTS>2.0.CO;2). Most changes to cloud schemes have been more incremental, so saying that there have “been some improvements in underlying algorithms” and to call out assumed-PDF schemes is perhaps overstating things. Note also that Chapter 7 contains lengthy descriptions of advances in cloud representation since AR4. [Robert Pincus, USA]	Taken into account -This discussion has been moved to Chapter 7.
9-259	9	9	19	9	21	This point might be expressed more clearly. One significant change between CMIP3 and CMIP5 is that many models now include prognostic aerosols that interact with clouds by affecting drop size, which then requires more sophisticated treatments of cloud microphysics. [Robert Pincus, USA]	Taken into account -This discussion has been moved to Chapter 7.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-260	9	9	21	9	23	This statement is somewhere between misleading and inaccurate. It is true that radiation schemes may have changed between CMIP3 and CMIP5 but the schemes used in CMIP3 accounted for the impacts of aerosols on the radiation field. [Robert Pincus, USA]	Taken into account -This discussion has been moved to Chapter 7.
9-261	9	9	23	9	24	Solely development or have the models also been improved? [Farahnaz Khosrawi, Sweden]	Taken into account -This discussion has been moved to Chapter 7.
9-262	9	9	23	9	24	To be specific, it is the treatment of variability in cloud properties, particularly that caused by multiple partly-cloudy layers, that has been improved since CMIP3. (The reference to the original algorithm referred to here is doi:10.1029/2002JD003322) [Robert Pincus, USA]	Taken into account -This discussion has been moved to Chapter 7.
9-263	9	9	24			Evidence suggests that including radiative treatments of precipitation may be important to obtain proper representation of radiative heating profiles and associated circulation dependencies (Waliser et al. 2011). Waliser, D. E., J.-L. Li, T. L'Ecuyer, and W.-T. Chen, 2011: The Impact of Precipitating Ice and Snow on the Radiation Balance in Global Climate Models. Geophysical Research Letters, 38, L06802, doi:10.1029/2010GL046478. [Duane Waliser, USA]	Taken into account -This discussion has been moved to Chapter 7.
9-264	9	9	34	9	35	After: The realistic treatment of the stable boundary layer remains difficult (Beare et al., 2006; Cuxart et al., 2006; Svensson and Holtslag, 2009). In fact, even at clear skies the modelling of the diurnal cycle over land remains a difficult issue (Svensson et al, 2011; Holtslag et al, 2012) References: Svensson G., A.A.M. Holtslag, V. Kumar, T. Mauritsen, G.J. Steeneveld, W. M. Angevine, E. Bazile, A. Beljaars, E.I.F. de Bruijn, A. Cheng, L. Conangla, J. Cuxart, M. Ek, M. J. Falk, F. Freedman, H. Kitagawa, V. E. Larson, A. Lock, J. Mailhot, V. Masson, S. Park, J. Pleim, S. Söderberg, M. Zampieri and W. Weng, 2011: Evaluation of the diurnal cycle in the atmospheric boundary layer over land as represented by a variety of single column models – the second GABLS experiment. Boundary Layer Meteorology, 140, 177-206; Holtslag,A.A.M., G. Svensson, S. Basu, B. Beare, F.C. Bosveld, J. Cuxart, 2012: Overview of the GEWEX Atmospheric Boundary Layer Study. ECMWF-GABLS workshop proceedings, ECMWF, Reading, UK, [Albert A.M. Holtslag, Netherlands]	Taken into account - The discussion of improvements to the stable boundery layer has been improved and the papers have been assessed and added to the reference list as appropriate.
9-265	9	9	34	9	35	This could be rephrased to: "The realstic treatment of the stable boundary layers (refs) and transitions (Svensson et al., 2011) remain difficult. (Svensson, G., A.A.M. Holtslag, V. Kumar, T. Mauritsen, G.J. Steeneveld, W. M. Angevine, E. Bazile, A. Beljaars, E.I.F. de Bruijn, A. Cheng, L. Conangla, J. Cuxart, M. Ek, M. J. Falk, F. Freedman, H. Kitagawa, V. E. Larson, A. Lock, J. Mailhot, V. Masson, S. Park, J. Pleim, S. Söderberg, M. Zampieri and W. Weng, 2011: Evaluation of the diurnal cycle in the atmospheric boundary layer over land as represented by a variety of single column models – the second GABLS experiment. Boundary-Layer Meteorology, 140, 177-206.) [Gunilla Svensson, Sweden]	Taken into account - The discussion of improvements to the stable boundery layer has been improved and the papers have been assessed and added to the reference list as appropriate.
9-266	9	9	37	9	44	There are also strong indications that gravity wave drag is even important on much smaller scales than currently represented (eg Steeneveld et al, 2008): Reference: Steeneveld, G. J., A. A. M. Holtslag, C. J. Nappo, B. J. H. van de Wiel, L. Mahrt, 2008: Exploring the Possible Role of Small-Scale Terrain Drag on Stable Boundary Layers over Land. J. Appl. Meteor. Climatol., 47, 2518–2530. doi: http://dx.doi.org/10.1175/2008JAMC1816.1 [Albert A.M. Holtslag, Netherlands]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-267	9	9	37	9	44	saturation unclear to what it is refereed to. The listed recent development concern the treatment of the sources of the waves. Please re-write, distinguishing between progress in representing the effect of the waves on the mean flow and the progress in representing the sources of the waves. Also, all literature listed is pre AR4, hence "more recently" does not apply to improvements since AR4. Please explain the context. [Elisa Manzini, Germany]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-268	9	9	37	9	44	This section put foreward very old results. The phrase "More recently" does not go very well with references from 1997 - 2000. This is an area that recent results (unpublished so far to my knowledge) suggest should receive more attention. [Gunilla Svensson, Sweden]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-269	9	9	37	9	54	The references here are all from year 2000 or previous years; there are a lot of newer publications available, e.g. Alexander et al., 2010 (review article); Geller et al., 2011; etc. Please update and use also some recent	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						peer-reviewed papers. [Martin Dameris, Germany]	recent advances have been added in revision.
9-270	9	9	37	10	16	As with the atmosphere you cannot deal wilh flow of the oceans because of the restraints of the basic model assumptions which are so far removed from reality [VINCENT GRAY, NEW ZEALAND]	Rejected - As discussed in Section 9.1.3.1 ocean models are based on known physical laws that represent reality. See textbooks about numerical modelling of the ocean, e.g. Numerical Modeling of Ocean Circulation, Ed. 1. ISBN : 0521781825or_9780521781824 Author : Miller, Robert N. Publisher : CAMBRIDGE UNIVERSITY PRESS 2007
9-271	9	9	46	9	46	"is becoming a common feature of GCMs" this is a sentence that would have been actual in the late 1990s, not in 2012. Indeed, middle atmosphere models with non-orographic gravity wave parameterizations were already used in the 2006 ozone assessment, see table 1 of Eyring, V., et al. (2006), Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, J. Geophys. Res., 111, D22308, doi:10.1029/2006JD007327 (note that a number of high top CMIP5 models can be traced to the underlying GCM of the 2006 ozone assessment) [Elisa Manzini, Germany]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-272	9	9	46	9	54	in this paragraph there are no references for the recent period: so which is the real upgrade from mid-90s? [ANNALISA CHERCHI, Italy]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-273	9	9	46	9	54	I have 2 main comments: (1) Is there a role for a non-orographic gravity waves in a "low-top model"? I ask this, because it is not clear if this discussion on non-orographic gravity waves is better placed here or in 9.1.3.3.7. (2) There is a body of literature on middle atmosphere models that can be assessed here (or in 9.1.3.3.7), with the aim of documenting what is the physical foundation of the high top models in CMIP5. This is fully missing. As a source of information to trace the relevant literature, see Eyring, V., et al. (2006), Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, J. Geophys. Res., 111, D22308, doi:10.1029/2006JD007327. But not only. [Elisa Manzini, Germany]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-274	9	9	47	9	47	Skip "i.e", the middle atmosphere is just composed of stratosphere and mesosphere. [Farahnaz Khosrawi, Sweden]	Editorial
9-275	9	9	49	9	49	non-orographic simply means that the sources of these waves are not linked to subgrid variations of orographic features. I suggest to keep it simple. [Elisa Manzini, Germany]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-276	9	9	51	9	51	... such GWD is essential to simulating realistically the temperature and wind structure in the high latitude middle atmosphere (...)), to the driving ... [Marco Giorgetta, Germany]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-277	9	9	53		54	Throw-away sentence on the stratosphere. Given the detail in previous paragraphs about other aspects a little more info, and some references, would be appropriate. Could delete this sentence in favour of section 9.1.3.3.7 but that doesn't have any references apart from CCMVal. [Joanna Haigh, UK]	Taken into account - In the discussion of gravity wave treatments, two recent review papers that summarize recent advances have been added in revision.
9-278	9	10	1			Units must be given! [Andreas Sterl, Netherlands]	I do not see where units are needed on this line. I suspect the reviewer entered the incorrect location for the comment. Maybe for degree to resolve eddies
9-279	9	10	1			these seem like relevant references as they found that better resolving ocean eddies in the tropical pacific improved the cold tongue bias in SST : Jochum, M., R. Murtugudde, R. Ferrari, and P. Malanotte-Rizzoli (2004), The impact of horizontal resolution on the heat budget of the mixed layer in ocean general circulation models, Journal of Climate, 18, 841-851. Jochum, M., and R. Murtugudde (2005), Temperature advection by tropical instability waves, Journal of Physical Oceanography, In press. [Duane Waliser, USA]	Taken into account and text revised Here is the edited text: Ocean components in CMIP5 models generally have horizontal resolution that is too coarse to admit mesoscale eddies **in the middle and high latitudes, though some provide a reasonable represenation of tropical eddies (Jochum et al 2004, 2005).**

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-280	9	10	2	10	3	The citations are somewhat displaced. [Farahnaz Khosrawi, Sweden]	Editorial. The citations are adjacent to the parameterization discussed, so that the reader knows what citations refer to which parameterization.
9-281	9	10	2			Bias – I am not sure whether most people would understand this. Better to write "difference between MME and observations". [Andreas Sterl, Netherlands]	I believe the reviewer has miss-cited the line number, as there is no discussion here of biases or MME.
9-282	9	10	6	10	7	Put the citations together. [Farahnaz Khosrawi, Sweden]	Editorial. The citations are adjacent to the parameterization discussed, so that the reader knows what citations refer to which parameterization.
9-283	9	10	7	10	8	"Another focus concerns specification of the eddy diffusivity, with many more CMIP5 models employing flow dependent diffusivities than CMIP3 models." Another focus concerns specification of sub-grid-scale mixing, with many more CMIP5 models employing flow dependent parameters than CMIP3 models. (Note that this is the only place the term "eddy diffusivity" occurs in reference to the ocean.) [James Christian, Canada]	Rejected. The term "eddy diffusivity" is technical but more precise than sub-grid-scale mixing. Non expert readers are refered to ocean modelling text books.
9-284	9	10	13	10	16	Please clarify and rephrase this sentence. It becomes not clear what the authors want to say. [Farahnaz Khosrawi, Sweden]	Editorial. Text rewritten to clarify meaning.
9-285	9	10	14	10	15	Skip "with" and "representative of this effort" and put the references in brackets [Farahnaz Khosrawi, Sweden]	Editorial. Agreed.
9-286	9	10	19	10	19	What is "dianeutral"? [Farahnaz Khosrawi, Sweden]	Rejected. The term "dianeutral" is indeed technical (belongs to cross-isopycnal physics) and non-expert readers are refered to ocean modelling text books.
9-287	9	10	21	10	23	"This has led to upgrades in the treatment of atmospheric radiation modules (Rotstayn et al., 2010), so that the radiative effects of aerosols can be included in a physically consistent fashion. The treatment of radiative effects of clouds has also been a signifcant development." Please clarify the referencing to indicate that the first online treatment of radiation transfer through discrete, size- and composition-resolved aerosols and discrete, size- and composition-resolved clouds in a global climate model was in Jacobson, M. Z., Control of fossil-fuel particulate black carbon plus organic matter, possibly the most effective method of slowing global warming, J. Geophys. Res., 107, (D19), 4410, doi:10.1029/ 2001JD001376, 2002 and in Jacobson, M.Z., The climate response of fossil-fuel and biofuel soot, accounting for soot's feedback to snow and sea ice albedo and emissivity, J. Geophys. Res., 109, D21201, doi:10.1029/2004JD004945, 2004 and in Jacobson, M.Z., Effects of absorption by soot inclusions within clouds and precipitation on global climate, J. Phys. Chem., 110, 6860-6873, 2006. [Mark Z. Jacobson, U.S.A.]	Rejected - This section is about assessing progress since AR4 and is not a review of all literature on this subject.
9-288	9	10	26	10	28	this is already written in 9.1.3.3.9 [ANNALISA CHERCHI, Italy]	Taken into account - text revised. These lines have been removed because their content is indeed duplicated in 9.1.3.3.9.
9-289	9	10	30	10	45	Too many scientific terms that are not explained are used in this two paragraphs. It is difficult for one outside the OBGC community to follow. [Farahnaz Khosrawi, Sweden]	Taken into account. Care has been taken to define necessary technical terms clearly in the chapter.
9-290	9	10	31	10	31	"Oceanic uptake of CO2 is highly variable" Ocean-atmosphere exchange of CO2 is highly variable in space and time [James Christian, Canada]	Accepted - text revised.
9-291	9	10	32	10	32	NPZD: expand the acronym [ANNALISA CHERCHI, Italy]	Rejected - this acronym is expanded in the same sentence that it is introduced.
9-292	9	10	32	10	32	"Most CMIP5 OBGC models are based on so-called NPZD-type models" This is not exactly true. About half of CMIP5 models that include ocean biogeochemistry are NPZD models. There are several that include multiple plankton groups, and some that appear to use the older HamOCC-style approach with only parameterized biology, and one that I can't really tell because all they have submitted is bulk quantities like primary production. There are several other labs that have multiple plankton groups but haven't submitted any data yet. [James Christian, Canada]	Accepted - text revised to "About half of the CMIP5 OBGC models are based on so called NPZD-type models...".
9-293	9	10	36	10	36	PFTs (would be good to cite some of the CMIP5 models that include PFTs here) [James Christian, Canada]	Rejected- insufficient space to list individual model

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							PFT classifications.
9-294	9	10	39	10	43	"Ocean acidification and the associated decrease in calcification in many marine organisms provides a negative feedback on atmospheric CO2". This is the biotic component: the abiotic component IS OF OPPOSITE SIGN, i.e., lower pH means higher pCO2(o) and less ocean uptake of atmospheric CO2. The biotic part may ultimately prove to be larger but it is less well constrained and may not materialize until some threshold is crossed (e.g. aragonite undersaturation in near-surface waters). The abiotic process is global, monotonic (over a range of pH that would take a huge amount of anthropogenic CO2 to exceed) and already well underway. Also, at present not all OBGC models that include calcification include acidification feedbacks on calcification rate. For example, CanESM (Christian et al 2010) does not. [James Christian, Canada]	Accepted - text revised. This paragraph has been rewritten to recognise the differing impacts of acidification on biotic and abiotic uptake of CO2.
9-295	9	10	44			which models [Frank Kreienkamp, Germany]	Accepted - text revised to list the CMIP5 models that include a sediment carbon reservoir.
9-296	9	10	50	10	50	"surface albedo and evapotranspiration". There is an obvious effect on the surface temperture and the long wave radiative balance [Eric Martin, France]	Noted. However, this is why we have written "..particularly through their effects on surface albedo and evapotranspiration".
9-297	9	10	50	10	52	As no individual episode or event in climate system should be directly (or even causally) connected with the statistical characteristics and their evolution (climate change, including changes in the state of the land-surface), the link between European drought 2003 and the state of the land-surface should not be declared so directly or unequivocally, as declared in the text. For this reason I suggest to use formulation "... might influence the severity..." or similar "weaker" formulation. [Ladislav Metelka, Czech Republic]	Accepted - text revised to "..for example, changes in the state of the land-surface may have played a part in the severity and length of the 2003 European drought".
9-298	9	10	50	10	52	Seneviratne et al. (2006, Nature) provided a more quantitative estimate that up to 60% of summer temperature variability was due to soil moisture-climate feedbacks in the Mediterranean region in late 20th-century climate, and a similar fraction in Central and Eastern Europe in late 21st century climate. Ref: Seneviratne, S.I., et al. Nature, 443, 205-209. [Sonia Seneviratne, Switzerland]	Accepted - text revised to cite this study.
9-299	9	11	4	11	4	If you are going to mention lack of interactive vegetation here you could forward-reference the discussion of dynamic vegetation in section 9.1.3.3.4. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Accepted - text revised. The following paragraph now mentions that some AR5 models do now include dynamic vegetation.
9-300	9	11	7	11	10	Several models include now vegetation dynamics, thus overcoming the porblem of prescribed vegetation maps. This aspect needs to be mentioned here becasue it has direct implications for the surface climate. [Marco Giorgetta, Germany]	Accepted - text revised. See response to comment 9-299.
9-301	9	11	12	11	13	True. But if a LSM behaves well in standalone version, it does not mean it will behave also well when coupled. The coupling can enhance some model deficiencies. Thus the success of the experiment with the standalone version of the LSM can be regarded as a condition necessary, but not sufficient. [Claudio Cassardo, Italy]	Noted.
9-302	9	11	12	11	13	This is not true if there is no observational data to evaluate these schemes (!). [Sonia Seneviratne, Switzerland]	Accepted - text revised. We now include a caveat to explain that although local validation data is available (e.g. from FLUXNET sites), there is still an issue with large-scale validation data.
9-303	9	11	13	11	15	This is a little too optimistic. In the case of some land variables, e.g. soil moisture, ground observations are still lacking in most part of the world (Seneviratne et al. 2010, Earth-Science Reviews; Dorigo et al. 2011, HESS). In the case of the Fluxnet network (Baldocchi et al. 2001), there are also still large parts of the globe without or only little observations. Ref: Dorigo, W.A., et al 2011, HESS, 15, 1675-1698. [Sonia Seneviratne, Switzerland]	Accepted - text revised. See response to comment 9-302.
9-304	9	11	14			It is wrong to use the word "validation" in this context, please use "evaluation" [Gunilla Svensson, Sweden]	Accepted - text revised to replace "validation" with "evaluation".
9-305	9	11	18	11	18	A section dealing with the representation of snow and permafrost in AOGCMs should be added here. [Thierry Fichefet, Belgium]	Taken into account -- the section is renamed as "Sea ice"
9-306	9	11	19	11	19	Ice: in the truth, it is only Sea ice [Claudio Cassardo, Italy]	Accepted -- the section is renamed.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-307	9	11	19	11	19	The title of this section should be "Sea Ice". [Thierry Fichefet, Belgium]	Accepted -- the section is renamed.
9-308	9	11	22	11	23	There is growing evidence that this classical rheological and dynamical representation, used in climate models, is unable to correctly model sea ice dynamics and deformation, especially towards small scales: Coon et al., JGR-C, 112, C11S90, 2007; Weiss et al., EPSL, 255, 1-8, 2007; Girard et al., JGR-C, 144, C08015, 2009; Hutchings et al., Ann. Glac., 52, 360, 2011, ... [Jerome WEISS, France]	Accepted -- text revised
9-309	9	11	25	11	27	new developments in sea ice rheology are also addressed. [Jerome WEISS, France]	Accepted - references added
9-310	9	11	30	11	43	What about sea ice rheology? [Thierry Fichefet, Belgium]	Accepted - references added
9-311	9	11	31	11	37	references are missing in this paragraph [ANNALISA CHERCHI, Italy]	Accepted -- references added
9-312	9	11	39			Pleasae see the following for an up to date evaluation of sea ice in cmip3: Kwok, R. 2011. Observational assessment of Arctic Ocean sea ice motion, export, and thickness in CMIP3 climate simulations. Journal of Geophysical Research 116: 10.1029/2011JC007004. [Duane Waliser, USA]	Noted
9-313	9	11	42	11	42	Remove the space between accurate and the comma. [Farahnaz Khosrawi, Sweden]	Editorial
9-314	9	11	43			What does 'but not for tracers' mean here? And why not? More explanation would be helpful. [Nathan Gillett, Canada]	N/A - text excluded
9-315	9	11	44	11	55	Thermodynamics clearly does not apply to this system, which is certainly not in equilibrium [VINCENT GRAY, NEW ZEALAND]	Noted. Indeed, classical thermodynamics deals with equilibrium states, concentrating on initial and final configurations, not on the processes involved in evolution. However, here, as well as in the vast scientific lterature, processes that affect the growth and melt of sea ice are referred to as thermodynamics.
9-316	9	12	1	12	1	Perhaps worth mentioning leads, which are also small-scale features [Graham Feingold, United States of America]	Rejected as not fitting the context here.
9-317	9	12	5	12	5	.... melting conditions, (e.g., Pedersen et al., 2009; Vancoppenolle et al., 2009b). More ... [Pedersen, C. A., E. Roeckner, M. Lu¨thje, and J.-G. Winther (2009), A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model, J. Geophys. Res., 114, D08101, doi:10.1029/2008JD010440.] [Marco Giorgetta, Germany]	Accepted --refernce added.
9-318	9	12	8	12	12	The following review paper should be added here: Hunke et al., 2011, The multiphase physics of sea ice : A review for model developers, The Cryosphere, 5, 989-1009, doi:10.5194/tc-5-989-2011. [Thierry Fichefet, Belgium]	Accepted - refernce added
9-319	9	12	12	12	14	Another important paper about this issue is: Vancoppenolle et al., 2010, Interactions between brine motion, nutrients and primary production in sea ice, J. Geophys. Res., 115, C02005, doi:10.1029/2009JC005369. [Thierry Fichefet, Belgium]	Accepted - refernce added
9-320	9	12	13	12	13	"and" has a different font size than the rest of the text. [Farahnaz Khosrawi, Sweden]	Editorial
9-321	9	12	20			It would be good to add a comment on the fidelity of modelled ice-atmosphere interactions here as this may be largely overlooked see for example papers by Clara Deser [Adam Scaife, United Kingdom of Great Britain & Northern Ireland]	Taken into account -- interaction with atmosphere and ocean is mentioned in the last paragrapph of the section. It is also covered in 9.4.3.
9-322	9	12	20			There are other strong feedback mechanisms operating besides the ice-albedo feedback. I suggest "…because of strong feedbacks in the system, theses errors are amplified. [Gunilla Svensson, Sweden]	Accepted - text modified as suggested
9-323	9	12	22	12	23	The second part of the title "Biogeochemical Feedbacks and the emerge of Earth System Modelling" does not really fit to the conent of this section. What you describe here is the coupling of surface (biogeochemical processes) and the stratosphere to the CCMs that only have considered the troposphere. I would suggest to	Accepted

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						skip "Biogeochemical Feedbacks" and keep "Emergence of Earth System Models" since they are the new models that resulted from the coupling of the processes. [Farahnaz Khosrawi, Sweden]	
9-324	9	12	22	12	23	The section title "New Components and Couplings: Biogeochemical Feedbacks and the Emergence of Earth System Modelling" suggests that in AR5 new modeling components are primarily in ESMs. While ESMs are a major new development, conventional AOGCMS also include many new model components and are still a major input for AR5, as also recognised in the subsections that follow. Some rewording of the section title may be useful. [Annarita Mariotti, U.S.A.]	Taken into account -- combined with comment 9-323.
9-325	9	12	22	15	25	In this section in particular it would be helpful to refer to models by name. For example 9.1.3.2 says that several ESMs are able to simulate a range of aerosol properties, but it doesn't say which ones in CMIP5. 9.1.3.3 says that some ESMs incorporate prognostic methane, without saying whether any of these are in CMIP5, and which models they are. This information is vital if this information on model formulation is to be useful in interpreting model output. [Nathan Gillett, Canada]	Accepte. Table of chemical/biogeochemical properties of ESMs in the CMIP5 archive added.
9-326	9	12	22			sec.9.1.3.3: it is just as important to list interactions between new components themselves as well as between a new component and climate. e.g. you list Carbon-cycle – climate interactions, and aerosol-climate interactions, but not aerosol-carbon cycle interactions. It would be a mistake to represent the Earth System as a central hub of physical climate with many "spokes" out to isolated ES components. Much better to show a fully interacting "web" of complexity – see, e.g. figure 3 in Collins et al., 2011, GMD paper on the HadGEM2-ES model. (dust-ocean BGC, ocean DMS-aerosol-clouds, veg-chemistry interactions... the list goes on) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Accepted -- web of connectivity among processes is now discussed in 9.1.3.3.10
9-327	9	12	25	12	25	I see discussion of Carbon Cycle here, but should Water Cycle also be discussed somewhere in this chapter (not in this sub-session)? [Hui Su, United States of America]	Noted -- Comment not pertinent to this section.
9-328	9	12	25	14	29	There are too many vague statements: "many of the important biogeochemical cycles", "some models ...", "in some cases.."; try to tie to specific information, link to table 9.1 if possible. [Martin Juckes, UK]	Taken into account -- combined with comment 9-325.
9-329	9	12	28	12	29	"the natural sources and sinks of CO2 and CH4, the two most important long-lived greenhouse gases" I think N2O should be mentioned here too. [James Christian, Canada]	Taken into account -- The sentence is correct as stands since it refers to the top two WMGHGs.
9-330	9	12	29			'ESMs incorporate many .. processes' please provide information not phrases [Frank Kreienkamp, Germany]	Taken into account -- combined with comment 9-325.
9-331	9	12	29			The abbreviation LLGHG is introduced and never used again, so should be omitted. [Philip Mote, USA]	Accepted -- acronym removed.
9-332	9	12	31		32	'Alternatively, when forced with specified concentrations, one can diagnose these sources (with feedbacks included)' --- this sentence includes no information [Frank Kreienkamp, Germany]	Rejected -- This experimental protocol is detailed at length in Hibbard et al 2007 (EOS).
9-333	9	12	35	12	36	these seem fairly random references to coupling carbon cycle into GCMs. Better to cite an overview paper like Friedlingstein et al (2006, J.Clim) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account -- Overview papers subsequent to Friedlingstein have been cited.
9-334	9	12	37	12	37	... change (Schurgers et al., 2008; Jungclaus et al. 2010 [Jungclaus et al., Clim. Past, 6, 723–737, 2010, doi:10.5194/cp-6-723-2010] [Marco Giorgetta, Germany]	Taken into account -- combined with 9-333.
9-335	9	12	39	12	48	"...several ESMs are currently capable of simulating the mass, number, size distribution, and mixing state of interacting multicomponent aerosols (Bauer et al., 2008)." Please clarify that this has been done for awhile in some models. The first model to treat these properties was the GATOR-GCMOM model, in Jacobson, M. Z., Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols, Nature, 409, 695-697, 2001 and Jacobson, M. Z., Control of fossil-fuel particulate black carbon plus organic matter, possibly the most effective method of slowing global warming, J. Geophys. Res., 107, (D19), 4410, doi:10.1029/2001JD001376, 2002 and Jacobson, M.Z., Effects of absorption by soot inclusions within clouds and precipitation on global climate, J. Phys. Chem., 110, 6860-6873, 2006. Also, the Bauer et al. study did not include discrete size resolution (they treated modes), whereas the studies above did. [Mark Z. Jacobson, U.S.A.]	Taken into account -- Focus is on models used in CMIPs
9-336	9	12	40	12	42	any references? [ANNALISA CHERCHI, Italy]	Taken into account -- combined with comment 9-325.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-337	9	12	40	12	42	"Many ESMs now include the basic features…so represent both the direct effect of sulphate aerosols, along with some of the more complex indirect effects." Please clarify that this has been done in some models for over a decade. The first explicit treatment of indirect effects (e.g.,growth of vapor onto size- and composition-resolved aerosol particles to form size- and composition resoled hydrometeor particles) and other cloud feedbacks in a global model was in Jacobson, M. Z., Control of fossil-fuel particulate black carbon plus organic matter, possibly the most effective method of slowing global warming, J. Geophys. Res., 107, (D19), 4410, doi:10.1029/ 2001JD001376, 2002. [Mark Z. Jacobson, U.S.A.]	Taken into account --Focus is on models used in CMIPs
9-338	9	12	46	12	48	It is not clear why the representation of aerosols is an "important source of uncertainty in the simulation of climate sensitivity". Not directly at least, it is a source of uncertainty in the simulation of the aerosol RF which makes it difficult to constrain the climate sensitivity from observed changes. But overall aerosol effects probably have a limited impact on the most important feedbacks. [Olivier Boucher, France]	Accepted -- Reference to climate sensitivity was to compensating errors discussed in Kiehl paper. However, the reference is confusing in this context and has been removed.
9-339	9	12	48	12	48	Cross reference Chapter 7, section 4 [Graham Feingold, United States of America]	Accepted -- cross reference added
9-340	9	13	1	13	20	Does the use of DGVMs integrated into ESMs (a) improve precision of projected warming, (b) provide better modelling of the impacts than can be achieved with off-line models and/or (c) improve understanding of uncertainties? [Martin Juckes, UK]	Accepted -- discussion of feedback from shifting vegetation on physical climate from Clark et al (2011) on the JULES model added.
9-341	9	13	1			Section: 9.1.3.3.4: How does vegetation influence climate or trace gases in general? [Farahnaz Khosrawi, Sweden]	Taken into account -- Impacts on trace gases is already are listed at end of this subsection.
9-342	9	13	2			First sentence could derive some general backup from : Bergengren, J. C., S. L. Thompson, D. Pollard, and R. M. DeConto (2001), Modeling global climate-vegetation interactions in a doubled CO2 world, Climatic Change, 50(1-2), 31-75. Bergengren, J.C., D. E. Waliser, Y.L. Yung, 2011: Ecological Sensitivity: A Biospheric View of Climate Change, Climatic Change, 107:433–457, DOI 10.1007/s10584-011-0065-1. [Duane Waliser, USA]	Accepted -- references added.
9-343	9	13	5	13	5	Put the citations together into one set of brackets. [Farahnaz Khosrawi, Sweden]	Editorial
9-344	9	13	6	13	6	Isn't that something also the other models are capable of? [Farahnaz Khosrawi, Sweden]	Taken into account -- There is no implicit or explicit claim of exclusivity associated with DGVMs.
9-345	9	13	7	13	7	Incorperation into which kind of models? ECMs, GCMs or AOGCMs. Please add this information. [Farahnaz Khosrawi, Sweden]	Taken into account -- the kinds of models incorporating DGVMs is stated two sentences earlier in the paragraph.
9-346	9	13	8	13	8	Why flora? Why is it important? [Farahnaz Khosrawi, Sweden]	Taken into account -- "flora" replaced with "vegetation"
9-347	9	13	19			Fires might be included in some ESMs, but are they included in the CMIP5 simulations? In the case of CanESM2, they are not. [Nathan Gillett, Canada]	Taken into account -- METAFOR database does not include a specific question regarding fire parameterizations
9-348	9	13	23	21	29	The abbreviation LULCC is introduced on page 13, used once, and then reintroduced on page 21. For ease of reading, I suggest omitting and just using the words land use and land-cover change in all 4 instances. [Philip Mote, USA]	Taken into account -- combined with comment 9-349
9-349	9	13	24	13	25	"RCP" and "LULCC": Are these abbreviations really necessary or useful? In my opinion they make the text more difficult to read. [Farahnaz Khosrawi, Sweden]	Accepted - LULCC acronym has been dropped. However, RCP is an acronym for a core component of CMIP5 and has therefore been retained.
9-350	9	13	33	13	53	This section should explain whether the term "CCM" is use to refer to models with tropospheric chemistry, stratospheric chemistry, or both. [William Collins, United Kingdom of Great Britain & Northern Ireland]	Accepted -- Term clarified. 13 of 15 models considered in CCMVal extend into the mesosphere.
9-351	9	13	33	13	53	Are the results of CCM simulations used in other chapters of AR5? See coments above. [Martin Dameris, Germany]	Taken into account -- Pertinence is established by sentence in this section reading "Several of the stratospheric chemistry-climate models 46 evaluated

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							in CCMVal have been incorporated into ESMs and are part of the CMIP5 ensemble.".
9-352	9	13	33	13	53	"Chemistry-climate models." Please refer to Zhang, Y., Online coupled meteorological and chemistry models: history, current status, and outlook, Atmos. Chem. Phys., 8, 2895-2932, 2008 for a discussion of online-coupled chemistry-climate models and their history. [Mark Z. Jacobson, U.S.A.]	Accepted -- reference added.
9-353	9	13	34			May affect? A clear statement would be necessary. [Martin Dameris, Germany]	Taken into account -- combined with comment 9-355.
9-354	9	13	37			Shift in tropopause height? In which direction? [Martin Dameris, Germany]	Accepted -- direction of shift is now specified.
9-355	9	13	48			ozone may increase? A clear statement is necessary. Please provide relevant references [Martin Dameris, Germany]	Taken into account -- A clear statement is not possible, since the future strat-trop exchange cannot be projected with certainty. References have been added.
9-356	9	13	49	13	49	Influx of what? Trace gases (if yes which)? Aerosols? [Farahnaz Khosrawi, Sweden]	Accepted -- phrase changed to "influx of ozone"
9-357	9	13	50	13	50	[tbc]? [Claudio Cassardo, Italy]	Editorial
9-358	9	13	55	14	5	The first high-top/low-top papers are being submitted (e.g. Hardiman et al. "The effect of a well resolved stratosphere on surface limate: Differences between CMIP5 simulations with high and low top versions of the Met Office climate model" submitted to J. Climate). The high-top models are found necessary to represent the QBO, the impact of stratospheric sudden waramings on surface climate, and some ENSO teleconnections. [William Collins, United Kingdom of Great Britain & Northern Ireland]	Taken into account -- this paper is now referenced in the text.
9-359	9	13	56	14	5	There are a number of publications that directly address the role of stratospheric dynamics on climate change, so far however not on CMIP5 (on CMIP5: in plan). Hence, a more focused assessment of literature is needed. Here I list two relevant papers, and I suggest to the authors to also assess the literature referred in these papers: Karpechko, A. Y. and E. Manzini, 2011: Stratospheric influence on tropospheric climate change in the Northern Hemisphere. J. Geophys. Res (in press) and Scaife, A., T. Spangehl, D. Fereday, U. Cubasch, U. Langematz, H. Akiyoshi, S. Bekki, P. Braesicke, N. Butchart and M. Chipperfield, et al. (2011), Climate change projections and stratosphere–troposphere interaction . Clim. Dyn., DOI: 10.1007/s00382-011-1080-7 [Elisa Manzini, Germany]	Accepted -- these publications have been traced forward from their citations
9-360	9	13	57			Is it not more correct to say that it influences the "tropospheric circulation" than "surface climate". [Gunilla Svensson, Sweden]	Accepted - text revised.
9-361	9	14	3	14	4	"subset … high top" not clear that is actually a "subset". Obviously, table is incomplete. But a rough look at is seems to suggest that at least half of the CMIP5 models are actually "high top". This is possibly a point that might be of interest to be assessed further, as metadata are becoming available. [Elisa Manzini, Germany]	Noted
9-362	9	14	8	14	8	The amount of melt water [Graham Feingold, United States of America]	Accepted - text revised.
9-363	9	14	8	14	14	It should be noted that the there are large uncertainties attached to the modelling of ice sheets and the acceleartion of ice loss - the processe are complex and occur on scales not yet resolved by ESMs - one of the importnat topics is the way oceans interact with ice sheets - a refernce to ch 4 may be needed [Graeme Stephens, USA]	Accepted - text revised and reference to Chapter 4 and process complexity added.
9-364	9	14	9	14	12	The wording here might be read as implying that coupled ice sheets are included in the CMIP5 simulations, but as far as I know, they are not. [Nathan Gillett, Canada]	Accepted -- text revised to eliminate msireading / misinterpretation of this sentence.
9-365	9	14	24			Additional support comes from: Murtugudde, R., J. Beauchamp, C. McClain, M. Lewis, and A. Busalacchi (2002), Effects of penetrative radiation on the upper tropical ocean circulation, J. Climate, 15, 470-486. [Duane Waliser, USA]	Accepted -- reference added.
9-366	9	14	31			Section 9.1.3.4: there should be mention of vertical resolution as well as horizontal resolution, as vertical resolution (not just the number of levels but their actual spacing) can make at least as much difference to	Taken into account - text revised. A mention on vertical resolution is added.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						model results as horizontal resolution. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	
9-367	9	14	33	15	12	Please give some information about vertical resolution as well as horizontal. [Paul Matthews, United Kingdom of Great Britain & Northern Ireland]	Taken into account - text revised. A mention on vertical resolution is added.
9-368	9	14	33			Section 9.1.3.4.1: Why does the high resolution not give the desired effect? [Farahnaz Khosrawi, Sweden]	Because of uncertainty in parameterizations. Text revised to clarify it.
9-369	9	14	35		47	supporting reference may be useful [Muhammad Amjad, Pakistan]	Table 9.1 supports this statement.
9-370	9	14	35			I think it would be good to mention that this concerns both horizontal and vertical grid [Gunilla Svensson, Sweden]	Taken into account - text revised. A mention on vertical resolution is added.
9-371	9	14	40			suggest that you change to "are not automatically (or immediately) realized" [Gunilla Svensson, Sweden]	Taken into account - text revised in a way consistent with the nuance suggested.
9-372	9	14	42	14	43	Here and throughout the chapter, "resolution" and "grid spacing" are used interchangeably. However, these two terms mean different things. "Resolution" refers to the smallest feature that can reasonably be resolved with good numerical accuracy; it is typically an order of magnitude larger than the "grid spacing". For example, the grid spacing might be "roughly 1 to 2 degrees for the atmospheric component", but the resolution is more like 10 to 20 degrees. [Paul Williams, UK]	Taken into account - text revised. Indeed, the term 'resolution' here is meant to refer to grid spacing, which is one of conventional uses of the term. More explicit definition of the term is added.
9-373	9	14	42			Would be nice with some general information on vertical resolution also and if any changes since CMIP3 [Gunilla Svensson, Sweden]	Taken into account - text revised. A mention on vertical resolution is added.
9-374	9	14	45	14	45	For comparison with the resolution of the regional climate models, here the global climate model resolutionof 1/2 degree should also be given in km. [Annarita Mariotti, U.S.A.]	Noted - but decided not to give the conversion in the text. We think most readers don't need it.
9-375	9	14	52			I disagree that this can be described as a transition from laminar to turbulent, only large scale eddies are resolved and the flow is basically laminar. Eddy permitting is the word that should be used. [Gunilla Svensson, Sweden]	Accepted - text revised. "turbulent" is changed to "eddy permitting".
9-376	9	14	55	14	55	Skip breackets around McClean citation. [Farahnaz Khosrawi, Sweden]	Editorial
9-377	9	14	56	14	56	Here it sounds that this high resolution is worth to be applied since some effects are better captured. So, what is the problem? [Farahnaz Khosrawi, Sweden]	Noted - The "problem" is that poorly resolved flow, namely those that do not resolve mesoscale eddies, fail to capture the full dynamical range relevant to ocean climate phenomena. As the models are pushed to finer resolution, they have shown enhanced abilities to simulate such important features as boundary currents, eddy-topography interactions, air-sea interactions so long as atmospheric models are also fine scale. These details are fully discussed in the noted citations, and exposing more for the IPCC AR5 is beyond its scope.
9-378	9	15	1	15	1	remove "general" [ANNALISA CHERCHI, Italy]	Accepted - text revised.
9-379	9	15	4	15	12	A series of papers appeared recently that noted the benefits and limitations of using higher spatial resolution in global atmospheric models. Jung et al. (Jung, T., M. J. Miller, T. N. Palmer, P. Towers, N. Wedi, D. Achuthavarier, J. M. Adams, E. L. Altshuler, B. A. Cash, J. L. Kinter III, L. Marx, C. Stan, K. I. Hodges, 2011: High-Resolution Global Climate Simulations with the ECMWF Model in the Athena Project: Experimental Design, Model Climate and Seasonal Forecast Skill. J. Climate (online). http://dx.doi.org/10.1175/JCLI-D-11-00265.1) showed that higher resolution is required to accurately simulate blocking, which is a major determinant of European climate in both winter and summer. Dirmeyer et al. (Dirmeyer, P. A. B. A. Cash, J. L. Kinter III, T. Jung, L. Marx, M. Satoh, C. Stan, H. Tomita, P. Towers, N. Wedi, D. Achuthavarier, J. M. Adams, E. L. Altshuler, B. Huang, E. K. Jin, and J. Manganello, 2011: Simulating the diurnal cycle of rainfall in global climate models: Resolution versus parameterization. Climate Dyn. DOI 10.1007/s00382-011-1127-9) found	Noted – blocking and diurnal cycle are discussed in 9.5.1. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						that higher resolution improves the representation of the diurnal cycle of precipitation, but cannot overcome the limitations of the convective parameterization. Manganello et al. (Manganello, J. V., K. I. Hodges, J. L. Kinter III, B. A. Cash, L. Marx, T. Jung, D. Achuthavarier, J. M. Adams, E. L. Altshuler, B. Huang, E. K. Jin, C. Stan, P. Towers and N. Wedi, 2012: Tropical Cyclone Climatology in a 10-km Global Atmospheric GCM: Toward Weather-Resolving Climate Modeling. J. Climate doi: http://dx.doi.org/10.1175/JCLI-D-11-00346.1) showed that increasing resolution improves the statistics and structure of tropical cyclones, even down to a grid spacing of 10 km, despite the limitations of hydrostatic dynamics and parameterized convection. [James Kinter, United States of America]	
9-380	9	15	9			An important caveat to this comes from the following which discussed the how one initailized experiment from a very active MJO state doesn't represent a robust test of a model's ability to represent the MJO: Sperber, K. R., J. M. Slingo, D. E. Waliser, and P. M. Inness (2008), Coarse-resolution models only partly cloudy (Comment on “A Madden-Julian oscillation event simulated by a global cloud-resolving model” by Miura et al.), Science (Washington), 320, 612. [Duane Waliser, USA]	Noted - This part is largely removed in SOD to reduce overlap with Chapter 7.
9-381	9	15	14	15	25	Add comment on statistical downscaling for completeness, as this technique remains in extensive use. [Robert Colman, Australia]	Taken into account. We mentioned SD 9.1.2, but as for RCMs, do the substantive discussion in 9.6.
9-382	9	15	14	15	25	Please mention that some regional climate models are nested from the global to regional or urban scale so have consistent boundary conditions (Jacobson, M. Z., GATOR-GCMM: A global through urban scale air pollution and weather forecast model. 1. Model design and treatment of subgrid soil, vegetation, roads, rooftops, water, sea ice, and snow., J. Geophys. Res., 106, 5385-5402, 2001; Jacobson, M.Z., Y.J. Kaufmann, Y. Rudich, Examining feedbacks of aerosols to urban climate with a model that treats 3-D clouds with aerosol inclusions, J. Geophys. Res., 112, D24205, doi:10.1029/2007JD008922, 2007; Jacobson, M.Z., The short-term effects of agriculture on air pollution and climate in California, J. Geophys. Res., 113, D23101, doi:10.1029/2008JD010689, 2008) [Mark Z. Jacobson, U.S.A.]	Noted. Many of these applications relate more to impacts and thus are at the remit of WGII. The overarching discussion of issues related to boundary conditions was in section 9.6.
9-383	9	15	14	15	25	Note that chapter 3 of the IPCC SREX report (2012) has a relatively large section on downscaling (Section 3.2.3) [Sonia Seneviratne, Switzerland]	Taken into account. We mentioned SREX when discussing downscaling in Chapter 9.6. The SREX section on downscaling is much on projections and thus relevant not least to other WGI and WGII chapters.
9-384	9	15	14		25	The section downscaling methods must include at least a hint towards empircal-statistical downscaling methods Like: Two strategies are currently used to downscale global climate model results to local scale. Regional climate models ... Empirical-statistical downscaling methods.... If you need some one to write the statistical part you can find my e-mail on the previous tab [Frank Kreienkamp, Germany]	Taken into account. We mentioned empirical and statistical downscaling in section 9.1.2. The more substantive discussion is in Section 9.6.
9-385	9	15	14			section 9.1.3.4.2: line 22: "Biases in the former inevitably lead to biases in the latter." There is no consensus about this, neither about the following line. There seems to be two schools of thought and the majority view is presumibly the one expressed in this text. For a different point of view see Veljovic et al (2010), briefly discussed in section 9.6.3.1. [Ramon de Elia, Canada]	Accepted. We moved the discussion into Section 9.6 where it can be addressed in more substantive terms.
9-386	9	15	14			Mention high-resolution time slice runs with global climate models somewhere in this section, since resolutions in some case approach those of the regional models and provide complementary information. [J. David Neelin, United States]	Taken into account. We mentioned AGCMs and variable-resolution GCMs in section 9.1.2.
9-387	9	15	15	15	19	There are also now a limited number of studies running regional climate models at convection-permitting scales. For example Kendon et al (2012) carried out a 20-year length simulation with a 1.5km RCM over a region of the UK; and Wakazuki et al (2008) carried out 10-year time-slice experiments with a 5km RCM over Japan. [See comment 5 for references] [Elizabeth Kendon, United Kingdom of Great Britain & Northern Ireland]	Taken into account. We expanded on this in the discussion of model resolution.
9-388	9	15	15	15	25	Section 9.1.3.4.2 Downscaling methods. The time-slice experiments, basically global downscaling experiments, should be discussed here along with regional downscaling.The section righty points the common issues of global and regional climate models. However it should also mention that 1) compared with global	Taken into account. Regional-scale information from AOGCMs, AGCMs, variable-resolution GCMs, SD and RCMs was expanded in Section 9.6.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						climate models, regional climate models add an artificial boundary in the simulation, an additional source of uncertanty 2) conventional regional climate models provide no feedback on the global scales, unlike the global high resolution models 3) the value added by regional climate models when these are used in projection/prediction mode is very different from when "perfect BC" are applied. Overall, given the high interest in regional-scale information, this section should be improved. [Annarita Mariotti, U.S.A.]	
9-389	9	15	15		25	RCMs have different parameterizations from GCMs, such as deep convection that does not need to consider mesoscale effects. The limitations in this type of model can be very different. Higher resolution does not guarantee better results. The writing in this paragraph implicitly assume that this is the case. [Kuan-Man Xu, USA]	Noted. The writing does not take stand on whether and to what extent higher resolution provides better results. When it comes to regional scale information, added value and skill are discussed in 9.6, which will be reorganised and expanded in SOD to add clarity.
9-390	9	15	16	15	19	That these models employ a higher resolution is due to the fact that a smaller grid is used and thus less computer power is needed. [Farahnaz Khosrawi, Sweden]	Noted.
9-391	9	15	24			The sentence that begin with "Rather…" is not possible to understand [Gunilla Svensson, Sweden]	Noted. The sentence continued the line of reasoning in the previous sentence and expresses that regional climate modelling targets processes and features that are characterised by higher resolution than representatively dealt with in AOGCMs. The sentence was, however, removed in the revision of the text.
9-392	9	15	27	15	57	Model evaluation is also carried out at modelling centres. The resulting literature is extensive (and much of it already referred to, as new representations of physical processes are not introduced without evaluation). The IS-ENES project has provided a list of observational datasets used in evaluation of ESMs. A compilation of the datasets used by a range of modelling centres to evaluate their models is available here: https://www.enes.org/models/evaluation-portal – I'm not sure how you can review this material, but it is important to refer to the process level validation work which has been done by the modelling centres. The references given in this section are an extremely minor component of the overall effort to validate models – though perhaps representative of the model intercomparison component of model validation work. [Martin Juckes, UK]	Noted. We have added a table that provides an overview of the observations used in this chapter. We have also added to the text that evaluation work is done at the modelling centers. However, the indivudal papers that are important in the context of chapter 9 are cited throughout the chapter, so have not been added here. The references cited list the those studies that have explored performance metrics in the mutli-model context.
9-393	9	15	31	15	37	I agree completely with these concelts. However, the climate models are mainly used for their climate projections, thus the ultimate but fundamental goal of the model evaluation is to assess their credibility. So I am suggesting to puctualize better this fact by anticipating the last sentence. [Claudio Cassardo, Italy]	Accepted. Text revised.
9-394	9	15	31	15	37	These are mere opinions. There is no intention to find out whether they are capable of forecasting the climate [VINCENT GRAY, NEW ZEALAND]	Rejected - The assessment provided here is firmly rooted in the published climate-science literature.
9-395	9	15	32	15	32	I would suggest to write "physical and biogeochemical processes as well as feedbacks". [Farahnaz Khosrawi, Sweden]	Accepted. Text revised.
9-396	9	15	32	15	34	It should be stated that the spread of models is at most a lower bound of the uncertainty we face, as the report says on page 9-20, line 1. Also, the problems of uncertainty quantification should be mendtioned or refered to. Alternatively, you might consider deleting the sentence altogether. Or, simply report: "... have been used as a means of ...." rather than "can serve as a means ...". [Gregor Betz, Germany]	Accepted. Sentence deleted since this is discussed elsewhere in the chapter.
9-397	9	15	32	15	34	Inter-model spread is not and can not be a measure of true uncertainty, regardless as to whether observational measures are available. One can at best say that some measure (not, in all likelihood, the range) of the distribution of model results can provide a lower bound on the true uncertainty. [Robert Pincus, USA]	Accepted. This is discussed in another place of the chapter; sentence deleted here.
9-398	9	15	33			the spread of models alone without observational tests is a VERY LIMITED test of model uncertainty. need to mention this. [Bruce Wielicki, USA]	Accepted. This is discussed in another place of the chapter; sentence deleted here.
9-399	9	15	34	15	36	Another major objective of model evaluation is to inform future model development (e.g. Jakob, 2010, BAMS) [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Accepted. Text revised to reflect this comment.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-400	9	15	36	15	37	This point is expressed more fully and more accurately in section 9.8.3, but this sentence is misleading. The link between a model's ability to simulate the historical record and its ability to make future projections is tenuous (see, for example, doi:10.1029/2009GL038082). Perhaps the safest course of action is to remove this sentence. [Robert Pincus, USA]	Accepted. Since it is certainly an objective of model evalaution we kept but reformulated the sentence and also included a link to Section 9.8 where this is discussed in more detail.
9-401	9	15	39	15	39	"forcing used in most" [George Kiladis, USA]	Noted.
9-402	9	15	44			After Waugh and Eyring 2008 insert "Sahany et al. 2012" [J. David Neelin, United States]	Accepted. Reference added.
9-403	9	15	46	15	49	Concerning weighting of projections, a more careful formulation is required, e.g. "have been used" rather than "can be used". Note that the IPCC Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections is very critical concerning such weighting (I quote: " A robust approach to assigning weights to individual model projections of climate change has yet to be identified. Extensive research is needed to develop justifiable methods for constructing indices that can be used for weighting model projections for a particular purpose." and: "It is problematic to regard the behavior of a weighted model ensemble as a probability density function (PDF). The range spanned by the models, the sampling within that range and the statistical interpretation of the ensemble need to be considered."). Therefore, it should be stated clearly that weighting is problematic. [Gregor Betz, Germany]	Taken into account by adding a link to 9.8.3 where this is discussed in more detail.
9-404	9	15	47			On first reading, it seems that references are needed here. Certainly a link to the later section 9.8.3 would help. [Ian Watterson, Australia]	Taken into account. A link to Section 9.8.3 added
9-405	9	15	48			As discussed below, the term "process-oriented evaluation" is inaccurate. [Robert Pincus, USA]	Taken into account. Reformulated.
9-406	9	15	49			After "Knutti et al. 2010b" insert ", Neelin et al. 2010" [J. David Neelin, United States]	Accepted. Reference added.
9-407	9	15	50	15	50	page 9.15, § 9.2.1, line 50 In fact, the so called, the observations are made by instrument and calculated by the physical representation of the interaction of these instruments with the process (i.e. speed of a current, or of o a wind at such or such small region and duration of time) or state variable (Temperature, or salinity or humidity at such or such small region and finite duration of time) one want to measure. So, what we call observation are functional , linear or not, on the variables and other one of the phenomena. So, we have in fact five levels of the phenomena we want to describe by a so called "model". -1: The true physical-chemical system we want to describe., with the concepts of physics and chemistry which evolutes in time and space. It is in fact a collection of concept, depending on the level of description we want (i.e., statistical physics or thermodynamics) and the category of questions at which we want to reply -2: The observed system, which is a collection of numbers, with margins of uncertainties, coming from the functional on the results of the instrument. We have to point that in order to solve this functional, we have to accept a first description of the system and its processes. Statistical treatment of observed data. Assimilation methods of the data. etc. -3: The mathematical models of -1: including the stationery stochastic treatment of certain uncertainties and stochastic processes in time to take care of irregularity by the mathematical theory of measure. 4: the numerical methods representations and algorithms of computation of the models -3: 5: Its translation in a model with physical numbers and using -2: [Robert DAUTRAY, France]	Noted.
9-408	9	15	51	15	52	Please provide relevant references (after the sentence ending with "reasons"). [Martin Dameris, Germany]	Accepted. References added.
9-409	9	15	51	15	53	It may not be obvious to readers why climate model evaluation is so different from the evaluation of weather forecasting models. It would be useful to be explicit here: forecasting models make specific predictions for specific times, and so can be tested against any available observations, while climate models are only expected to reproduce climatological distributions, and so require long records from both model and observations to reduce sampling noise. [Robert Pincus, USA]	Taken into account. Text revised to clarify this point.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-410	9	15	51	16	7	An elaborate excuse why you cannot use the models in practice. The "long term" claim is wishful thinking. You are not even able to do as well as routine weather forecasts [VINCENT GRAY, NEW ZEALAND]	Rejected. The text here describes the methods and data requirements for evaluating climate simulations, and these differ fundamentally from the requirements for routine weather forecasts.
9-411	9	15	51			I don't know if you will find it useful, but I made an evaluation of climate model skill (in the sense understood in NWP), which I think is quite a neat demonstration of two things - that climate models have predictive skill, and that we cannot calculate what it is for the current flock of models. See J.C. Hargreaves, Skill and uncertainty in climate models, 2010, WIREs, 1(4), 1757-7799, OI: 10.1002/wcc.58. [Julia Hargreaves, Japan]	Taken into account. Reference to the paper added.
9-412	9	15	53			Its not just a longer range of time scales that differentiate ESMs from weather models its also the broader spectrum of processes that interplay across space and time [Graeme Stephens, USA]	Taken into account. Text revised to clarify this point.
9-413	9	15	57			Suggest reference to "obs4MIPs" here: Gleckler, P. R. Ferraro, D. E. Waliser, 2011: Better use of satellite data in evaluating climate models contributing to CMIP and assessed by IPCC, Meeting Summary, EOS, Vol. 92, No. 20, 17 May 2011. [Duane Waliser, USA]	Accepted. Reference added.
9-414	9	16	3	16	3	I don't agree with this statement. Usually errors are given for observational data sets. [Farahnaz Khosrawi, Sweden]	Taken into account. Sentence deleted.
9-415	9	16	3	16	3	What do you mean with multiple estimates? [Farahnaz Khosrawi, Sweden]	Taken into account. Sentence deleted.
9-416	9	16	4	16	4	"based on the same underlying measurements": among different datasets, the majority of the data will be exactly the same, while some other data may have been chosen by a particular dataset and not by another one, for any reason. I suggest to add mostly: based mostly on the same underlying measurements [Claudio Cassardo, Italy]	Taken into account. Sentence deleted.
9-417	9	16	9	16	9	This is a good place to have a sub-section to summarize the use of multiple satellite observations, especially A-Train, in model evaluation approaches. Many A-Train instruments, e.g. CloudSat and CALIPSO, were not available during the AR4 period, and thus should be considered new in AR5 model evaluations (e.g. Jiang et al. 2012). The most straightforward approach to evaluate models is to compare overall simulated fields (e.g., global distributions of temperature, precipitation etc.) with corresponding observations…. It should be included that new approaches using collocated satellite datasets such as examining relations between difference parameters and cloud profiles in large scale regimes (e.g. Su, H., J.H. Jiang, D.G. Vane, and G.L. Stephens, "Observed Vertical Structure of Tropical Oceanic Clouds Sorted in Large-scale Regimes ," Geophys. Res. Lett. 35, doi:10.1029/2008GL035888, 2008; Su et al 2011, "Comparison of Regime-Sorted Tropical Cloud Profiles Observed by CloudSat with GEOS5 Analyses and Two General Circulation Model Simulations," J. Geophys. Res. 116, D0910, doi:10.1029/2010JD014971) are being used since AR4. [Hui Su, United States of America]	Taken into account - The general use of cloud observations is described in Chapter 7 (Section 7.2.2.2) as well as in the Instrument Simulator section in Chapter 9 (Section 9.2.2.3). Regime-oriented approaches are discussed in Section 9.2.2.2. We have included the assessment of the papers suggested by the reviewer in those sections.
9-418	9	16	11	16		A new set of metric has been reported recently, which contains physically related components. This comprehensive metric can objectively assess the performance of GCMs in reproducing a climate phenomenon of interest. "A comprehensive metric is proposed to objectively assess the performance of Beijing Climate Center Atmospheric General Circulation Model (BCC_AGCM) in simulating the interdecadal changes of East Asian climate. The metric includes the consideration of both rainfall and its associated large-scale environments, which exhibit strong coherent interdecadal changes in observations. The physically related multi-components ensure the evaluation to be comprehensive. (Chen et al., 2011, The coherent interdecadal changes of East Asia climate in mid-summer simulated by BCC_AGCM 2.0.1. Climate Dynamics)" [Rucong Yu, China]	Noted – however, what is called 'metric' in this paper does not really fit in the context of this section. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-419	9	16	13	16	14	The most straightforward approach to evaluate models is to compare... global temperatures... with observations. In this I concur. The "high estimate" of the FAR report has exceeded global temperature data and exceeded SST data even more, and this every year for past 22 years except the year of the exceptional 1998 El Niño peak. 22 years is a sufficient period of time to raise questions about the inputs and ingredients entering the GCM models. This means that the climate sensitivity retained in IPCC reports has to be reduced, at least its maximum value. This is also why the increasing divergence of the model predictions with	Noted - the consistency between observed and projected temperature in recent decades is discussed in Chapter 10 and, as a historical review, in Chapter 1. Also, note that there is evidence in the published literature that recent hiatus of temperature rise is consistent with our basic understanding of climate

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						temperature indicators (global, SST, ocean heat content, deceleration of increase of sea level) which show a plateau during the past several years, and some data even up to 13 years, raises a fundamental and embarrassing problem to the models. [François GERVAIS, France]	change (Easterling and Wehner, 2009, GRL; Meehl et al., 2011, Nature Climate Change).
9-420	9	16	13	16	14	Cont. - Climate modelers should face the problem and cease ignoring the more and more diverging recent observations. Schmittner et al, Science 334 (2011) 1385, arrive to the same conclusion of decreasing the climate sensitivity from paleoclimatic data. They reevaluate it to a value between 1.7 and 2.6°C for CO2 doubling. The new forcing suggested is even not considered in the models. Another paper which was not available when the AR5 draft was written deserves inspection : N. Scafetta, J. Atm. & Solar-Terrestrial Phys. doi.org/10.1016/j.jastp.2011.12.005. It is based upon an analysis of the decadal variability of the temperature since 1850. It fits the data since 2000 much better than models with exagerated radiative forcing of CO2 (in the sense of my comment about Chapter 8 Page 97 Fig. 8.16), retained in the AR5 draft. Of course, the Scafetta model is empirical but the "model tuning" procedures admitted throughout the Chapter 9 also implies a dose of empiricism. [François GERVAIS, France]	Noted - the suggested paper will be discussed in Chapter 10.
9-421	9	16	13	16	14	Cont. - To calculate a more realistic CO2 anthropogenic component of climate change by comparison with experimental data, I suggest running the models with an anthropogenic CO2 radiative forcing limited to 0.0025 W/m2 per year. This suggestion is based upon two simple but sound arguments which do not seem to have been considered in the AR5 draft. 0.0025 W/m2 is evaluated by retaining ~ 0.01 W/m2 per year corresponding to the anthropogenic fraction measured in Fig. 8.16 nowadays, viz. similar to the lowest value measured in 1992 (see comment above about Chapter 2 Page 44), divided by 4 to take into account the presumably exagerated radiative transfer component retained in the AR5 models. [François GERVAIS, France]	Rejected - the argument is not based on scientific literature. Also, note that an IPCC assessment is an assessment of existing literature and not a means of undertaking new experiments.
9-422	9	16	13	16	14	Cont. - The factor 1/4 comes from the ratio of the thermal emissivity of the earth measured by satellite at the relevant wavelength of the vibrational CO2 bending mode (15 micrometers) related to the blackbody profile at the same temperature and frequency, as explained in the introduction. This factor accommodates the fact that ~ 3/4 of the remaining additional molecules of CO2 emitted by anthropogenic activities cannot "see" the earth radiation which is already stopped (fraction not reemitted, heat dissipated by thermal conduction and convection) at the relevant frequencies by molecules lying below, and cannot, therefore, heat and contribute to the greenhouse effect. [François GERVAIS, France]	Rejected - the argument is not based on scientific literature. Also, note that an IPCC assessment is an assessment of existing literature and not a means of undertaking new experiments.
9-423	9	16	13	16	14	The IPCC GCMs do not reproduce the global surface temperature decadal and multidecadal cycles, see Scafetta (2011) for a very detailed check. See Scafetta, N., Testing an astronomically based decadal-scale empirical harmonic climate model versus the IPCC (2007) general.... Journal of Atmospheric and Solar-Terrestrial Physics (2011), doi:10.1016/j.jastp.2011.12.005 [James Wanliss, USA]	Noted - the suggested paper will be discussed in Chapter 10.
9-424	9	16	13	16	15	It would be worth being explicit here: comparisons are normally made between time-mean fields, perhaps resolved into a climatological seasonal cycle. [Robert Pincus, USA]	Noted - it is stated when specific application is described below including 9.4.1.2. Here it seems that a more general description is enough.
9-425	9	16	13	16	17	I think Murphy et al 2004 (and maybe some of the other references cited on 9-70 lines 51-53), which was the first paper to use one number based on a basket of 32 metrics to evaluate models in an application (see 9-72 l52 as well), should be added to or replace some of the references in this list. [David Sexton, UK]	Taken into account - a mention on aggregated metric is added.
9-426	9	16	16			After "correlations" insert "or conditional averages or regressions of an evaluation variable on other variables known to affect it, such as for ENSO teleconnections" [J. David Neelin, United States]	Taken into account - we understand that this comment essentially requests a mention on the evaluation of variability modes. A text is added.
9-427	9	16	17			Add to list of citations some references relevant to these methods, e.g. "Wallace et al. 1998, Neelin 2008" [J. David Neelin, United States]	Taken into account - we understand that this comment essentially requests a mention on the evaluation of variability modes. Citations, though different from suggested ones, are added.
9-428	9	16	23	16	31	"If each model is a random and independent sample from a distribution of possible models centred on the observations": in my opinion, in general, two elements of the statistical sample of the model results are not completely independent from each other, as models can share some basic physics and initial conditions. This may be one of the reasons for which the errors do not converge zero. This fact is also underlined at page 21,	Noted - it is mentioned in 9.2.2.7.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						lines 19-25. [Claudio Cassardo, Italy]	
9-429	9	16	23		31	It would seem appropriate to actually mention the rank histogram approach introduced (in this context) by Annan and Hargreaves GRL 2010, as this is very much intended as a technique to analyse ensembles as a whole. Yokohata et al 2011 (Climate dynamics) extends this idea substantially (many more variables) and is surely worth discussing somewhere in the context of the analysis of CMIP3 models. [James Annan, Japan]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-430	9	16	23		31	What conclusions do the authors draw from this pair of findings? [Larry Thomason, United States of America]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-431	9	16	24	16	25	This is a highly problematic assumption; and that should be made clear. In addition to the statistical evidence against it, the assumption is conceptually flawed, because it is not clear at all what the set of all possible models is in the first place. [Gregor Betz, Germany]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-432	9	16	26	16	31	The last sentence begins with By contrast. I don't see that this statement contrasts the earlier statement. [Gunilla Svensson, Sweden]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-433	9	16	27	16	27	Slower?? Please clarify and rephrase what you mean with slower. [Farahnaz Khosrawi, Sweden]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-434	9	16	28	16	31	Sentence is opaque (to me at least!) [Robert Colman, Australia]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-435	9	16	28	16	31	I think the results of Annan and Hargreaves (2010) were consistent with the hypothesis that the truth and CMIP3 simulations were drawn from the same distribution - this was something they demonstrated based on analysis of model output, not an assumption of their analysis as implied here. [Nathan Gillett, Canada]	Noted - the part is deleted and related discussion is left to 9.8.3.
9-436	9	16	31	16	31	Another statistical tool used to verify the model simulations or hindcasts/predictions with multi-ensemble member is an empirical orthogonal function analysis with a maximized signal-to-noise ratio. With this technique, the most predictable or reliable patterns are isolated. We have applied this method to isolate the most predictable patterns of NCEP Climate Prediction System (CFS) in the tropical Atlantic (Hu and Huang 2007), and in East Asia summer monsoon (Liang et al. 2009; Gao et al. 2011). We belive this method may be also very useful to pick up the signals from multi-model and multi-ensemble member runs of CMIP5 experiments. [Zeng-Zhen HU, USA]	Noted – however, the papers are on predictability of seasonal variation and do not really fit in the context of this section. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-437	9	16	31	16	31	Gao, H., S. Yang, A. Kumar, Z.-Z. Hu, B. Huang, Y. Li, and B. Jha, 2011: Variations of the East Asian Mei-yu and simulations and prediction by the NCEP Climate Forecast System. J. Climate, 24 (1), 94-108, DOI: 10.1175/2010JCLI3540. [Zeng-Zhen HU, USA]	Noted – however, the paper is on predictability of seasonal variation and does not really fit in the context of this section. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-438	9	16	31	16	31	Liang, J., S. Yang, Z.-Z. Hu, B. Huang, A. Kumar, and Z. Zhang, 2009: Predictable patterns of the Asian and Indo-Pacific summer precipitation in NCEP CFS. Clim. Dyn., 32: 989-1001, DOI: 10.1007/s00382-008-0420-8. [Zeng-Zhen HU, USA]	Noted – however, the paper is on predictability of seasonal variation and does not really fit in the context of this section. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-439	9	16	31	16	31	Hu, Z.-Z. and B. Huang, 2007: The predictive skill and the most predictable pattern in the tropical Atlantic: The effect of ENSO. Mon. Wea. Rev., 135 (5), 1786-1806. [Zeng-Zhen HU, USA]	Noted – however, the paper is on predictability of seasonal variation and does not really fit in the context of this section. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-440	9	16	33			Section 9.2.2.2: “Isolating Processes” in “New Developments in Model Evaluation Approaches” could mention the link between the water vapour feedback on climate change & control climate established by Ingram (2010), if only for the theoretical advantage of such a link, as no practical implementation has been achieved. Text could be like “One direct link between an important component of climate change and control climate has been established: most of the water vapour feedback on climate change is attributable to a component which	Taken into account - Water vapour feedbacks, including the issues commented upon here are discussed in Chapter 7.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						depends only on control climate (Ingram, 2010). This component is not directly observable, however, and does not seem to account for a large fraction of the uncertainty in climate sensitivity (Ingram, 2012)." The references are Ingram, W., 2010: A very simple model for the water vapour feedback on climate change. Q. J. R. Meteorol. Soc., 136, 30-40. available at http://onlinelibrary.wiley.com/doi/10.1002/qj.546/abstract and W. J. Ingram, 2012: A new way of quantifying GCM water vapour feedback. Clim. Dyn. 2012, 10.1007/s00382-012-1294-3 (in press) - proofs available for IPCC review purposes as pub/user/ingram/pS/Ingram2012newway.pdf by anonymous ftp into ftp.atm.ox.ac.uk. (This comment also applies to 9.7.4.3.1 & is repeated there.) [William Ingram, UK]	
9-441	9	16	33			Section 9.2.2.2: You might want to include discussion of the use of model families (such as HadGEM2) in which there are several available configurations with varying levels of complexity. These offer the chance to isolate the impact of particular components, e.g. the middle atmosphere, chemistry, carbon cycle, on the model simulation, as well as to provide some quantification of the uncertainty that might arise from different levels of complexity between models in the MME. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account - A sentence noting the use of model families and modelling techniques was added.
9-442	9	16	35	16	35	? Processes parametrisations [Robert Colman, Australia]	Editorial
9-443	9	16	35			climate models rely heavily on process parameterizations(deleted "es" after process) [Jiemjai kreasuwun, Thailand]	Editorial
9-444	9	16	41			"distinct" is more accurate than "important" here. [Robert Pincus, USA]	Accepted
9-445	9	16	41			The utility of sorting by circulation regimes dates at least to doi:10.1175/1520-0442(2001)014<2540:ITFEGC>2.0.CO;2. [Robert Pincus, USA]	Rejected - As the focus of this section is on assessing the literature since AR4 we did not include a citation to this paper.
9-446	9	16	44			After "Williams and Brooks 2008)" add "or conditional averages of precipitation onset statistics as a function of water vapor and temperature (Neelin et al. 2009, Sahany et al. 2012)" [J. David Neelin, United States]	Noted - these papers have been assessed and the Sahany et al. reference has been added.
9-447	9	16	51	16	54	VOCALS-ReX was a major international modeling campaign to improve our understanding/modeling since the AR4. A reference to this experiment should be added. [Annarita Mariotti, U.S.A.]	Taken into account - A reference to Wood, R. and Coauthors, 2011: The VAMOS Ocean-Cloud-Atmosphere-Land Study Regional Experiment (VOCALS-REx): goals, platforms, and field operations. ACP, 11, 627–654. has been added.
9-448	9	16				Lines 17-19 and lines 19-21 are both accurate but are misplaced - they don't follow from a description of how fields are compared with one another. [Robert Pincus, USA]	Noted - the former part is deleted and left to 9.8.3, while the latter part is kept here as we think it fits in the context.
9-449	9	17	1	17	26	In additon to satellite simulator, which focuses mostly on clouds, recent studies have used standard satellite data to evaluate model simulation for more than one parameter (e.g. clouds and water vapor) and examine relationship between different parameters and compare with the observed relationship. These should be valuable to the AR5. [Hui Su, United States of America]	Taken into account - The cloud evaluation section in Section 9.4.1 has been extended and now assesses those studies. We also added references to recent studies not using simulators in the first part of this section.
9-450	9	17	1	17	27	In additon to satellite simulator, which focuses mostly on clouds, recent studies have used standard satellite data to evaluate model simulation for more than one parameter (e.g. clouds and water vapor) and examine relationship between different parameters and compare with the observed relationship. [Jonathan Jiang, United States of America]	Taken into account - The cloud evaluation section in Section 9.4.1 has been extended and now assesses those studies. We also added references to recent studies not using simulators in the first part of this section.
9-451	9	17	1			As noted above, the term "instrument simulators" reflect the scientific community's language but are inaccurate. Note, too, that this technique is most frequently applied to observations of clouds - the people computing temperature trends observable from microwave sensors, for example, are content to apply vertical weighting functions to monthly-mean data. [Robert Pincus, USA]	Taken into account - We acknowledge the issue raised by the reviewer. However, the term is now in such wide use, that we adopt it here as well. We have changed the text to better explain the use of the term as generating "observation equivalent" from models,

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							rather than strictly simulating satellite measurements.
9-452	9	17	3	17	3	"meteorological conditions"? They sample rather meteorological parameters, but also trace gases and aerosols. [Farahnaz Khosrawi, Sweden]	Taken into account - The text has been adjusted to make it clearer that we refer to sampling across many conditions.
9-453	9	17	3			Section: 9.2.2.3: This is a quite complicated description of the retrieval process. It is true that there are uncertainties in the parameters or species derived from satellites that should be considered and mentioned in this report. However, as the section is written now it is quite difficult to follow and is rather confusing than helpful. [Farahnaz Khosrawi, Sweden]	Taken into account - The section has been reworded to simplify it.
9-454	9	17	6		6	It seems that "reality" is a better word for "performance" here. [Kuan-Man Xu, USA]	Taken into account. - The sentence has been rephrased.
9-455	9	17	9		11	Those limitations are not always true for both the passive and active sensors. Please expand this sentence to include different limitations for these two sets of sensors. [Kuan-Man Xu, USA]	Taken into account - we now note that the limitation vary across satellites and instruments. As this report is not a science review, we did not include a lengthy discussion of all possible issues one could encounter and rather provide examples to help the understanding.
9-456	9	17	13	17	26	Some caveat should be given that for some quantities there needs to be caution exercised in terms of utilization of simlators. For example cloudsat reflectively is sensitive to cloud and falling hdyrometeors (e.g. rain and snow) but most CMIP models do not simulate profiles of these quantities. Thus when using a simulator this part is either ignored or significant assumptions have to be made about how to account for these missing components in order to compare the model/simulator results to observations. [Duane Waliser, USA]	Taken into account - We have added a sentence and references to papers discussing the limitations of simulators.
9-457	9	17	13		26	Calling out some specific instruments for which simulators exist would be helpful. [Larry Thomason, United States of America]	Rejected - The list of such simulators is very long and growing. Including a list here would become very technical and not in the spirit of an IPCC assessment. We trust that the more technical readers can make use of the references provided to acquire more details.
9-458	9	17	24			Readers may not know what the "ISCCP simulator" is; indeed, they may not know what ISCCP is. [Robert Pincus, USA]	Taken into account - The sentence has been rephrased.
9-459	9	17	24			the text here is needlessly detailed. [Robert Pincus, USA]	Accepted - The sentence has been simplified.
9-460	9	17	28	17	53	This section breaks the flow and would prefer to have it later in the section. [Gunilla Svensson, Sweden]	Rejected. The section is on the same line as the previous section and before the regional and multi model approaches that can be applied to any climate.
9-461	9	17	40		42	This sentence is not clear to me. [Larry Thomason, United States of America]	taken into accont. The section has been rewritten
9-462	9	17	43	17	43	change "also to be accounted" with "be taken into account" [ANNALISA CHERCHI, Italy]	Accepted
9-463	9	17	45			It would be useful to point out that the forward modeling of proxy measurements here is essentially the same as the "instrument simulator" strategy discussed in section 9.2.2.3. [Robert Pincus, USA]	Accepted. This is a good point .
9-464	9	17	52	17	52	TYPO: "or" instead of "of" [Ladislav Metelka, Czech Republic]	Rejected. Unclear
9-465	9	17	55	18	24	9.2.2.5 Use of Data Assimilation and Initial Value Techniques. This section should also mention that re-analyses, from assimilation of consistent assimilation of observations on a stationary model platform, represent an important source of observational information for model evaluation. [Annarita Mariotti, U.S.A.]	Rejected - This section is specifically about using intial condition techniques to assess model error. Reanalyses are discussed in Chapter 2.
9-466	9	18	1			Somewhere in Section 9.2.2.5 the recently suggested approach by Eden et al (Eden, Widmann, Grawe, Rast, 2012, J Climate, doi: 10.1175/JCLI-D-11-00254.1) should be mentioned. They nudge the ECHAM5 AGCM to	Taken into account - A brief discussion of the nudging apporach has been added to this section. The paper

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						reanalysis data and observed SST and provide an upper limit for the precipitation error in the model that stems from precipitation parameterisations. [Douglas Maraun, Germany]	has been assessed and has been included.
9-467	9	18	9	18	10	This sentence does not follow from the previous paragraph. [Robert Pincus, USA]	Taken into account - The sentence has been rephrased.
9-468	9	18	10	18	10	Include "for model performance" after advantages. [Farahnaz Khosrawi, Sweden]	Editorial
9-469	9	18	12	18	17	It should be noted that this approach requires the model to have its own data assimilation system. The majority of climate models do not have DA systems and they are very expensive to develop. [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account - A sentence to make this point has been added.
9-470	9	18	14	18	15	The Rodwell and Palmer paper shows that data assimilation can identify parameter values that make the forecasting system less consistent with observations. This does not mean the values can be excluded, especially since the parameter value most consistent with observations might be the one which best compensates for other errors. The fact that the best value of the particular parameter used in this paper led to lower climate sensitivity is not relevant. [Robert Pincus, USA]	Taken into accout - The sentence has been rephrased to describe the study as an attempt at excluding parameter settings rather than prrof that it can be done.
9-471	9	18	15			What is 'ensemble data assimilation'? Can this be briefly explained? [Nathan Gillett, Canada]	Rejected - The term is in common use.
9-472	9	18	19	18	19	What do you mean with "initial value simulations"? [Farahnaz Khosrawi, Sweden]	Rejected - The term is in common use.
9-473	9	18	19	18	24	there are also results emphasizing the positive contribution of ocean data assimilation to seasonal prediction (Alessandri et al., 2010). The reference is "Alessandri A, Borrelli A, Masina S, Cherchi A, Gualdi S, Navarra A, Di Pietro P, Carril AF (2010) The INGV-CMCC Seasonal Prediction System: improved ocean initial conditions. Monthly Weather Review, 138(7), 2930-2952 DOI: 10.1175/2010MWR3178.1" [ANNALISA CHERCHI, Italy]	Rejected - This section does not discuss the success or failure of predictions, but the use of initial value techniques in identifying model errors.The proposed reference is not addressing this issue.
9-474	9	18	26	18	47	Add also use of station (or other single locality) data for RCM evaluation. [Robert Colman, Australia]	Noted. The direct use of station data is not a major RCM evaluation technique, but certainly used in some cases. We assess it to fall under the overall use of observational data in model evaluation.
9-475	9	18	26	18	47	"A difficulty in the evaluation of RCMs….is the relative sparseness of observational networks." This is not really correct. There are many high-resolution datasets and field campaigns that RCMs have been compared with (e.g., Jacobson, M. Z., GATOR-GCMM: 2. A study of day- and nighttime ozone layers aloft, ozone in national parks, and weather during the SARMAP Field Campaign, J. Geophys. Res., 106, 5403-5420, 2001; Jacobson, M.Z., Y.J. Kaufmann, Y. Rudich, Examining feedbacks of aerosols to urban climate with a model that treats 3-D clouds with aerosol inclusions, J. Geophys. Res., 112, D24205, doi:10.1029/2007JD008922, 2007; Jacobson, M.Z., The short-term effects of agriculture on air pollution and climate in California, J. Geophys. Res., 113, D23101, doi:10.1029/2008JD010689, 2008). [Mark Z. Jacobson, U.S.A.]	Noted. There is relative sparseness of observational networks when considering the many different regions and longer time periods as a whole, for which RCMs are nowadays used. For some regions and periods more data is available, which facilitates the evaluation of how RCMs perform in specific locations, times and with respect to specific processes.
9-476	9	18	26	18	47	I don't think this text is necessary to have here. RCMs are evaluated the same way as other models. I agree that there are some special techniques that are special but I think it would be better to describe them in the section that is devoted to RCM evaluations. [Gunilla Svensson, Sweden]	Taken into account. The discussion was shortened but, in addition to noting on the similarities with AOGCM evaluation, we assess that it is useful to introduce already here the dependency on the boundary conditions as well as the possibility to use reanalyses, thus facilitating more direct comparison with observations.
9-477	9	18	26			section 9.2.2.6: For evaluation techniques particular to RCMs there is also the "Big Brother" experiment (see e.g. Laprise et al 2008, already quoted) also mentioned in page 57, line 55. In addition we can mention sensitivity experiments with parameters that do not exist in GCMs, like domain size, nudging technique etc. (e.g. de Elia and Cote (2010) Climate and climate change sensitivity to model configuration in the Canadian RCM over North America. Meteorologische Zeitschrift 19:325-339) [Ramon de Elia, Canada]	Taken into account. We do a substantive discussionon these methods in section 9.6.
9-478	9	18	32	18	32	"a RCM" instead of "an RCM" and "result" instead of "results". [Farahnaz Khosrawi, Sweden]	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-479	9	18	35	18	37	A cautionary note should be added regarding the fact that regional climate model evaluation in the perfect model setting should not translate into expectations that similar fidelity will be obtained in projections, which are heavilty affected by the boundary conditions. [Annarita Mariotti, U.S.A.]	Accepted. We drew attention to the difference between RCM simulations when using boundary conditions from AOGCMs and from reanalyses.
9-480	9	18	37			A reference to the spectral nudging technique should be added, including a discussion; e.g., von Storch, Langenberg, Feser (2000), Mon. Wea. Rev. 128: 3664-3673. [Douglas Maraun, Germany]	Noted. Spectral nudging is not necessarily an evaluation technique as such and was discussed in Section 9.6.
9-481	9	18	37			I am missing a thorough discussion of limitations and problems of validation data sets at a central place of the chapter. Often reanalysis data sets are used or gridded data sets, and all of them have their limitations. E.g., Hofstra, New and My Sweeney (2010), Clim. Dynam., 35: 841-858; Kysely and Plavcova (2010) J. Geophys. Res., D23118; and Maraun, Osborn and Rust (2012), Clim Dynam., DOI 10.1007/s00382-011-1176-0 all find serious problems of the E-OBS data set, which is routinely used for validation of regional climate models over Europe. [Douglas Maraun, Germany]	Taken into account. An assessment of the quality of datasets is at the remit of Ch02.
9-482	9	18	44	18	47	Although high resolution gridded observations might not be available, model output at varying resolutions could still be compared with observations at indvidual weather stations. This would be a particularly relevant test if the model were being used to derive projections for a particular location - say a city. [Nathan Gillett, Canada]	Noted. (Cf. response to comment #9-474.) Indeed, if and when downscaling results are used for point-scale applications, appropriate evaluation should be made. However, as RCMs operate in a gridded sense, it is more appropriate to compare their results to gridded observational products.
9-483	9	18	44	18	47	I do not see why this would be special for RCMs, a global model would be just as difficult to evaluate for these regions. [Gunilla Svensson, Sweden]	Noted. Available literature does highlight these issues in the regional climate modelling context. However, we agree that the basic issue is shared by GCMs and RCMs and omitted specific discussion here.
9-484	9	18	49	19	2	This paragraph needs some more explanation! Difficult to follow. [Martin Dameris, Germany]	Noted. Significant updates were included in this section.
9-485	9	18	49			This section does not fully capture the benefits of the perturbed physics ensemble or the benefits of using information from both perturbed physics ensembles and multimodel ensembles that are out there in the literature and have been for a few years now. Most important omission is that the perturbed physics ensemble underpins Bayesian frameworks such as that outlined by Rougier 2007, which provide a rigorous mathematical framework including the key uncertainties for a given set of boundary conditions: parametric uncertainty, observational uncertainty and structural uncertainty (called “discrepancy” in Rougier 2007). The main point to bring out here is that the presence of structural uncertainty i.e. a non-zero discrepancy variance affects the ability of a metric to discern a good model from a bad model because of model imperfections, thereby weakening the constraint and avoiding over-confident constraints. Most studies assume zero discrepancy and so constraints are over-confident. In this chapter that point has to be discussed and I would suggest that accounting for the effects of structural uncertainty/model imperfections on model evaluation and constraining projections provides a challenge for us all. Its importance has been demonstrated recently. The framework has been applied in Sexton et al (2011 online in Climate Dynamics - see first comment for REF) where multimodel and perturbed physics runs were used to specify a discrepancy and explore parametric uncertainty and constrain climate sensitivity and response of some variables to doubled CO2. Sanderson (in review) also use perturbed physics and multimodel ensembles together to provide more reliable constraints on climate sensitivity and have an alternative approach for inferring the degree of systematic error from the multimodel ensemble. That's two papers and so it is early days as far as applications go, but the impact of structural error on model evaluation is well known and should be presented as so. [David Sexton, UK]	Accepted. This has been updated to include these comments by David Sexton as a CA.
9-486	9	19	4	19	4	Although the emphasis … is the multi-model evaluation…'. Justify taking this approach (here, and indeed in the chapter generally). [Robert Colman, Australia]	Noted. This text was substantially revised and this sentence removed.
9-487	9	19	4	19	11	It would be good to mention that with these ensemble approaches, there is another light touch way to use observations but nevertheless very important way in terms of saving resources and generating deceptively large uncertainty ranges, and that is to use observations to screen out model variants that are very poor at simulating historical climate. Stainforth et al (2005) use TOA imbalance to narrow climate sensitivity range in	Accepted - text revised and references included.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						one of their figures, Shiogama et al (in review I think) has recently used TOA imbalance of PPE driven by prescribed SSTs to identify model variants suitable for coupling to a ocean component. I am aware Covey et al have used screening procedures but again I think the paper is in review. Edwards et al (2011) have provided a statistically rigorous way of screening out model variants, and call the method “pre-calibration”. REFERENCE Title: Precalibrating an intermediate complexity climate model Author(s): Edwards Neil R.; Cameron David; Rougier Jonathan Source: CLIMATE DYNAMICS Volume: 37 Issue: 7-8 Pages: 1469-1482 DOI: 10.1007/s00382-010-0921-0 [David Sexton, UK]	
9-488	9	19	4	19	11	In this paragraph, it would help to use the word “weighting” which seems to be the “elephant in the room” as it is described on line 7 but not mentioned explicitly. I have found the word “weighting” to be more accessible to non-experts as well as experts even if the former don't understand the technical details. [David Sexton, UK]	Accepted. This discussion is revised to include discussion of weights as suggested.
9-489	9	19	8	19	9	The following paper should be added here: Loutre et al., 2011, Evaluating climate model performance with various parameter sets using observations over the last centuries, Climate of the Past, 7, 511-526, doi:10.5194/cp-7-511-2011. [Thierry Fichefet, Belgium]	Accepted. This citation will be included.
9-490	9	19	10	19	11	doi:10.1175/2011JCLI4193.1 also reports on a perturbed-parameter ensemble; this may be worth citing because the underlying model differs from the other studies listed here. [Robert Pincus, USA]	Accepted - citations added.
9-491	9	19	10	19	11	Murphy et al (2004) and Sexton et al (2011a) should be added to this list of references. Also Piani et al (2005). Sexton et al is new but Murphy et al isn't and I am not sure why it has not included. Is it because they used a basket of metrics instead of a “specific observational constraint”? If so, seems a bit strange to only consider specific constraints and not multivariate constraints and besides, Forest et al use upper atmospheric temperatures and surface temperatures in their constraints. Either these papers need to be added to the list of references or another sentence has to be added about use of multivariate observations to constrain uncertainty. [David Sexton, UK]	Accepted. Citations added.
9-492	9	19	10			Surprising that the phrase “now being applied to more complex models” makes no fewer than 5 references to analyses of one specific model (HadCM3/SM3) and ignores other work both past and recent involving other GCMs, such as Annan et al SOLA 2005 (et seq), Jackson et al 2008, Klocke and Pincus, and doubtless others. [James Annan, Japan]	Accepted. Additional citations are included.
9-493	9	19	13	19	13	Better than what? The context suggests than perturbed physics ensembles, but I think this should be than single model results. [Robert Colman, Australia]	Noted. This discussion has been removed.
9-494	9	19	13	19	16	As the sentence is written now it is ununderstandable. [Farahnaz Khosrawi, Sweden]	Noted. This discussion has been revised and does not include this material.
9-495	9	19	13	19	25	It is not clear why this paragraph is in a chapter on model evaluation. Is the intent to describe the evaluation of the multi-model mean or some other ensemble estimate? If so, this link should be made explicitly. [Robert Pincus, USA]	This discussion pertains to the use of the MME as a measure of uncertainty in the climate models and is not directly addressing whether the MME mean is a better match to the observations. Regarding the uncertainty quantification, this identifies the MME as a lower bound on the uncertainty.This section will be combined with material in FOD Section 9.7 and 9.4.1 regarding the use and interpretation of the MME.
9-496	9	19	13		14	Actually, Annan and Hargreaves (J Clim 2011) demonstrated that this was due to a trivial algebraic identity, their eqn 1: "this simple algebraic identity (which appears to have folkloric status within the wider ensemble prediction community, e.g., Stephenson and Doblas-Reyes 2000, but is perhaps not so widely known elsewhere) makes no appeal to any properties of the errors in the underlying physics of the models, nor how the ensemble was generated, and does not rely on any probabilistic interpretation of sampling distributions or use of an expectation operator, rather applying immediately to any finite sample. Furthermore, it does not even depend on where the observations happen to lie relative to the ensemble and therefore *is not contingent on the oft-	Noted. This section has been clarified to address the characterization of uncertainty by the MME. This has been combined with material in FOD Sections 9.4.1 and 9.7 to more appropriately address the MME mean concern raised here.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						cited property that the errors of different models have a tendency to cancel*. On the contrary, this result holds even if the signs of the errors are the same for all ensemble members." [James Annan, Japan]	
9-497	9	19	23	19	25	You should say more about the limits of PPE ("likelihood based" ensembles). They don't take the full uncertainty into account because they rely on one specific model and its corresponding structural assumptions. This seems to be acknowledged by the IPCC Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections: "Finally, it should be noted that few studies have addressed the issue of structural model inadequacies, i.e., errors which are common to all general circulation models (GCMs)." [Gregor Betz, Germany]	Noted. This has been rewritten to address additional uncertainties.
9-498	9	19	27	19	27	The title of this section makes not clear what the purpose of this section is. I would suggest to rename it into: "Overall Summary of model evaluation approaches". [Farahnaz Khosrawi, Sweden]	Accepted. Title changed.
9-499	9	19	29	19	30	Isn't that natural do use observational data for model evaluation since this is how it should be done? If one would use models for the evaluation it wouldn't be an evaluation. It would rather be a comparison. So, I would suggest to skip this sentence. [Farahnaz Khosrawi, Sweden]	Accepted. Sentence removed.
9-500	9	19	30			“observationally-based product” is quite a mouthful of jargon. Presumably what’s meant here are things like reanalyses. [Robert Pincus, USA]	Accepted. Sentence reworded.
9-501	9	19	33	19	33	Intermodal? [Robert Colman, Australia]	Accepted. Replaced with 'inter-model'
9-502	9	19	33	19	33	intermodal or intermodel? [Graham Feingold, United States of America]	Accepted. Replaced with 'inter-model'
9-503	9	19	38	19	38	" extreme behaviour of particular relevance to society". Too vague. What is meant by "behaviour". Also justify why this approach is taken, which seems sudden and ad hoc: shouldn't the focus be on those processes which shed most light on the reliability of the models? [Robert Colman, Australia]	Accepted. The revsied version just refers to extremes.
9-504	9	19	57	20	1	This is a key statement that should be articulated more prominently in above sections that report limitations and also in the summary. You might consider citing here: 1. Stainforth, D. A., M. R. Allen, et al. (2007). "Confidence, uncertainty and decision-support relevance in climate predictions." Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences 365(1857): 2145-2161 and 2. Gregor Betz, "What’s the worst case?", Analyse und Kritik, 32(1), 2010, pp. 87-106. [Gregor Betz, Germany]	Accepted. Rather than adding a referenc here we refer to Section 9.2.2.7 where this is discussed in more detail.
9-505	9	19	57			Here and elsewhere the spread or range is used to characterize the distribution of model projections. This has historical value but is statistically quite fragile. Might another metric (5-95% confidence bounds, as one example) be more useful? [Robert Pincus, USA]	Accepted. Rather than adding a referenc here we refer to Section 9.2.2.7 where this is discussed in more detail.
9-506	9	20	1	20	1	provide a lower bound estimate of model uncertainty'. Not necessarily. For example, there may be "outliers" which can be eliminated as "wrong", producing a smaller, but more tightly agreeing sample. Statistical combination of models may also conceivably reduce spread, as might use of downscaling techniques. [Robert Colman, Australia]	Agreed. Lower bound removed and link to Section 9.2.2.7 added, where this is discussed in more detail.
9-507	9	20	6	20	8	I suggest “Use of Observations in Support of Climate Evaluation” and “The Role of Model-Observation comparisons” may be developed as a sub-section here. For example, satellite observational data can be used to evaluate spatial patterns of the mean state of atmospheric moisture, clouds and radiation in simulations of recent climate. [Hui Su, United States of America]	Taken into account -- the role of model-observation comparisons is described in greater detail earlier in the chapter in sections 9.2.2.1 (evaluating the overall model results) and 9.2.2.3 (instrument emulators).
9-508	9	20	8	20	23	Model intercomparisons have proliferated in recent years. Some (like CMIP5) are very valuable, others not so much. Section 9.3.1 could say more about the many intercomparisons and their conclusions. Perhaps a table would be useful. [David Randall, USA]	Taken into account -- utiliziation of MIPs in several chapters is now noted in the text.
9-509	9	20	10	20	33	Model intercomparison merely ensures that all models are equally wrong [VINCENT GRAY, NEW ZEALAND]	Noted
9-510	9	20	18	20	18	"More realistic processes" is badly/vaguely worded, suggesting that previously they were "unrealistic", or even that they were the wrong processes. I suggest avoiding the use of realistic completely, perhaps saying "as representation of key processes is improved". [Robert Colman, Australia]	Accepted - text revised

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-511	9	20	20	20	20	"All models suffer from errors...". True, by definition, as the concept of a 'model' is understood in the scientific community, but this is an open-ended statement with potential for misquotation or misunderstanding. This could equally well be written: "No model is perfect, and model evaluation is a necessary step towards identification of model errors.", or simply "Model evaluation is a necessary step towards identification of model error". [Robert Colman, Australia]	Accepted - text revised
9-512	9	20	31			Insert 'and AOGCMs' after 'ESMs'. [Nathan Gillett, Canada]	Accepted - text revised
9-513	9	20	34	20	34	Definition of "effective climate sensitivity". Not sure you can just leave this important concept to the glossary. [Gareth S Jones, UK]	Taken into account -- The point of this sentence is not to dwell on the precise meaning of effective climate sensitivitiy, but to point out that the response of the model has been shown to depend on its initial state. Effective climate sensivivity is defined in the glossary.
9-514	9	20	34	20	36	This statement is seemingly inconsistent with the following from chapter 9: "Showing anomalies is reasonable since while the models exhibit differing biases in their means, climate sensitivity is not a strong function of the mean state in climate models (Stainforth et al., 2005)." (Chapter 10, p14, lines 53-55). The respective authors should resolve any inconsistencies. [Chris Roberts, Uk]	Agreed -- the linkage reproduction of the observed state and climate sensitivity has been removed.
9-515	9	20	35	20	36	This must depend on the variable being predicted. For some variables perhaps mean state errors are less important. Stainforth et al. (2005) did not find a strong relationship between mean state and climate sensitivity in a perturbed physics ensemble. [Nathan Gillett, Canada]	Taken into account -- combined with comment 9-514.
9-516	9	20	37	20	41	The discussion of uncertainty in projectsions because of uncertainties in the historical and present states would benefit from a more complete explanation, similar to that in the sensitivity discussion on pg 67, sec 9.7.4.3 [David Bader, USA]	Accepted -- this statement has been supported by a reference to Kocke, Pincus, and Quaas (2011) making the same point.
9-517	9	20	38	20	41	Interpreting 'the observational record' to mean all observable properties of the climate system, agreement with the observational record is the best we can hope for for a climate model, and we cannot expect to narrow uncertainties further. As written the statement implies that there may be other ways of narrow the range of projections, but there are not. Any further narrowing of the uncertainties not based on observations would be spurious and lead to false confidence in projections. [Nathan Gillett, Canada]	Taken into account -- Better theory and better computational formulation are two additional avenues (i.e., the two other major types of sciences besides empirical/experimental) for improving the projections from climate models. The sentence is correct as it appears in FOD and has not been modified.
9-518	9	20	38	20	41	agree with the uncertainties in forcing: but limits because of changes in ocean heat storage and coupled processes? puzzling. ultimately tests of models must be against observations. PPEs and varying initial condition tests are useful additions, but are not as fundamental tests as observations: in particular decadal change observations (much less so for mean state observations). [Bruce Wielicki, USA]	Accepted -- text revised to clarify
9-519	9	20	40			Add "e.g.": "uncertainties in prjections due to, e.g., remaining…" [Tor Eldevik, Norway]	Editorial
9-520	9	20	43	20	44	Which three classes of experiments? [Farahnaz Khosrawi, Sweden]	Accepted -- text revised to clarify. The three classes of experiments are: (1. Long-range simulations with AOGCMs; (2. Long-range simulations with ESMs; and (3. Decadal-length projections.
9-521	9	21	5	21	5	... 1850 to 2005 (van Vuuren et al.). ... Which publication is meant? [Marco Giorgetta, Germany]	Editorial
9-522	9	21	5	21	5	van Vuuren et al. : the year is missing [Eric Martin, France]	Editorial
9-523	9	21	5	21	5	Van Vuuren et al. ???? Year is missing [HASIBUR RAHAMAN, India]	Editorial
9-524	9	21	6	21	6	... (Moss et al., 2010; Van Vuuren et al., 2011). ... [D. P. van Vuuren et al., The representative concentration pathways: an overview, Climatic Change (2011) 109:5–31, DOI 10.1007/s10584-011-0148-z] Describes the updated RCP scenarios. This overview paper is part of a special issue on the RCP scenarios with further publications for the single RCP scenarios. [Marco Giorgetta, Germany]	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-525	9	21	7	21	7	... Extension Concentration Pathways (ECPs; Meinshausen et al., 2011). ... [Meinshausen et al., The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change (2011) 109:213–241DOI 10.1007/s10584-011-0156-z] [Marco Giorgetta, Germany]	Editorial
9-526	9	21	7			The abbreviation ECP is introduced and never used again, so should be omitted. [Philip Mote, USA]	Accepted - text revised
9-527	9	21	11	21	16	"were left to the discretion of the individual modelling groups" this sound very odd and not true for a number of GHG concentrations. Please be precise, list and distinguish between which forcing were/are specified in CMIP3/CMIP5 and which are now better specified. [Elisa Manzini, Germany]	Accepted -- We now reference table 10.1 in the WG1 AR4 report to thoroughly document the diversity present in CMIP3 forcing agents.
9-528	9	21	14	21	14	As stated already in one of my previous comments I am not fond of this abbreviation. [Farahnaz Khosrawi, Sweden]	Accepted - text revised
9-529	9	21	18	21	19	The models should treat natural and anthropogenic CO2 forcing separately. Is it done ? How is it discriminated ? [François GERVAIS, France]	Taken into account -- The separation of increased levels of CO2 from anthropogenic emissions and natural feedbacks is part of the experimental design for CMIP5 (Hibbard, Meehl, Cox, and Friedlingstein, EOS Transactions, 2007).
9-530	9	21	18	21	35	The amount of detail here seems inappropriate. [Robert Pincus, USA]	Taken into account -- The much more complete prescription of radiatively active species and their chemical precursors is a distinguishing feature of the RCPs and the CMIP5 protocol in general relative to the SRES scenarios in CMIP5 and the AR4. In order to highlight that advance, we would prefer to keep the detailed list of species.
9-531	9	21	24	21	25	As stated elsewhere, of course emissions are only prescribed in emissions-driven simulations - all ESMs are also run in prescribed concentrtaion mode too. [Nathan Gillett, Canada]	Accepted - text revised
9-532	9	21	33	21	33	Delete "and" [Farahnaz Khosrawi, Sweden]	Editorial
9-533	9	21	43	21	46	I was under the impression that many centers will be doing annual initialized decadal prediction experiments. [Annarita Mariotti, U.S.A.]	Accepted - text revised
9-534	9	21	48	21	48	Somewhere in this paragraph the work of Crook and Forster need to be incorporated (J. Geophys. Res., doi:10.1029/2011JD015924). Their AOGCM evaluation seems very relevant for this para. [Bram (Abraham) Bregman, Netherlands]	Taken into account. Suggested reference has been considered in Section 9.7
9-535	9	21	52	22	2	This section could include a discussion on reanalysis data, see comment above [Gunilla Svensson, Sweden]	Taken into account. Text has been modified.
9-536	9	21	53	21	53	... to confront models ... --> ... to evaluate models ... The earlier text on the model development uses "evaluation", and this is the word that should be used here again. [Marco Giorgetta, Germany]	Taken into account. Text has been modified.
9-537	9	21	53	22	2	This introduction makes it sound as if the modelling community has set its priorities in an ad hoc way from availability of observations, and also begs the question that there may be model evaluation that has not been done that is fundamental for climate sensitivity, for example. It is important to point out that emphasis on atmospheric model evaluation has been motivated by a number of additional factors, including our understanding of critical physical processes (for example that clouds, atmospheric temperatures (lapse rate) and water vapour are critical in determining climate sensitivity), and because many of the impacts we care about from climate change result primarily from atmospheric changes (from temperature, rainfall, extreme wind/rain/temperature events etc). [Robert Colman, Australia]	Taken into account. Text has been modified.
9-538	9	21	54			cloud cover and cloud ice - (ice values from cloudsat or cloudsat/calipso are robust enough for model evaluation) [Duane Waliser, USA]	Taken into account - The text was changed to "cloud cover and cloud condensate"
9-539	9	21		24		Consider including in section 9.4.1.1 the following: Zhang, Y., D. J. Seidel, J.-C. Golaz, C. Deser, R. A. Tomas, 2011: Climatological characteristics of Arctic and Antarctic surface-based inversions. J. Climate, 24,	Noted. This literature has been considered.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						5167-5186, doi: 10.1175/2011JCLI4004.1. which compares surface-based inversions in the Arctic in several models with radiosonde data and finds significant model biases. [Melissa Free, USA]	
9-540	9	21		30		While the results for the reliability at the LGM (from Hargreaves et al, CP, 2011) have been neatly summarised in section 9.4.1.5, I'm surprised to find that the much more comprehensive analysis for the present day of Yokohata et al, Clim Dyn, 2011 has been omitted. Perhaps the results may be summarised around section 9.4.1.2. In short, we found that the CMIP3 ensembles (ie slab and coupled ocean) were reliable on broad temporal and physical scales, over a large number of variables, including the trend in 20th century temperature. I can't believe you left out these encouraging results! Work with CMIP5 is presently underway, but not yet submitted. [Julia Hargreaves, Japan]	Taken into account. Ensemble reliability is now addressed in Section 9.8
9-541	9	21		40		You have split the world into various components but have no place for discussing combinations of things together. I suspect that such discussions are hidden in the various subsections, which would not make for easy comprehension of the material. I have only read targeted sections in detail but immediately noticed one such case, mentioned in my next comment, where the LGM SST is discussed in the atmosphere section. [Julia Hargreaves, Japan]	Noted. The results of our assessment are brought together as a collective in Section 9.8
9-542	9	22	4			Section 9.4.1.1: Having mentioned the observational uncertainty in section 9.2, it is not mentioned in this section, and is not mentioned again until section 9.4.1.2 where metrics are shown. It would be helpful to have some indication of the relative magnitude of the MME biases and the range of observational estimates. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Text has been modified
9-543	9	22	4			Section 9.4.1.1: Since both CMIP5 and CMIP3 results are used in these sections, it might be worthwhile showing both where possible. For instance, at the moment, all the results discussed in 9.4.1.1.2 are from CMIP3, whereas in 9.4.1.1.1 we have seen surface air temperature and precipitation only from CMIP5. I realise this reflects the early stage of this draft, but ulimately we would have more confidence in this interchangeability of results were we to be shown the surface air temperatures and precipitation from CMIP3 as well as CMIP5 in section 9.4.1.1.1. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. CMIP3 and CMIP5 results are shown in the revised Fig 9.8.
9-544	9	22	6	22	25	Models are tested on their ability to reproduce the spatial pattern of observed means. But you do not discuss any work on simulations of the spatial pattern of observed trends, which seems to me to be the more relevant test of model quality. The only paper I know of is McKitrick and Tole (2011 see cell 67) who show that the AR4 models generally do a very poor job of reproducing the spatial pattern of gridded trends. Most models fail to provied significant explanatory power for the spatial pattern of trends over land when compared to a simple vector of socioeconomic variables, and even when a Bayesian Model Averaging framework is used that evaluates every possible linear combination of models, most fail to exhibit any explanatory power and the variables with the highest probability of obtaining support in the data remain socioeconomic observations rather than GCM-generated temperatures. This paper is being revised for Climate Dynamics. [Ross McKitrick, Canada]	Noted. The pattern of temperature trends is discussed in Chapter 10.
9-545	9	22	6	22	25	McKitrick, Ross R. and Lise Tole (2011) "Evaluating Explanatory Models of the Spatial Pattern of Surface Climate Trends using Model Selection and Bayesian Averaging Methods" submitted to Climate Dynamics [Ross McKitrick, Canada]	Noted
9-546	9	22	7	22	7	Avoid 'realistic'. This is a difficult word: what does this mean? [Robert Colman, Australia]	Noted
9-547	9	22	16	22	25	See comments regarding Fig. 9.3 p 128 [Peter Guttorp, USA]	Noted.
9-548	9	22	16	22	25	"The multimodel …" this sentence does not contribute to the understanding. Please discuss the results in the figure instead. The entire paragraph is not very well written and raises questions Why "also" in the sentence about maximum annual temperatures. The overall bias is that all models seem to show to cold temperatures over the ocean. South America also show large biases. The average CMIP5 model error subfigure is very interesting. [Gunilla Svensson, Sweden]	Taken into account. The figure and corresponding text have been completely revised.
9-549	9	22	17	22	18	This sentence hard to understand. [Robert Colman, Australia]	Editorial
9-550	9	22	17	22	18	What does the sentence about 'distinctive gradients of observed temperatures' mean? [Nathan Gillett,	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Canada]	
9-551	9	22	17	22	18	What does the sentence "The multi-model ensemble .... decrease with latitude" mean? Is it to say that observed latitudinal gradients are reproduced? Is the reproduction becoming worse (decreasing) with latitude? Or doe observed gradients decrease with latitude? Which I very much doubt – latitudinal gradients are higher at high latitudes. [Andreas Sterl, Netherlands]	Taken into account. The figure and corresponding text have been completely revised.
9-552	9	22	20	22	20	it should be useful to mention/specify theorigin of the data from Jones et al., 1999 [ANNALISA CHERCHI, Italy]	Noted
9-553	9	22	20	22	20	Why are you using the observations as in Jones (1999) given the updated records provided by chapter 2 of this report? [Celeste Saulo, Argentina]	Taken into account. Observations have been updated
9-554	9	22	22	22	22	"were the biases…" --> "where the biases…" [Hai Lin, Canada]	Editorial
9-555	9	22	27			Figure 9.3: the red color scale in the bottom figure is hard to read quantitatively: use rainbow color scale as in top figures? [Bruce Wielicki, USA]	Taken into account. The figure has been completely revised.
9-556	9	22	28			I would prefer deg C instead of Kelvin for this figure. Lay people usually do not know what Kelvin is. [Andreas Sterl, Netherlands]	Taken into account. Units have been change.
9-557	9	22	30	22	30	the reference is Jones et al., 1999 [ANNALISA CHERCHI, Italy]	Noted
9-558	9	22	30			Bias – I am not sure whether most people would understand this. Better to write "difference between MME and observations". [Andreas Sterl, Netherlands]	Taken into account. Terminology is clarified.
9-559	9	22	32	41	3	Simulation is a euphemism fo correlation, which can never prove cause and effect, however convincing the fit [VINCENT GRAY, NEW ZEALAND]	Rejected. Computer simulations performed with general circulation models are based on physical laws.
9-560	9	22	34	22	34	Change "Figure 9.3" with "Figure 9.4" [ANNALISA CHERCHI, Italy]	Noted
9-561	9	22	34	22	34	Fig. 9.4 [Graham Feingold, United States of America]	Noted
9-562	9	22	34	22	34	TYPO: "Figure 9.4." instead of "Figure 9.3." [Ladislav Metelka, Czech Republic]	Editorial
9-563	9	22	34	22	34	The top panel of Figure 9.3 it should be Figure 9.4 [HASIBUR RAHAMAN, India]	Editorial
9-564	9	22	34	22	34	it should be Figure 9.4 [Celeste Saulo, Argentina]	Editorial
9-565	9	22	34			typo: fig 9.4 [Ramon de Elia, Canada]	Editorial
9-566	9	22	36	22	36	change "middle" with "bottom": the figure has only two panels [ANNALISA CHERCHI, Italy]	Editorial
9-567	9	22	36	22	36	The bottom panel [Graham Feingold, United States of America]	Editorial
9-568	9	22	38	22	39	"Over land, the models tend to overestimate the temperature range with magnitudes noticeably larger than over the oceans". Actually, I notice few spots of large bias (in bsolute sign) in some specific and small land areas, mainly at high latitudes. Those over deserts and close (Black and Caspi) seas are lower and comparable (in absolute intensity) with the biases over oceans. [Claudio Cassardo, Italy]	Taken into account. Figure and text have been completely revised.
9-569	9	22	38	22	39	Over land, models tend to overestimate the cycle -except over Siberia- (or perhaps you prefer some "warning" sentence regarding lack of data/poor reliability over that portion of Asia?) [Celeste Saulo, Argentina]	Taken into account. Figure and text have been completely revised.
9-570	9	22	42			Figure 9.4: missing the third bottom figure on CMIP average error to be consistent with figure 9.3 [Bruce Wielicki, USA]	Taken into account. Figure has been completely revised.
9-571	9	22	43	22	43	Here figure 9.3 is included, but where is it described in the text? [Farahnaz Khosrawi, Sweden]	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-572	9	22	47	22	55	Another study that evaluate model performance in the regions of the tropical Pacific convergence zones include: Irving, DB et al. 2011 in Climate Research Vol 49(3) [Brad Murphy, Australia]	Taken into account. Suggested literature has been considered.
9-573	9	22	47	22	55	The noted challenge in representing preciptation is good but it it too superficial - the long standing double ITCZ bias in accumulation remains an issue but comparisons of accumulation alone is no real test of model preciptation physics becauseaccumulation is controlled by climate energetics - the breakdown of accumulation into frequency and intensity reveal marked biases that is a little too glossed over in lines 2-9 on 9-43 as drizzle bias - I feel some comment is needed to add a little more depth to the comparison of fig 9-5. [Graeme Stephens, USA]	Taken into account - An assessment of the literature regarding the model "drizzle" problem has been added to this paragraph.
9-574	9	22	54	22	55	Could also add a citation to: Brown, J. R., S. B. Power, F. P. Delage, R. A. Colman, A. F. Moise and B. F. Murphy (2011), Evaluation of the South Pacific Convergence Zone in IPCC AR4 climate model simulations of the 20th century, Journal of Climate, 24, 1565-1582. This study evaluates simulation of SPCZ in CMIP3 models. [Josephine Brown, Australia]	Taken into account. Suggested literature has been considered.
9-575	9	22	57	22	57	fig. 9.5: it is more convenient to put precipitation data in mm/day [ANNALISA CHERCHI, Italy]	Taken into account. Units for precipitation have been changed.
9-576	9	22	57			Figure 9.5: use of units where color scale values are "1e-05 to 0.00013" is awkward at best. Different unit scale or multiply by a constant would help. CMIP 5 average error has same issue as figure 9.3: red color scale is difficult to read quantitative values [Bruce Wielicki, USA]	Taken into account. The figure has been completely revised, including a change of units.
9-577	9	22				Fig.9.5: Because precip is zero-bounded its errors tend to scale with its mean, so showing mm/s (or similar) errors is not really very informative, except in the wettest regions. Percentage errors would be much better - I know they would be large over dry areas (this could be pointed out as less relevant or blanked out), but overall would be much more informative globally. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Figure has been revised.
9-578	9	22				Fig.9.5: Consider showing obs precip as well (then 2x2 panels in this Fig.) to more clearly support the text. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Suggestion considered.
9-579	9	23	1	23	1	As in comment #40: your reference data is "updated Adler (2003)" which is probably similar as that used in Chapter 2, but it would be better to use the same data set and reference it as in Chapter 2 (as you do with respect to radiation and clouds, following Chapter 7). [Celeste Saulo, Argentina]	Taken into account. Observations and reference are now consistent with Chapter 2.
9-580	9	23	1	23	3	Figure 9.5 The units are not those commonly used for precipitation like mm/d. [Annarita Mariotti, U.S.A.]	Taken into account. Units for precipitation have been changed.
9-581	9	23	1		3	The precipitation unit is very strange. It is much easier to understand the meaning of precipitation amount with a unit of mm day-1. [Kuan-Man Xu, USA]	Taken into account. Units for precipitation have been changed.
9-582	9	23	1			Figure 9.5: Only GPCP precipitation data is used to calculate model bias. The difference between CMAP and GPCP can be larger than model bias, e.g. in tropical Pacific. Therefore, the choice of precipitation data set should be justified and the uncertainty associated with CMAP versus GPCP difference should be discussed. [Josephine Brown, Australia]	Observational uncertainties are now noted. GPCP is the default because it has been judged in the literature to be more reliable.
9-583	9	23	5	24	29	Section 9.4.1.1.2 should make more reference to Chapters 7 and 8. [David Randall, USA]	Taken into account - References to Chapters 7 and 8 have been added as appropriate.
9-584	9	23	6	23	6	"total moisture content" --> "total moisture content (water vapor, liquid water and ice)" [Claudio Cassardo, Italy]	Rejected -- Precipitable water is conventionally defined just as water vapor, not condensed water. Text that follows only concerns water vapor.
9-585	9	23	6	23	19	There is a paper in revision for JGR by Jiang et al (2012) that is very relevant here showing the joinf statiticts of UT humidity and cloud ice and the performace of CMIP3 and CMIP 5 models wrt to these joint statistics - it reveals biases in models both in UTH and ice and that biases dont logically follow one another (ie dry uth less ice) - these baises are important for a number of reasons including issue wrt feedbacks - I can have authors send the revised version to appropriate LAs [Graeme Stephens, USA] JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 117, D14105, 24 PP., 2012 doi:10.1029/2011JD017237	Taken into account - The paper has been assessed and its conclusions have been added to this subsection.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-586	9	23	10	23	11	"…sub-grid vertical transport" it should be "sub-grid scale vertical transport .." [HASIBUR RAHAMAN, India]	Accepted. Text revised.
9-587	9	23	13			This bias of 100% is presumably at an individual latitude and height. When I first read this I interpreted this as a bias in the water content of the whole of the free troposphere of 100%, but I don't think it can be. Clarify. [Nathan Gillett, Canada]	Accepted. Text revised to clarify.
9-588	9	23	21	23	21	flues --> fluxes [Claudio Cassardo, Italy]	Editorial -- text corrected.
9-589	9	23	21	23	21	"flues"? I guess you mean "fluxes" [Farahnaz Khosrawi, Sweden]	Editorial -- text corrected.
9-590	9	23	21	23	28	The paper by Bender et al 2006 paper motivates a more modest wording of this paragraph. Note that tis paper, like Pincus et al. is based on CMIP3 models. Bender, F. A-M., Rodhe, H., Charlson, R.J., Ekman, A.M.L. and Loeb, N. 2006. 22 views of the global albedo – comparison between 20 GCMs and two satellites. Tellus A 58, 320-330. [Henning Rodhe, Sweden]	Taken into account - The paper has been assessed and its conclusion included.
9-591	9	23	21	23	53	lines 46-53 more naturally follow line 21-28 then they do lines 30-44. [Robert Pincus, USA]	Accepted -- paragraph (lines 48-53) moved to just after paragraph recommended (lines 21-28)
9-592	9	23	21		22	spatial patterns and annual cycle of the radiative flues at the top of the atmosphere. It would seem that these should be shown and compared to obs, say in zonal mean; also difference plots. [Stephen E Schwartz, USA]	Noted - As the mainsource of errors for these fluxes are clouds we decided to only show figures of errors in cloud radiative effect. Articles cited to support include Bender 2011 and Donahue and Battisti 2011.
9-593	9	23	22			Given that this section is mainly (exclusively?) discussing CMIP3 results, should "current" be replaced by "CMIP3"? Is it possible to insert results for CMIP5 yet? Otherwise the section could be trimmed. [Philip Mote, USA]	Will add discussion of "J.-L. F. Li, D. E. Waliser, G. Stephens, Seungwon Lee, T. L'Ecuyer, Seiji Kato, Seungwon Lee, W-L Lee, (2012), Characterizing and Understanding Cloud Water and Radiation Budget Biases in CMIP3/CMIP5 GCMs, Contemporary GCMs and Reanalyses", (JGR, to be submitted) once it becomes available.
9-594	9	23	23		23	The data used in Pincus et al. (2008) were very old version of CERES data. The most updated version of the data is called EBAF version 2.6 (Loeb et al. 2009). The CERES data should be updated for this figure. [Kuan-Man Xu, USA]	Taken into account - Figure 9.6 has been redrawn using the new data set and CMIP5 models.
9-595	9	23	27	23	28	This is inaccurate. The Pincus et al. paper shows that climate model errors in cloud radiative effects are commensurate with those from a free run of a current (at the time) weather forecasting system. Clouds in the ERA-40 reanalysis have significantly higher errors. [Robert Pincus, USA]	Taken into account - The sentence has been modified to point to the discussion of CRE in Figure 9.6.
9-596	9	23	28			Name these individual models. [Nathan Gillett, Canada]	Noted -- the sentence in question discussing the fidelity of individual models relative to ECMWF analysis has been removed in the SOD.
9-597	9	23	30	23	30	surface fluxes --> surface components of radiation fluxes [Claudio Cassardo, Italy]	Accepted -- text revised.
9-598	9	23	30	23	31	"the downward all-sky shortwave flux" --> "the downward shortwave flux" (all-sky seems useless here) [Claudio Cassardo, Italy]	Rejected -- the distinction between clear-sky and all-sky is an important distinction in this context.
9-599	9	23	30	23	33	Giving mean model radiation errors is not helpful without 1) giving some idea of the observed value. 6W/m2 is a big error if fluxes are 10W, irrelevant if they are 300W. 2) giving some idea of model spread. Do models encompass the observational estimates? What are the estimated errors on obs? [Robert Colman, Australia]	Accepted -- Information regarding observed values and model spread has been added.
9-600	9	23	30	23	44	I was asked to write a review article for nature geosciences on the energy balance of the planet and in it is the most comprehensive review of fluxes & uncertainties and in ig I compare obs to CMIP5 -the paper is currently in review - the numbers quoted in this paragraph will need revision - for example the observationally infered solar flux to the surface and the model fluxes are much closer than the 12 Wm-2 Martin Wilds work has suggested and quoted in the paragraph and hence arguments for effects of absorbing aerosl etc are less	Accepted -- These papers are now cited in this portion of the text and the numbers have been revised.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						dramatic. The longwave flux to the surface is 8Wm-2 larger in observations than models and interestingly reanalysis products (mostly model based) are biased even lower (this is discussed in my paper Stephens et al., 2012 J CLimate now appearing and I can send both the review and the J CLiamte paper to relevant LAs - there are other knock-on effets to other surface fluxes, notably evaporation which I argue islarge than previously quoted. Martin Wild is a co-author on both papers [Graeme Stephens, USA]	
9-601	9	23	31	23	31	"the corresponding downward longwave flux" --> "the downward longwave flux": why "corresponding? Longwave radiation flux does not "correspond" to shortwave one. [Claudio Cassardo, Italy]	Editorial -- "corresponding" removed.
9-602	9	23	32			Replace '-5.6' by '5.6'. As written, this sentence says an underestimate by -5.6 W/m^2, which is an overestimate. [Nathan Gillett, Canada]	Editorial -- Sign fixed.
9-603	9	23	35			Another notable and systematic bias relevant to this discussion is the positive bias in surface downwelling and outgoing longwave flux associated with the ITCZ (Trenberth and Fasullo, 2009). There is evidence that this systematic bias stems from ignoring the impact of precipitating hydrometeors in the radiation calculations in the CMIP3 models (Waliser et al. 2011). Trenberth, K. E., and J. T. Fasullo (2009), Simulation of present day and 21st century energy budgets of the southern oceans, Journal of Climate, 23, No. 2, 440-454. Waliser, D. E., J.-L. Li, T. L'Ecuyer, and W.-T. Chen (2011), The Impact of Precipitating Ice and Snow on the Radiation Balance in Global Climate Models. , Geophysical Research Letters, 38, L06802, doi:10.1029/2010GL046478. [Duane Waliser, USA]	Noted -- In Waliser et al, the bias due to ignoring ice precip falls beween 1 and 3 W/^2 but it is not quantified in the article.
9-604	9	23	39			typo: "to THE omission" [Ramon de Elia, Canada]	Editorial -- typo fixed.
9-605	9	23	43	23	44	Maybe a discussion about the reason for which surface albedo differs in the various models can be useful here. [Claudio Cassardo, Italy]	Accepted --new estimates of the variation in surface albedos among the CMIP3 models have been added here.
9-606	9	23	46	24	10	results from CMIP5 should be included, even in fig. 9.6 [ANNALISA CHERCHI, Italy]	Taken into account - CMIP5 model results have been included in Figure 9.6.
9-607	9	23	50	23	50	Skip the abbreviation "CRE", it is not really useful. [Farahnaz Khosrawi, Sweden]	Rejected - As the abbreviation needs to be used frequently, it needs to be introduced here.
9-608	9	23	55			Why are CERES ES-4 data being used in preference to the CERES EBAF data? [Robert Pincus, USA]	Taken into account - Figure 9.6 has been redrawn using the new data set and CMIP5 models.
9-609	9	23	55			the comparisons of climate model CRE and observations should be done to the more recent 10 years of CERES EBAF data, which is much higher accuracy than the ES-4 ERBE-Like CERES data. Use of the old ERBE S-4G 5-year data set should be phased out. The newer data is factors of 2 to 3 more accurate. [Bruce Wielicki, USA]	Taken into account - Figure 9.6 has been redrawn using the new data set and CMIP5 models.
9-610	9	23	55			The comparison of climate models to observations in general lacks any method to define "how accurate is good enough". This is true of mean state comparisons, seasonal comparisons, ENSO comparisons, etc. A large part of the problem is that these comparisons are only loosely related to the desired decadal prediction accuracy. A good summary of this issue can be found in the recent paper by Klocke, Pincus, and Quaas, J. Climate, 2011. Needs to be mentioned as a current limitation. [Bruce Wielicki, USA]	Noted - We appreciate the reviewers concern. However, we feel it would be inappropriate to single out cloud radiative effects in this respect. A general discussion of the question on how to translate model evaluation into reliability for projections, including the Klocke et al. paper are carried out in Section 9.8.
9-611	9	23	56	24	2	For people working for CERES, CERES ES-4 and ERBE S-4G are not data to be used for IPCC because the cloud radiative effects are not very accurate. The right data for use is called EBAF (Loeb et al. 2009) version 2.6. [Kuan-Man Xu, USA]	Taken into account - Figure 9.6 has been redrawn using the new data set and CMIP5 models.
9-612	9	24	6	24	7	"A large underestimation … over the sub-polar oceans": if the figure shows MOD-OBS, since those areas are red (=positive), it means MOD>OBS, thus it should be an overestimation. [Claudio Cassardo, Italy]	Taken into account - as shortwave CRE is defined as negative, an overestimation in the mathematical sense is an underestimation of the effects clouds should have. The sentence was changed to clarify this.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-613	9	24	12	24	25	This is a great deal of detail and not very much synthesis. [Robert Pincus, USA]	Noted - We disagree. The notion that it is process interactions and cloud phase that drive errors is synthesis. The evaluation of clouds has been given more prominence in the SOD and this paragraph has been incorporated into this new discussion.
9-614	9	24	16	24	16	The subtropical low cloud biases are extremely large for almost all current models, particularly in the Atlantic (Huang et al. 2007, Wahl et al. 2009; Hu et al. 2011), which may result in the SST zonal gradient reversion along the Equatorial Atlantic Ocean. [Zeng-Zhen HU, USA]	Noted - The suggested papers have been assessed. As they deal with individual models neither of which are included in this report, only a sentence on the possible consequences of the low cloud errors for SST has been added.
9-615	9	24	16	24	16	Wahl, S., M. Latif, W. Park, and N. Keenlyside (2009): On the Tropical Atlantic SST warm bias in the Kiel Climate Model. Climate Dynamics, DOI:10.1007/s00382-009-0690-9. [Zeng-Zhen HU, USA]	Noted - The suggested papers have been assessed. As they deal with individual models neither of which are included in this report, only a sentence on the possible consequences of the low cloud errors for SST has been added.
9-616	9	24	16	24	16	Huang, B., Z.-Z. Hu, and B. Jha, 2007: Evolution of model systematic errors in the tropical Atlantic basin from the NCEP coupled hindcasts. Clim. Dyn., 28 (7/8), 661-682, DOI: 10.1007/s00382-006-0223-8. [Zeng-Zhen HU, USA]	Noted - The suggested papers have been assessed. As they deal with individual models neither of which are included in this report, only a sentence on the possible consequences of the low cloud errors for SST has been added.
9-617	9	24	16	24	16	Hu, Z.-Z., B. Huang, Y.-T. Hou, W. Wang, F. Yang, C. Stan, and E. K. Schneider, 2011: Sensitivity of tropical climate to low-level clouds in the NCEP Climate Forecast System. Clim. Dyn., 36 (9-10), 1795-1811, DOI: 10.1007/s00382-010-0797-z. [Zeng-Zhen HU, USA]	Noted - The suggested papers have been assessed. As they deal with individual models neither of which are included in this report, only a sentence on the possible consequences of the low cloud errors for SST has been added.
9-618	9	24	16			"Li et al. (2012) concluded that significant disagreements of cloud liquid water path are found among CMIP3 models, reanalyses and satellite derived estimates." [Li, J.-L. F., D. E. Waliser, and J. H. Jiang (2011), Correction to “Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations,” Geophys. Res. Lett., 38, L24807, doi:10.1029/2011GL049956.] [Richard Allan, UK]	Taken into account - The evaluation of clouds has been given more prominence in the SOD and the suggested paper has been assessed and its conclusions included in that new discussion.
9-619	9	24	22	9	25	The papers cited here examine cloud properties in various regimes, defined either in terms of the large-scale circulation or in terms of the clouds properties themselves. This does not make them “process-oriented” methods for model evaluation. [Robert Pincus, USA]	Accepted - The word was changed to "Regime-oriented"
9-620	9	24	22			We are about to submit an update to the Waliser et al. 2009 paper that provides a similar but more robust analysis of cloud ice for cmip5 and a comparison between cmip3 and cmip5 performance - of which the latter is significantly better (50% improvement in RMSE of MME), although as a whole and individually with many shortcomings (LI, Waliser et al. 2012; abstract is below). ---- One thing that I think should be stressed in this discussion is that while the global radiation budget is often consistent with observations, it is based on much tuning of clouds based on (roughly) their cloud cover, particle size and ice/liquid mass. Until recently, good observations through the column of only clouds were arguably available leaving great flexibility with the other quantities. Recent observations from cloudsat/calipso are starting to provide key constraints on cloud mass profiles leading to improved representations.------- An Observation-Based Evaluation of Cloud Ice Water in CMIP3 and CMIP5 GCMs and Contemporary Analyses------- Abstract ------- Tropospheric ice clouds play a key role in the energy and water cycles of our climate, and their proper representation in our global models are necessary to achieving robust weather and climate simulations/forecasts. Moreover, due to the significance role of cloud climate feedback in determining present and future climate, it is crucial that global climate models (GCMs) demonstrate realistic tropospheric ice clouds to minimize global climate change projection uncertainties. In this study, we utilize a considerable heritage of knowledge and experience to perform a robust evaluation of the cloud ice water content (CIWC) and cloud ice water path (CIWP) from 20th century	Noted - The paper will be assessed when it becomes available.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						CMIP5 simulations and compare these results to the same analysis on CMIP3 and four other GCMs of interest (GEOS-5 AGCM, UCLA CGCM GFDL AM3 and CM3). To account for observational uncertainty we use three different CloudSat+CALIPSO retrieval schemes for tropospheric ice water content and two very different methods to remove the contribution to ice mass from falling hydrometeors so that a robust estimate for CIWC and CIWP can be obtained for comparison to the GCMs. As an additional benchmarking measure, we include parallel evaluation of two recent reanalysis data sets (i.e. MERRA and ERA-Interim). ------- The results of the comparison show that even for annual mean CIWP, there are easily factors of 2-10 in the differences between observations and modeled values for a majority of the GCMs and for a number of regions. However, there are a number of CMIP5 models, including CNRM-CM5, MRI and CanESM2, as well as the UCLA CGCM, that perform relatively well compared to our past evaluations. Examination of the vertical structure indicates that most of the systematic biases in CIWC occur below the mid-troposphere where the models tend to overestimate CIWC, with this bias arising mostly from the mid and high latitudes. Disparities in the tropics also include significant differences in the level of maximum CIWC in the models, which ranges from about 250-550 hPa. Based on a number of diagnostics, our study finds the ensemble behavior of CMIP5 has improved considerably relative to CMIP3, including a 50% or more reduction in the multi-model mean bias and RMSE of CIWP. Despite this improvement, neither the CMIP5 ensemble mean nor any individual model demonstrates Taylor metric values within the range of observational uncertainty. In addition, there are still a number of models that still exhibit very large cloud ice biases despite the availability of relevant observations for a number of years now. While the CMIP5 models may be providing roughly the correct radiative energy budget, the results of our evaluation suggest that all, or at least many, of the models are accomplishing it by means of unrealistic cloud ice mass and structure, and thus are likely to have compensating errors in cloud cover, particle sizes, etc. This issue is discussed in more detail in the summary, along with caveats associated with the observed estimates, model and observation representations of both the precipitating and cloudy components of the ice, relevant physical processes and their parameterization, and the implications of these on the calculations of the radiation field. [Duane Waliser, USA]	
9-621	9	24	25	24	25	What do you mean with medium-term? [Farahnaz Khosrawi, Sweden]	Taken into account - This part of the sentence was removed.
9-622	9	24	27	24	29	If it is known that, due to significant errors in the model simulation of clouds, the results of climate projections will be of dubious validity, wouldn't it be more reasonable scientifically to stop the writing of the AR5 report there and avoid dubious projections ? [François GERVAIS, France]	Rejected - While the simulation of clouds is not perfect and while it affects the simlulations and projections, the uncertainty introduced by cloud feedbacks is well quantified (see Chapter 7) and smaller than the overall climate change signal.
9-623	9	24	31	24	42	This paragraph is unclear! What is your message? What are the consequences of the findings presented in the last part of this paragraph? [Martin Dameris, Germany]	Taken into account. The paragraph was a placeholder since ozone from CMIP5 models was not available at the time of the FOD. The SOD discusses CMIP5 ozone.
9-624	9	24	31	25	4	This section is completely un-understandable! Please rewrite! [Andreas Sterl, Netherlands]	Taken into account. The paragraph was a placeholder since ozone from CMIP5 models was not available at the time of the FOD. The SOD discusses CMIP5 ozone.
9-625	9	24	31	25	23	This discussion is too long in my opinion, see comment above. [Gunilla Svensson, Sweden]	Taken into account. Discussion on non-CMIP5 results shortened. This was a placeholder since ozone from the CMIP5 models was not available at the time of the FOD. . The SOD discusses CMIP5 ozone.
9-626	9	24	31			Section 9.4.1.1.3: This section seems a little out of place: it would help to know what the subset of CMIP5 models is that use model-generated interactive ozone compared with those that use forced ozone. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. The paragraph was a placeholder since ozone from CMIP5 models was not available at the time of the FOD. The SOD discusses CMIP5 ozone.
9-627	9	24	33	24	32	Though ODSs is an abbreviation commonly used I would rather skip it in this report since it is only used in a	Accepted. Abbreviation deleted.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						minority of sections. [Farahnaz Khosrawi, Sweden]	
9-628	9	24	36	24	38	This sentence seems incomplete and does not make sense. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Accepted. Text revided.
9-629	9	24	36	24	42	The scale of the anomalies in the right part of the plot (d and especially b) is too large. I understand that a clear inspection of the figure and a careful interpretation of the text will reveal the details, but I think that the aim to put a figure is to provide an immediate evidence of the concept mentioned in the text. Figure 9.7b, at a first view, does not show immediately the bias above Antarctica and South pole, giving thus the impression of a perfect model performance there. For the same reason, also the scale of the left part of the figure could be reduced, as the ozone values do not exceed 200 (min) and 340 (max) DU. [Claudio Cassardo, Italy]	Accepted. This figure has been removed from the SOD.
9-630	9	24	39	24	39	There is no panel (e,f) in the figure I think it will be (c,d) [HASIBUR RAHAMAN, India]	Editorial
9-631	9	24	43			I am missing a paragraph about water vapor! [Martin Dameris, Germany]	Noted. Stratospheric water vapour is discussed in Chapter 8 since it is important for RF. We refer to it in the revised version.
9-632	9	24	44			Figure 9.7 need to be clearer that a) and b) are strat ozone while c) and d) are tropospheric ozone [Bruce Wielicki, USA]	Figure has been removed.
9-633	9	24	51	25	5	It would be usefull to show the historical evolution of surface ozone in the models (e.g. figure from Lamarque et al. 2010). There have been many assessments of the performance of the suite of tropospheric chemistry models under present day conditions (e.g. Fiore et al. 2008). Arlene Fiore is also preparing a paper on the evaluation of the tropospheric ozone in the CMIP5 models as part of ACCMIP [William Collins, United Kingdom of Great Britain & Northern Ireland]	Rejected. We have not included a figure on surface ozone, but discuss tropospheric ozone and refer to Chapter 8 for details on radiative forcing agents. The focus of Chapter 9 is on evaluating CMIP5 models that are used in the later projection chapters.
9-634	9	25	2	25	5	slightly lower, broadly consistent, some deviations: These are very general statements which must be clarified; references are needed! [Martin Dameris, Germany]	Taken into account. The paragraph was a placeholder since ozone from CMIP5 models was not available at the time of the FOD. The SOD discusses CMIP5 ozone.
9-635	9	25	7	25	23	It would be good to update this with the latest AEROCOM results (instead of 2007 AEROCOM A/B). I think the Liu et al study is somewhat anecdotal. [Olivier Boucher, France]	Taken into account. However, this is the latest paper on the topic produced by AEROCOM in the published literature, verified by checking the up-to-date pubs list at the AEROCOM website and by forward citation search from Textor et al (2007). Reference/discussion of Liu et al has been dropped.
9-636	9	25	7	25	23	No aerosol distributions from CMIP5 are shown or discussed. It might be helpful to do this. The discussion focuses on reasons for intermodel differences in aerosol distribution, but no such differences are shown or reported. [Nathan Gillett, Canada]	Taken into account. Literature on CMIP5 aerosols as of drafting of SOD is still extremely sparse. Pointer to graphs in section 9.4.6 has been added.
9-637	9	25	7	25	23	Aerosols and clouds have their own chapter, still there are sections on aerosols and cloud evaluations here but they seem to be short (at least the aerosol one) and this section on aerosols is confusing and is not really about evaluation. [Gunilla Svensson, Sweden]	Taken into account -- Combined with comment 9-636
9-638	9	25	7			Section 9.4.1.1.4: I assume that more is to be added to this section in terms of actual results from the CMIP5 models? [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account -- combined with 9-636.
9-639	9	25	11		17	Reference is made to comparisons but no indication is given of differences: 10%?; factor of 2? The magnitudes of these differences need to be specified, and their implications. How much do these differences translate into forcings? It is not enough simply to say that the quantities have been compared. [Stephen E Schwartz, USA]	Taken into account -- combined with 9-636.
9-640	9	25	14	25	14	ensemble of chemical transport models [Graham Feingold, United States of America]	Take into account -- combined with 9-641.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-641	9	25	14	25	14	"an ensemble chemical transport models": You should decide for singular or plural. I guess you mean singular. It sounds like you are talking about one model and not several. [Farahnaz Khosrawi, Sweden]	Editorial -- text corrected.
9-642	9	25	17			Replace 'intermodal' with 'intermodel'. [Nathan Gillett, Canada]	Editorial -- text corrected.
9-643	9	25	25	25	25	Again, satellite observational based metrics especially the use of bi-variable or multi-variable metrics should be a valuable contribution to AR5. [Jonathan Jiang, United States of America]	Taken into account. Suggested literature now assessed.
9-644	9	25	25	25	35	Performance metrics can be constructed to quantify what models simulate well and contrarily to demonstrate model performance deficiencies... and Figure 9.8 In additional to the traditional metrics of temperature and precipitation Taylor diagrams that used in CMIP3, it will be good in this AR5 report to also include new metrics (and figures) such as those based on new A-Train satellite observation (e.g. Jiang et al. 2012, "Evaluation of Cloud and Water Vapor Simulations in IPCC AR5 Climate Models Using NASA 'A-Train' Satellite Observations," J. Geophys. Res. http://mls.jpl.nasa.gov/library/JiangEtAl_JGR_IPCC.pdf). [Hui Su, United States of America]	Noted - The evaluation of clouds has been given more prominence in the SOD and the suggested paper has been assessed and its conclusions included in that new discussion. The metrics proposed in the reference have seen far less uptake by the community so far than those used here. As a consequence, their use in this report is not justified.
9-645	9	25	25			Section: 9.4.1.2: There has been a comment in ACP by Grewe and Sausen (2009) on the Waugh and Eyring (2008) paper with some critique on the application and quantification of model performance using these matrices. From the text in th AR5 draft it does not become clear what exactly has been done and how far this critique has been considered. Reference: Grewe and Sausen (2009), Comment on "Quantitative performance metrics for stratospheric-resolving chemistry-climate models" by Waugh and Eyring, Atm. Chem. Phys., 9, 9101-91110. [Farahnaz Khosrawi, Sweden]	Taken into account. The global metrics assessed in this section are fundamentally different to those applied in either publication mentioned, which were soley based on an area averaged bias.
9-646	9	25	25			This section is great but seems odd to have inserted into only the atmospheric section; doesn't some of this applies to all of 9.4.1.2-5. [Duane Waliser, USA]	Noted. Metrics are discussed in other sections, but are introduced here as a reflection of the majority of the literature to assist.
9-647	9	25	27	25	35	The discussion in this section does not correspond to the figure. What observations have been used here? It is not clear how this example illustrates that circulation is better than other model fields (I do agree that this is the case but the example needs to show this more convincing. [Gunilla Svensson, Sweden]	Taken into account. The figure and corresponding text have been completely revised.
9-648	9	25	28	25	28	Figure 9.8 is for CMIP3 or CMIP5 ? [HASIBUR RAHAMAN, India]	Taken into account. Figure now shows results for CMIP3 and CMIP5.
9-649	9	25	28	25	28	Please state whether the pattern correlation in Fig 9.8 is centred or uncentred. [David Sexton, UK]	Editorial
9-650	9	25	31	25	31	.... for 200 hPa zonal winds ... Figure 9.8 shows 850 hPa zonal wind [Marco Giorgetta, Germany]	Editorial
9-651	9	25	31	25	31	The text refers to 200 hPa wind whereas the Figure 9.8 states that it shows 850 hPa wind. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Editorial
9-652	9	25	33	25	35	For precip, it is not only the greater impact of parameterisations that degrades its pattern correlation, but also its closer link to surface fields (topography, coastline, vegetation) that leads to much greater spatial heterogeniety at regional scales, making it both harder to model and harder to validate against point-based observations. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Taken into account. The figure and corresponding text have been completely revised.
9-653	9	25	38	25	41	In the figure captions it is mentioned "see Table [9.x] "but there is no Table [9.x] [HASIBUR RAHAMAN, India]	Taken into account. Table is now included.
9-654	9	25	45	25	46	What is RMSE? The abbreviation hasn't been used and explained before. [Farahnaz Khosrawi, Sweden]	Taken into account. RMSE definded in Glossary.
9-655	9	25	49	25	52	The text could be improved so that it will be easier to follow. E.g. the sentence "The results in this figures mixed". Sounds like a text fragment. Only when reading it a second time I could get what is meant (and that just an " are " is missing). [Farahnaz Khosrawi, Sweden]	Editorial. Text has been made more clear.
9-656	9	25	50	25	50	remove "The results in this figure mixed." or explain better …. [ANNALISA CHERCHI, Italy]	Editorial. Text has been made more clear.
9-657	9	25	50	25	50	Results in this figure are mixed [Graham Feingold, United States of America]	Editorial. Text has been made more clear.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-658	9	25	50	25	50	Something is lacking in the sentence: "The results in this figure mixed." [Celeste Saulo, Argentina]	Editorial. Text has been made more clear.
9-659	9	25	50			typo: "figure ARE mixed" [Ramon de Elia, Canada]	Editorial. Text has been made more clear.
9-660	9	25	50			The sentence on this line is ungrammatical. [Robert Pincus, USA]	Editorial. Text has been made more clear.
9-661	9	25	50			typo or grammar error [Duane Waliser, USA]	Editorial. Text has been made more clear.
9-662	9	25	52	25	54	See comment #2 above - choice of precipitation data set can be important. E.g. for model evaluation over the tropical Pacific ocean, models are generally dry compared with CMAP and wet compared with GPCP. See Irving, D. B., S. E. Perkins, J. R. Brown, A. Sen Gupta, A. F. Moise, B. F. Murphy, L. C. Muir, R. A. Colman, S. B. Power, F. P. Delage and J. N. Brown (2011), Evaluating global climate models for the Pacific island region, Climate Research, 49, 169-187, doi:10.3354/cr01028. for example over western tropical Pacific. [Josephine Brown, Australia]	Noted.
9-663	9	25			23	Line 23 ("concentrations differ by a factor of 3") is better (than lines 11-17): factor of 3; but still need to know how this translates into forcing. This factor of 3 is of course a mininum; it is the same model with different met drivers. One wants to know the overall uncertainty, in global mean forcing, and spatially. [Stephen E Schwartz, USA]	Taken into account -- Discussion from Liu (2007) paper cited on climatic implications of factor of 3 range among models has been added.
9-664	9	26	2		11	It seems a little odd to describe Eqn 1 of Annan and Hargreaves (2011) as an “alternate” to anything. It is simply an algebraic identity. Furthermore, the extension to the mean being better than any single model does not rely on the ensemble being statistically indistinguishable. [James Annan, Japan]	Noted: The text reflects that the Annan & Hargreaves (2011) results are an alternative interpretation of the utility of CMIP3 ensemble mean as a performance metric. The text also reflects the statement in the second paragraph of the conclusions that the CMIP3 is consistent with a statistically indistinguishable ensemble based on the leave-one-out validation procedure.
9-665	9	26	5	26	5	replace "is" in "now evident that is applies" with "it" or "this" [Celeste Saulo, Argentina]	Accepted. Text revised.
9-665	9	26	5	26	5	replace "is" in "now evident that is applies" with "it" or "this" [Celeste Saulo, Argentina]	Editorial
9-666	9	26	5			typo: "that IT aplies" [Ramon de Elia, Canada]	Accepted. Text revised.
9-666	9	26	5			typo: "that IT aplies" [Ramon de Elia, Canada]	Editorial
9-667	9	26	6	26	11	I think the statistically indistinguishable hypothesis explains why the multi-model mean matches observations better than a typical model, but I don't think it explains why the mean matches the observations better than any individual model. I think this may have something to do with the high dimensionality of the measures used to evaluate consistency. [Nathan Gillett, Canada]	Noted.
9-668	9	26	6			"the multi-model mean compares so well with observations" - at face value this reads like a close fit (for what observations?); I presume the qualitative finding is that multi-model mean in genreal compares better (or less worse…) with observed climate than individual model runs [Tor Eldevik, Norway]	Noted. Yes, the multi-model mean has a closer fit to observations given the set of measures used in the respective studies and as compared with single model results.
9-669	9	26	9	26	9	For what is this important? Is it important for that mean is agreeing good with the observations? [Farahnaz Khosrawi, Sweden]	Noted. Yes, when the multi-model mean agrees better with the observations, this supports the approach of using the multi-model ensemble as a better representation of past climate changes when simulating the 20th century.
9-670	9	26	13	26	13	fig 9.9 should have CMIP5 results … [ANNALISA CHERCHI, Italy]	Taken into account. Figure now shows results for CMIP5.
9-671	9	26	13			Figure 9.9: need to define acronyms used on the y-axis of this figure. otherwise very nice summary chart	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						[Bruce Wielicki, USA]	
9-672	9	26	22	26	22	sentence is incomplete. [Graham Feingold, United States of America]	Editorial
9-673	9	26	22	26	29	Citations that are embedded in the text are given in parantheses. [Farahnaz Khosrawi, Sweden]	Editorial
9-674	9	26	23	26	23	"(Gleckler et al.,2008;Yokoi et al.,2011) "should be "Gleckler et al.,2008;Yokoi et al.,2011" [HASIBUR RAHAMAN, India]	Editorial
9-675	9	26	25	26	25	"..metrics , (Yokoi et al.,2011) identify 7…." should be "...metrics , Yokoi et al.,2011 have identified 7…" [HASIBUR RAHAMAN, India]	Editorial
9-676	9	26	28	26	29	"(Reichler and Kim, 2008) "should be "Reichler and Kim, 2008" [HASIBUR RAHAMAN, India]	Editorial
9-677	9	26	28	26	33	In this chapter there needs to be a clear distinction between a single measure, generated by pooling a basket of metrics as in Reichler and Kim or Murphy et al (2004) and a single measure based on a single metric. On line 28 it makes Reichler and Kim's work sound like the latter. There is a big difference between the two as the risk of erroneously rewarding a poor model is much greater in the latter single metric because it is not hard to have a good score due to some fortuitous compensation of errors. For a basket of metrics this becomes much harder to do if the basket covers a range of variables. This chapter needs to make that distinction clear. Sexton et al (2011b) tests how a multivariate metric with increasing amounts of independent information can gradually lead to tighter constraints on a projection even when structural uncertainty is accounted for. REFERENCE D. M. H. Sexton and James M. Murphy Multivariate prediction using imperfect climate models part II: robustness of methodological choices and consequences for climate sensitivity Clim. Dyn. 2011 10.1007/s00382-011-1209-8 [David Sexton, UK]	Accepted - This point will be reflected in the discussion of results comparing multiple measures against results using single measures. The discussion of Reichler and Kim will be adjusted to reflect these distinctions by indicating a single skill score is constructed from multiple measures.
9-678	9	26	28	26	33	What's this para supposed to say? Using performing metrics is hopeless? How does this para relate tot rest of the section? [Andreas Sterl, Netherlands]	Taken into account. This section has been completely revised.
9-679	9	26	28			It may be useful to add a bridge sentence here to indicate that the next paragraph describes efforts to summarize many metrics for a given model. [Robert Pincus, USA]	Taken into account. This section has been completely revised.
9-680	9	26	32	26	32	"and it is unclear to": the word "unclear" is misleading here. It is unclear because, until now, nobody has developed a methodology able to objectively create such an index. [Claudio Cassardo, Italy]	Taken into account. This section has been completely revised.
9-681	9	26	44	26	9	Figure 9.10 ostensibly shows the same thing as Figure 1.4, but they don't appear to agree with each other. Figure 1.4 places the observed anomalies in the upper half of TAR model outputs and mid-range of AR4 model outputs, and this point is emphasized in the text and summary for the chapter. But in Fig 9.10 the black observational line now looks to be in the low end of the AR5 model range. The model outputs end at between 0.3 and 1.0, and the observations end at 0.5. But, the black line tends to bounce around from the top of the model range to the bottom over the 20th century. Depending on where the normalization is done, the observations could be made to appear to run above or below model projections in recent decades. That raises the question of how the data in Fig 9.10 were normalized. Were the TAR and AR4 data in Figure 1.4 also shifted to a zero mean over 1961-1990? Then combining the text in Chapter 1 (page 7) with Figure 9.10 we could now say that observations were in the upper half of model outputs in the TAR, the middle range in the AR4 and have now fallen to the lower half in the AR5. This point should be brought into chapter 1. On the other hand, if the relative location of the observational line as of 2010 is an artifact of different anomaly bases, then there should be some discussion of the sensitivity of the results to the application of the offsets. For example, if all series in Fig 9.10 were adjusted to a zero mean over 1900 to 1930 the observations would probably be over the top end of the models, and if they were adjusted to a zero mean over 1940 to 1970 the observations would be under the low end of the models. All these considerations make me think Figure 1.4 is likely not as reliable as it is made out to be, and in any case the two figures should be reconciled to tell a coherent story regarding where the observations lie in the model ranges from TAR through AR5. [Ross McKitrick, Canada]	Noted - Figure 9.10 shows the historical runs from CMIP5, while Figure 1.4 shows the results assessed in previous IPCC reports. Hence, the two figures could not possibly show the same model results. The captions of both figures state that both show anomalies with respect to the 1961-1990 mean.
9-682	9	26	44	26	9	It appears that, judging by the placement of the gray lines in Figure 9.10, the GCMs all anticipate the volcanic explosions, especially in 1908 and 1991, because the temperature drop starts a year or more before the	Noted. Figure has been completely revised

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						volcano. Maybe the gray lines need to be moved back. [Ross McKitrick, Canada]	
9-683	9	26	51	26	51	Figure 9.10 shows the anomalies in relation to the 1961-1990 period, thus artificially minimizing the ensemble spread for these years though the models not necessarily simulate these years in better agreement between themselves than for other years. To overcome this limitation it would be valuable to add a panel showing the full global 2m temperature in comparison to the full observations. This would reveal the evolution of the true spread of simulated global mean 2m temperature. [Marco Giorgetta, Germany]	Taken into account. Revised figure also depicts absolute spread between models and observations.
9-684	9	26	51	26	51	Same concern as in comment #40: in Chapter 2 the global temperature anomalies are from the HadCRUT4 dataset, and here you are using HadCRUTT3v. [Celeste Saulo, Argentina]	Taken into account. Observations are now consistent with Chapter 2
9-685	9	26	51	26	57	In Figure 9.10 the mulit-model ensemble mean is shown and it is commented that that contain much less variability than the observed which is natural since variability in the model might cancel each other. There should be a better way to represent the variability of the models, especially since CMIP5 contain ensembles of model runs for the historical run. [Gunilla Svensson, Sweden]	Taken into account. Only one realization is now shown for each model so there is no cancellation of variability.
9-686	9	26	52	26	53	Is this meant to be apparent from the plot? This seems pretty hand-wavy. I suggest this be quantified in some way, some reference be given, or else it be dropped. [Robert Colman, Australia]	Taken into account. The figure and corresponding text have been completely revised.
9-687	9	26	53	26	55	The statement "The gradual warming evident in the observational record, particularly in the more recent decades, is also evident in the simulations although again there are some important differences among models" does not come to grips whether the models in CMIP5 were selectively employing forcings that compensated differences in sensitivity, as has been found for the AR4 models. Rather than wait for a Kiehl or Knutti or Huybers to discover these correlations in the AR5 runs, the Assessment should examine for these. That would be the beginning of an Assessment. [Stephen E Schwartz, USA]	Taken into account. The role of external forcings is addressed in Chaper 10. The role of the IPCC is to assess the literature, not to discover new correlations.
9-688	9	26	53			Again on the statement, "The gradual warming evident in the observational record, particularly in the more recent decades, is also evident in the simulations although again there are some important differences among models." An assessment of consistency of current understanding might be gained by running each of the models over the twentieth century with a range of forcings representative of current understanding. It would seem possible to do this using the Greens function technique of Hansen (as described in Hansen et al ACP 2011). So there is little excuse for not doing this based on the amount of computer resources that would be required. [Stephen E Schwartz, USA]	Rejected. The focus of the model evaluation in this assessment is on simulations (notably CMIP5) that have already been performed. It is not the role of the IPCC to perform new simulations
9-689	9	26				Fig.9.9: I trust this will be updated to CMIP5 models, with separate columns for the CMIP3 and CMIP5 multi-model means. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Figure now shows results for CMIP5.
9-690	9	26				This section (and/or the discussion on salinity) should cross reference the chapter that includes the patterns of change of salinity as described in recent work by Durack and colleagues (I cant find it in chapt 3 but think it is in chapter 10 which I didn't download) - this work shows how salinity patterns have changed over 50 years (as I recall) consistent with the notion of wet getting wetter and dry drier wrt to preciptation and futher compares cmip3 model patterns of change - the point is salinity gives a much broader indicator of the change in preciptation over oceans over a multi-decadal period for which we have no other observational evidence. [Graeme Stephens, USA]	Noted. Cross reference to Chapter 10 is now made.
9-691	9	27	1	27	1	"particularly after the 1991 eruption of Pinatubo": also for the weak eruption of 1982 almost all models show a consistent cooling, and also for that of 1883 most models show a weaker cooling. The only eruption for which most models do not performed well is that of the 1903: in that occasion, the data show that the temperature was already in a phase of cooling before the eruption, cooling that lasted for two years after the eruption, while for most models it was smaller and almost instantaneous. [Claudio Cassardo, Italy]	Noted. Model evaluation of response to volcanic eruptions is problematic in the pre-satellite era.
9-692	9	27	3	27	3	Connect the Derbyshere reference to the other references given before. [Farahnaz Khosrawi, Sweden]	editorial
9-693	9	27	7	27	9	and this result confirms the results in AR4, right? [ANNALISA CHERCHI, Italy]	Taken into account - have rephrased statements in calibrated uncertainty language.
9-694	9	27	7		10	Is this some sort of statement in a box or does it belong elsewhere? In any case is the font and italics intended? [Larry Thomason, United States of America]	Taken into account - have rephrased statements in calibrated uncertainty language.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-695	9	27	7			models broadly capture? Please specify! [Martin Dameris, Germany]	Taken into account - have rephrased statements in calibrated uncertainty language.
9-696	9	27	7			Shouldn't there be a caveat somewhere in the section regarding the recent work from Yi Ming of GFDL indicating that the models are not doing well in the historical record at certain aspects of the aerosol response? [J. David Neelin, United States]	Taken into account - have rephrased statements in calibrated uncertainty language.
9-697	9	27	12	27	12	"annual mean global average anomaly" --> "annual-mean global-average anomaly" [Claudio Cassardo, Italy]	Editorial. Text has been made more clear.
9-698	9	27	12			In fig. 9.10, the multi-model mean stops at 2004. Since there is great interest in comparing observed and model trend for the last 2 decades, it would be very helpful to see the model results to at least 2020 (so that the model trend can be more accurately determined for comparison to the latest available observed data). [Stephen Gaalema, USA]	Rejected. The focus of this chapter is on model evaluation with the historical record. Future projections are discussed and assessed in Chapter 12
9-699	9	27	16	27	17	What is 'merged surface temperature'? Is this SAT over land and SST over ocean? I haven't seen this used elsewhere, but I agree it might be more directly comparable with obs. This deserves some discussion. [Nathan Gillett, Canada]	Taken into account. Merged surface temperature was also used in the AR4.
9-700	9	27	20	27	20	"……near surface specific humidity over land …" should be " near surface specific humidity from CMIP3 over land …" [HASIBUR RAHAMAN, India]	Editorial. Text has been made more clear.
9-701	9	27	25	27	26	"Given the sparse data network in the Southern Hemisphere this discrepancy may result from a combination of model errors and observational sampling uncertainty": a way to avoid such kind of error could be to sample the model output using only a selection of grid points corresponding to the station coordinates (latitude and longitude). [Claudio Cassardo, Italy]	This material has been removed because of space limitations.
9-702	9	27	28	9	44	I don't know how many times I have to say this, but Santer et al. did NOT examine the issue Douglass et al. examined. We answered the question "IF a climate model has a surface trend agreeing with the observed surface trend, would the climate model's upper air temperature trend agree with observations?" The answer was and remains a resounding NO! Indeed the "flaws" were committed by Santer et la.: a) not normalizing the models's surface trend to the observed for an apples to apples comparison, b) using a contaminated SST dataset (ERSSTv2,v3 rather than v3b), c) excluding land surface temperature to compare with full tropical atmosphere (a second apples to oranges comparison), d) not recognizing the spurious warm shift in two of the radiosonde datasets that had been pulished. See further analysis in Christy et al. 2010 and other papers to verify this. Even though McKitrick et al. did not do the Douglass et al comparison by normalizing models to observations at the surface, they at least acknowledged as much (footnote 5) that Douglass et al. performed a different analysis. Leaving a statement such as this in the IPCC will provide easy fodder to substantiate IPCC bias. Also McKitrick et al. indicated Santer et al. were in error - the key point of their paper, not a discussion of end-points. I would like to point out, and this is another issue about which justifiable criticism can be brought to the IPCC, is that the authors and co-authors of Santer et al. have several author positions in the IPCC menagery whereas authors and co-authors who have demonstrated Santer et al. is flawed are not represented in the CLA, LA or CA list here or in Chap 2. This is clearly not conducive to objective thought, and this section is not objectively written. This can become one of the more notorious examples of gatekeeping - where authors of one view have total control (eventually) in judging comments related to their own work vs. that of their critics (see author list of Chap 2). [John Christy, USA]	Taken into account. The two paragraphs ranging from line 27 to line 52 have been completely rewritten. We have enlisted as CAs two experienced researchers who are experts in atmospheric temperature observations and statistics in atmospheric science, respectively. Neither of these two CAs (Jörg Schulz, Andreas Hense) has been involved in any of the papers contributing to the debate. Editorial control of the SOD revised paragraphs rests directly with the CLAs. Our assessment now follows this logic: 1. Most climate model simulations have been found to warm more in the upper tropical atmosphere than is shown in observational estimates. 2. There has been an extensive debate in the published literature as to whether this difference is significant, given observational uncertainty and natural variability. 3. An explicitly contentious issue has been about how to take into account serial correlation. Our assessment furthermore takes into account that, in contrast to the approach in some papers, it is not permissible simply to average the different observational datasets as if they contained uncorrelated error. Nor does the standard error of a model ensemble mean properly reflect natural variability.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-703	9	27	28	27	44	which is the variable involved in this discussion? not clear … [ANNALISA CHERCHI, Italy]	Taken into account - indeed the issue is confusing, because different papers consider different variables.
9-704	9	27	28	27	44	Dealing with such heated debates in one paragraph is not easy. The authors did this reasonably well although they showed their preference by calling errors in Douglass et al 'serious flaws'. When you read the climategate emails and when you know how Douglass et al were treated in the peer review system (http://www.americanthinker.com/2009/12/a_climatology_conspiracy.html), these words are far too harsh. You can also say that Santer et al made serious flaws by not updating his analysis up to 2008, as McKitrick et al showed. Christy (2010) introduced a scaling ratio for the warming ratio of the troposphere compared to the surface. It would be informative to introduce that here as well as it reduces the influence of start and end dates. A scaling ratio of 1.4 is expected in the Tropics. Available datasets give a scaling ratio in the order of 0.8. This suggests GCM's still have great difficulties in accurately simulating the tropical troposphere. [Marcel Crok, The Netherlands]	Taken into account - see response to 9-702
9-705	9	27	28	27	44	The opening sentence is biased in several ways. First the findings of discrepancies between models and observations are described as "detectable" rather than "statistically significant", whereas the findings of studies claiming the discrepancies are small are described as "statistically insignificant." If you are going to place weight on statistical test results (as you should), then use the proper terminology for both types of studies. Second, the Santer et al. paper only examined data from 1979 to 1999, whereas the McKitrick et al. paper examined data up to 2009; and the Thorne paper's findings on the post 1979 interval were "On an individual pressure level basis, agreement between models, theory, and observations within the troposphere is uncertain over 1979 to 2003 and nonexistent above 300 hPa." -- hardly supportive of the assertion being made. Finally, Santer et al. is not the examplar of "all uncertainties being taken into account." While they made a small improvement on the methods in the Douglass et al. work, it was clearly explained in McKitrick et al. that the Santer method is deficient on serval counts, namely that the method is inaccurate outside the AR1 case, and inaccurate even in the AR1 case for comparisons across different data sets. Only McKitrick et al. 2010 can be said to have taken all uncertainties into account by using a bandwidth-free multivariate trend estimation framework robust to all forms of autocorrelation within and between the data panels. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-706	9	27	28	27	44	The paragraph on the tropical troposphere raises a disturbing philosophical point about the IPCC. Having read several chapters so far, whenever data appear to agree with GCMs and/or point towards confirmation of prior IPCC claims, the uncertainties are brushed aside, out comes the "virtually certain" language and the existence of one or two studies contradicting the main message is swept aside as unimportant. But with regard to the tropical troposphere, you have multiple data sets in agreement with each other but conflicting with GCM projections, and multiple studies from multiple teams showing the models significantly exaggerate warming, yet in this case out comes the "uncertainty" language and you claim to be unable to draw any conclusion at all because researchers have to make methodological choices, data sets all have end points, and so forth. This intellectual modesty would appear less opportunistic if it were applied everywhere, including in the observation and detection chapters. The difficulties in assembling tropical troposphere data products as described in this chapter are no more severe than those associated with assembling global SST or land air temperature records as described in Chapter 2, but there is no hesitation in those cases in using the resulting products to make very strong conclusions. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-707	9	27	28	27	44	My remaining points about the tropical troposphere make reference to the following studies, referenced above but repeated here for convenience. [Ross McKitrick, Canada]	Noted.
9-708	9	27	28	27	44	McKitrick, Ross, Stephen McIntyre and Chad Herman (2010) "Panel and Multivariate Methods for Tests of Trend Equivalence in Climate Data Sets". Atmospheric Science Letters, DOI: 10.1002/asl.290. (http://onlinelibrary.wiley.com/doi/10.1002/asl.290/abstract) [Ross McKitrick, Canada]	Taken into account.
9-709	9	27	28	27	44	R. McKitrick, Stephen McIntyre, Chad Herman, Corrigendum, Atmospheric Science Letters, 2011, 12, 4 [Ross McKitrick, Canada]	Taken into account.
9-710	9	27	28	27	44	McKitrick, Ross R. and Timothy Vogelsang (2011) "Multivariate trend comparisons between autocorrelated climate series with general trend regressors" University of Guelph Economics Department Discussion Paper	Taken into account.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						2011-09 (http://www.uoguelph.ca/economics/sites/uoguelph.ca.economics/files/2011-09.pdf) In preparation. [Ross McKitrick, Canada]	
9-711	9	27	28	27	44	The paper referenced in cell 55 examines the LT and MT layer records in from RICH and HadAT over 1958-2010, allowing for a shift in the mean in 1978 to account for the Pacific Climate Shift, something Thorne et al. (2011) did not do, though they alluded to it with reference to the Seidel and Lanzante paper. The paper applies a HAC estimator for multivariate trend models (ie comparisons across data sets) which is the most accurate way to do these tests. Without the 1978 shift term the observed trends are significantly higher than the average model, and with the 1978 shift term the difference is extremely significant, with the radiosonde trends falling below the entire range of model trends. In an extension which we are currently preparing for submission we allow the break date to be unknown, which changes the critical values, but the model-observation discrepancy remains massively significant. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-712	9	27	28	27	44	The text says: "First, there remain significant observational uncertainties (Chapter 2, Thorne et al., 2011; Mears et al., 2011)." But the uncertainties do not stand in the way of drawing conclusions. McKitrick et al. (2010, cell 53) tested the balloons and satellite records and found that the two observational systems agreed with each other but differed significantly from models. Also, you need to take note of the Corrigendum to MMH2010, referenced in Cell 54. We found an error in the GCM data supplied by GISS, and some of the observational data were revised after we did our analysis. We submitted corrected results which showed, among other things, that all balloon and satellite systems, individually and jointly, exhibit a significant difference from models at both the LT and MT layers for the 1979-2009 intervals. This includes the RSS product. So any observational uncertainties are immaterial to the key finding of a significant discrepancy between models and observations. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-713	9	27	28	27	44	The text says "Second, the choice of metric and statistical method have been shown to crucially affect the conclusions, as demonstrated by (Santer et al., 2008) who found severe flaws in the statistics used by (Douglass et al., 2008)." If you are going to indulge in trash-talking other researchers based on statistical methodology, how about you show the Santer method to a half dozen time series experts and record their responses in this paragraph. Santer et al. use a simple AR1 approximation developed in the 1930s before computers were available. It only applies to univariate series, it is invalid for higher order AR processes and it is invalid for comparing trends across different data series. While they made a slight improvement on the Douglass et al. method, it is just as "severely" flawed, not least by terminating the data set in 1999, the year after a massive El Nino. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-714	9	27	28	27	44	The choice of statistical method does not affect the conclusions nearly as much as truncating the data set.But even still, on data ending in 1999, MMH2011 (cell 54) show that the model-observational differences are marginally significant. And there is no reason to truncate the data at 1999. On the 1979-2009 sample the model-observational differences are statistically significant for all data sources individually and jointly using either a panel estimator or a multvariate HAC estimator. The Santer method can be shown to be a restricted version of the method used in MMH, where the restrictions can be rejected. To say that some well-established results are "crucially affected" by the decision not to use a restricted model, when the restrictions are known to be invalid, is grasping at straws. If you are going to hold that position then Santer's objections to the Douglass et al. model become invalid, since Douglass et al. is a restricted version of the Santer method. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-715	9	27	28	27	44	The text says "It has also been shown that the identification of trends in short records is severely affected by end-point issues (McKitrick et al., 2010; Santer et al., 2011; Thorne et al., 2011)." Again with "severe". If what you say is true then you can't rely on the Santer critique of Douglass since Santer only used data up to 1999, a short sample ending at an El Nino. But more generally, the model trend terms do not differ greatly depending on whether one uses a 1979-2009 sample or a 1958-2010 sample. The observed trends differ, but then what matters is whether one allows for a mean shift in 1977/78 when using the long sample. If the break is included then the trends don't change much, they remain small and insignificant. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-716	9	27	28	27	44	The text says "For instance McKitrick et al. (2010) found a strong dependence of their conclusions on record length.." This misrepresents our findings. We only included the short sample to show that Santer's results were sensitive to truncating the record. We didn't base our conclusions on the artificially-shortened sample, we	Taken into account - see response to 9-702

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						based our conclusions on the full sample length. But even still, in our corrected results (cell 54) we show that the model-obs discrepancy is weakly significant on the short sample and significant on the strong sample. This is not the kind of difference that your summary implies. [Ross McKitrick, Canada]	
9-717	9	27	28	27	44	"... and Thorne et al. (2011) found much better agreement between models and radiosonde observations when using the full radiosonde record instead of the shorter record that overlaps with satellite observations." The language here is biased. They didn't find "much better agreement," they found some improvement, but still noted strong inconsistency above 300 mb. And they don't report proper trend comparison scores anyway. McKitrick and Vogelsang (2011 - cell 55) computed proper multivariate trend comparison scores for the MSU-equivalent LT and MT layers over the 1958-2010 interval and show that the model-observation discrepancy is significant in a linear comparison, and if a break term is added in 1977 the discrepancy becomes extremely significant ($p<0.0001$). There is no possibility here of concluding that the models agree with the observations over the 1958-2010 interval or the post-1979 interval. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-718	9	27	28	27	44	" In addition there are uncertainties in how the models are forced, in particular in the recent studies that used scenario simulations to represent the last decade from model simulations (Fu et al., 2011; McKitrick et al., 2010)." This is clutching at straws. The forcings used in the models are listed in Table I of McKitrick et al. 2010--where is the uncertainty? Not all models use all forcings, but that doesn't undermine the findings in these papers. [Ross McKitrick, Canada]	Accepted - sentence deleted from rewrite.
9-719	9	27	28	27	44	The paragraph as written should be deleted and replaced with: "Several studies have focused on the ability of models to simulate observed trends in the free troposphere, in particular those in key tropical latitudes where GCMs all predict maximum warming responses to GHG's and other external forcings. Santer et. al. (2008) argued there was no statistically significant difference between models and observations over the 1979-1999 interval, attributing a contrasting result in Douglass et al. (2008) to a failure to correct for autocorrelation. However the Santer method is only valid in a lag-1 model and does not account for cross-panel covariance, hence is not valid for comparisons across different data sets (McKitrick et al. 2010). McKitrick et al. used a multivariate heteroskedasticity and autocorrelation (HAC) robust estimator and showed that the model-observation discrepancy was marginally significant over the 1979-1999 interval and highly significant over the 1979-2009 interval. Evidence of significant discrepancies have also been reported in Christy et al. (2010) Bengtsson and Hodges (2011) and Fu et al. (2011). Thorne et al. (2011) reported that the discrepancies between models and observations were less severe below 300 mb on a 1958-2003 sample. Using a multivariate HAC estimator, McKitrick and Vogelsang have shown that the MSU-equivalent LT and MT layer radiosonde series over 1958-2010 exhibit significant differences with models, and if a step-change is used to account for the effect of the 1977 Pacific Climate Shift the differences are extremely significant ($p<0.0001$). Hence the finding of a discrepancy between models and observations in the tropical troposphere, in which models significantly overestimate warming by a factor of 2—4 times, has been established on multiple time scales, and on multiple data sets, using methods robust to the relevant uncertainties. [Ross McKitrick, Canada]	Rejected as a proposal to accept this rewrite - see, however, response to 9-702.
9-720	9	27	29	27	45	More quantified information would be helpful in this section, or possibly a figure could be added. The unfamiliar reader would have trouble understanding the scale and range of the trends, and the degree of dispute between the different studies. [David Bader, USA]	Taken into account. See response to 9-702; will include a figure if it is compelling enough.
9-721	9	27	33	27	34	Why is it harder to evaluate models' ability to simulate tropical upper tropospheric trends? As written the text suggests that the uncertainties are larger here. But I don't think this is the case. The issue is that there is apparent disagreement here, with the models warming more than the obs. The discussion in the literature is over whether or not the difference is within the bounds of internal variability plus observational uncertainty. I think this paragraph could be written a bit more straightforwadly and say that the models generally warm more than the obs in the tropical upper troposphere, though there is debate about whether the difference is significant in the light of observational uncertainty. As it is the apparent disagreement is noted briefly and then the rest of the discussion focuses on uncertainties. In other regions where models and obs agree there is much less discussion of uncertainties. [Nathan Gillett, Canada]	Accepted - see response to 9-702, which builds on this comment to quite some extent.
9-722	9	27	37	27	37	This is an assessment, and since this paper (Douglass) has been shown to be flawed, it should not be given credence (e.g. by citing it earlier). [Robert Colman, Australia]	Accepted - also see response to 9-702.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-723	9	27	37	27	37	Santer et al. (2008) instead of (Santer et al, 2008). [Farahnaz Khosrawi, Sweden]	Editorial
9-724	9	27	37	27	37	"…by (Douglass et al., 2008)" --> "…by Douglass et al. (2008)" [Hai Lin, Canada]	Editorial
9-725	9	27	42	27	44	But have any specific problems been found associated with merging the historical and SRES simulations? Ben Santer has done this in a recent paper, and generally the models do not exhibit forcing discontinuities. As it is this statement suggests there might be problems without providing evidence that there are. [Nathan Gillett, Canada]	Accepted - sentence deleted from rewrite.
9-726	9	27	46	27	52	this paragraph contains results that are not in agreement, which is the truth? [ANNALISA CHERCHI, Italy]	Taken into account - see response to 9-702
9-727	9	27	46	27	52	This paragraph relies on the Thorne graphs. But they point out that there was no warming in the pre-78 interval, and their Figure 13 shows that there is poor agreement in the amplification rates in the post-79 interval, so if the step-change at 1977 were controlled for, it is likely their results would change considerably. Consequently it is unwise to base any conclusions on the Thorne analysis of the 1958-2003 sample. [Ross McKitrick, Canada]	Taken into account - see response to 9-702
9-728	9	27	50	27	50	While there are discrepencies…' : This is much too strong a statement given the discussion in the preceeding paragraphs. A number of studies (e.g. Santer, Thorne) find no discrepency, and it is apparent that methodological problems cloud many of those that do. So this statement should be much more guarded, e.g. 'While it remains possible...' or even conclude that 'it is very far from clear that any such discrepency exists.' [Robert Colman, Australia]	Taken into account - see response to 9-702
9-729	9	27	54			The work of Randel et al. (2009) must be cited here: Randel, W. J., et al. (2009), An update of observed stratospheric temperature trends, J. Geophys. Res., 114, D02107, doi:10.1029/2008JD010421. [Martin Dameris, Germany]	Accepted. Reference added.
9-730	9	27	55	28	14	Surely this is a place to include some mention of the high-top models in CMIP5 and whether they benefit the simulation of stratospheric temperature, especially given the suggested link with the QBO and the Brewer Dobson circulation. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account and some discussion high-top low-top models mentioned.
9-731	9	27	56	28	1	The reference to Figure 2.16 is incorrect. I could not find which figure in Chapter 2 are you referring to. [Celeste Saulo, Argentina]	Editorial
9-732	9	28	1	28	1	I think figure 2.16 is wrongly quoted it may be figure 2.12. It will be good if Ch-2 also added here and else where when refering other chapters figures. [HASIBUR RAHAMAN, India]	Editorial
9-733	9	28	8	28	8	Another reference for the model overestimate of volcanic warming in the stratosphere is Free, Melissa and J. R. Lanzante, 2009: Effect of Volcanic Eruptions on the Vertical Temperature Profile in Radiosonde Data and Climate Models. Journal of Climate 22, 2925-39, DOI: 10.1175/2008JCLI2562.1 [Melissa Free, USA]	Accepted. Reference added.
9-734	9	28	11			Figure 9-10 shows the evolution of temperature anomalies. It would be useful to also show the time series of absolute temperature values - when I have seen this presented the impression of agreement is less marked. [Robert Pincus, USA]	Taken into account. Revised figure also shows absolute values.
9-735	9	28	13	28	14	This paragraph is not concerned with metrics related to the co-variability of variables, but rather the co-variability of trends. The first sentence should make this distinction clear. [Robert Pincus, USA]	Taken into account. Text is revised to be more clear
9-736	9	28	17	28	36	I do not know how much value this figure will add on the text, as the studied period is short and the values of models and data refer to different periods and thus are not completely comparable. [Claudio Cassardo, Italy]	Taken into account. The figure now shows results for CMIP5, with a condiderably longer record and a consistent period between models and observations.
9-737	9	28	19	28	31	Explain that W stands for precipitable water [Philip Mote, USA]	Taken into account. W is now clarified.
9-738	9	28	22	28	22	TO natural variability [Robert Colman, Australia]	Editorial
9-739	9	28	22	28	22	"…sensitive natural variability" --> "…sensitive to natural variability" [Hai Lin, Canada]	Editorial

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-740	9	28	22			typo: "sensitive TO natural" [Ramon de Elia, Canada]	Editorial
9-741	9	28	23	28	23	scaling ratio?' [Robert Colman, Australia]	Taken into account. Sentence rephrased.
9-742	9	28	25	28	27	It is not clear from Figure 9.11 that the scaling ratio as simulated by models is consistent with that of the observations. This is demonstrated in the Mears et al. (2007) paper but is not shown by this plot. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Editorial. Text has been made more clear and consistent with figure.
9-743	9	28	27			Figure 9.11: Caption and axis labels appear to differ on what exactly is being plotted here. [James Christian, Canada]	Editorial. Text has been made more clear and consistent with figure.
9-744	9	28	28	28	28	"Scatter plot of the variability of W as a function of the trend in W as a function of the TLT trend": this sentence seems confuse. I suggest to modify in: "Scatter plot of the trend in tropical precipitable water (W) as a function of the lower tropospheric temperaure (TLT) trend". [Claudio Cassardo, Italy]	Editorial. Text has been made more clear and consistent with figure.
9-745	9	28	28	28	28	Define W, TLT, UAH [Robert Colman, Australia]	Editorial. Text has been made more clear and consistent with figure.
9-746	9	28	28	28	36	Fig 9.11 uses a very obsolute version of UAH data. We are at v5.4 and have been for a good while. Why use 5.1 and 5.2? This is not useful. Indeed v5.1 was obsolete in AR4. [John Christy, USA]	Taken into account. Revised figure uses updated data.
9-747	9	28	28			Figure 9.11 caption: This figure caption is completely ununderstandable with all the abbreviations used that have not been introduced. What is W? What is UAH and what is TLT? [Farahnaz Khosrawi, Sweden]	Editorial. Text has been made more clear and consistent with figure.
9-748	9	28	29			Figure 3a ???? Does not exist in this report. [Andreas Sterl, Netherlands]	Editorial. Text has been corrected.
9-749	9	28	31	28	31	Figure 9.11 It should be made more explicit what is W and TLT. [Annarita Mariotti, U.S.A.]	Editorial. Text has been made more clear and consistent with figure.
9-750	9	28	31	28	31	Fig. caption: I think that it should be "Scatter plot of the trend in W as a function…" [Celeste Saulo, Argentina]	Editorial. Text has been made more clear and consistent with figure.
9-751	9	28	36	28	38	This is a weak sentence; could restate as "the plot illustrates the physical connection between the two quantities, and consequently the fact that different periods of record are used for climate model and reanalysis (1981-99) and satellite records (1988-2006) affects only the actual value of the trends, not the strength of the conclusion about the physical connection." or something along those lines. [Philip Mote, USA]	Editorial. Text has been made more clear and consistent with figure.
9-752	9	28	41	29	3	Again there is an opportunity here to discuss the high-top models in CMIP5 and whether they provide a better simulation of the influence of the changes in stratospheric ozone than the low-top models. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. High-top models discussed in the context of stratospheric temperature trends
9-753	9	28	41	29	3	Again a very long section describing CCMVal results… [Gunilla Svensson, Sweden]	Taken into account. The paragraph was a placeholder since results for ozone from CMIP5 models were not available at the time of the FOD.
9-754	9	28	46	28	47	vers is missing [ANNALISA CHERCHI, Italy]	Editorial
9-755	9	28	46	28	47	It is worth noting that the models without ozone changes exhibited weaker trends. [Nathan Gillett, Canada]	Taken into account. Discussion extended.
9-756	9	28	48	28	50	Even if climate models prescribe time-varying ozone, they normally prescribe zonal mean values and ignore the zonal asymmetric component (as it was in the CMIP3 models). There is a number of publications showing that the influence of the zonally asymmetric component of ozone field is not negligdible. I suggest mentioning this. Relevant references: (1) Crook, J. A., N. P. Gillett, and S. P. E. Keeley (2008), Sensitivity of Southern Hemisphere climate to zonal asymmetry in ozone, Geophys. Res. Lett., 35, L07806, doi:10.1029/2007GL032698. (2) Waugh, D. W., L. Oman, P. A. Newman, R. S. Stolarski, S. Pawson, J. E. Nielsen, and J. Perlwitz (2009), Effect of zonal asymmetries in stratospheric ozone on simulated Southern Hemisphere climate trends, Geophys. Res. Lett., 36, L18701, doi:10.1029/2009GL040419. [Alexey Karpechko, Finland]	Accepted. Reference added.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-757	9	28	49	28	50	a typo in these two lines of text: "however, in a subset .... than a time varying field." doesn't appear to make any sense. [Bruce Wielicki, USA]	Editorial
9-758	9	28	53	29	7	In the text pag. 28 line 53, or in the legenda of Fig. 9.12, a link to section 9.4.1.1.3 could be included in order to remember to the reader why the two databases do not show the same values of data. [Claudio Cassardo, Italy]	Noted. The section has been revised and ozone (mean and trends) is discussed in one paragraph in the SOD.
9-759	9	28	55	28	55	Skip the abbreviation SAM since it is used only here and not further in the text. [Farahnaz Khosrawi, Sweden]	noted
9-760	9	29	9	28	9	I would suggest to skip the question in the section and title and just name it: "Model-Data comparisons for the Last Glacial Maximum, the Mid-Holocene and the Last Millenium" and maybe add "and their impact on present day and future simulations" if you want to point out their impact on these simulations. [Farahnaz Khosrawi, Sweden]	accepted
9-761	9	29	12	29	12	"The Last Glacial Maximum (LGM) and mid-Holocene": I suggest to insert here the years of the simulated period. [Claudio Cassardo, Italy]	accepted
9-762	9	29	15	29	20	I am not fond of the abbreviations used here, like LGM, MWP, LIA. I don't see anything useful for applying them. It only makes the text more difficult to read. [Farahnaz Khosrawi, Sweden]	noted. Note that the millennium has been removed from this section. Reference for it is made to chapter 5
9-763	9	29	22	29	25	Shoud cite here Mann et al (2009) [Mann, M.E., Zhang Z., Rutherford, S., Bradley, R.S., Hughes, M.K., Shindell, D., Ammann, C., Falugevi, G., Ni, F., Global Signatures and Dynamical Origins of the “Little Ice Age” and “Medieval Climate Anomaly”, Science, 326, 1256-1260, 2009] assesses evidence from proxy reconstructions regarding the MCA-LIA transition, assessing the role of natural radiative forcings with regard to the observed spatial features and apparent dynamical responses. [Michael Mann, USA]	taken into accound. Note that the millennium has been removed from this section, and reference is made to chapter 5.
9-764	9	29	23	29	23	Use the more recently preferred terminology "Medieval Climate Anomaly (MCA)" in place of outdated (and potentially misleading) "MWP" moniker. [Michael Mann, USA]	taken into accound. Note that the millennium has been removed from this section, and reference is made to chapter 5
9-765	9	29	24	29	27	I didn't find the results presented in Fig 9.13 provided very strong evidence of model fidelity. What is the correlation between the patterns of air temperature change in each era in models with the obs? Centred and un-centred correlations could be calculated to evaluate whther the models have skill at simulating the mean change and the spatial pattern. In particular, the Holocene results appear to not be very similar bewteen model and reconstructions. [Nathan Gillett, Canada]	taken into account. The figure and the text have been revised and a new figure has been added to show quantitative comparisons
9-766	9	29	24	29	47	In this part there is a detailed interpretation of Fig. 9.13a, but I cannot find any comment about Fig. 9.13b, e.g. on precipitation differences. Looking on the map, these are sometimes very large, not only in the tropical areas but also over orth America, Europe, some Asian regione and few African areas. In my opinion, these differences deserve an intepretative explanation to justify the assertion about the good performance of the models, that otherwise could appear questionable. [Claudio Cassardo, Italy]	taken into account. The figure and the text have been revised and show quantitative comparisons
9-767	9	29	24	29	58	The description on lines 24-47 does not give me the impression that “These analyses all show that [...] models can reproduce large scale features ....”, nor does Figure 9.13 do. The figure shows that during the LGM it was colder than today nearly everywhere. This is reproduced i the models. The models also show patterns (larger cooling in the North than in the tropics), but they are not evident in the obs. On the contrary, the obs show a warming near Greenland, which is completely absent in the models. Likewise, obs show decreased rainfall over Europe in 6 ka, a signal which is not reproduced by the models. Increasing precip in the obs over eastern America is not present in the models, who show more precip over the Gulf Stream. [Andreas Sterl, Netherlands]	taken into account. The figure and the text have been revised and show quantitative comparisons
9-768	9	29	24		52	9.4.1 is about the atmosphere but this paragraph discusses a mixture of atmosphere and ocean results. This makes it a slightly confusing read. However, I would say the mistake is in making separate sections for atmosphere and ocean, as there is surely a place for discussing both together. [Julia Hargreaves, Japan]	taken into account. However this is the only place where this can be discussed
9-769	9	29	32		33	While it is implicitly understood for those who know what MARGO is, I wonder if there is room to add the caveat that we are talking about reliability on large scales and for steady state climate. [Julia Hargreaves, Japan]	taken into account.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-770	9	29	35	29	36	This wasn't clear to me from the figure. [Nathan Gillett, Canada]	noted
9-771	9	29	49	29	49	Why is CMIP3 mentioned here? It hadn't been discussed in this section (9.4.1.5). Are you refering here to a previous section? [Farahnaz Khosrawi, Sweden]	editorial - typo error
9-772	9	29	49	29	51	The inference in lines 50-51 is quite strong given the evidence supporting lines 49-50. It would be sufficient to point out that climate models are able to simulate very different climates than prevail today. [Robert Pincus, USA]	taken into account. The section has been revised to provide an assessment from quantitative model-data comparisons
9-773	9	29	52	29	55	"suggesting that they are likely to properly project future climate change" appears much too strong to this reviewer from the crude and very qualitative argument being presented [Tor Eldevik, Norway]	taken into account. The section has been revised to provide an assessment from quantitative model-data comparisons
9-774	9	29	53			I don't think this necessarily suggests that climate models are likely to "properly project future climate change". Instead, passing these tests simply strengthens our confidence in the models. [Adam Scaife, United Kingdom of Great Britain & Northern Ireland]	taken into account. The section has been revised to provide an assessment from quantitative model-data comparisons
9-775	9	29	55	29	58	Figure 9.13 caption: You should decide if you want to use the abbreviation LGM or not, then you want to use it you should do it consequently. In this line LGM should be added in parantheses. However, I would rather skip the usage of this abbreviation. [Farahnaz Khosrawi, Sweden]	editorial. Figure and caption have been revised
9-776	9	29	57			Figure 9.13: very hard to compare the colors in the dots on the two left side figures to the color maps. make dots larger? [Bruce Wielicki, USA]	Taken into account. Figure has been revised
9-777	9	30	5			Section 9.4.2: This section is only describing the results, but rarely giving some explanations where the differences are coming from. [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified accordingly, based on literature results
9-778	9	30	9	30	9	The meaning of the abbreviation ENSO has still not been introduced. [Farahnaz Khosrawi, Sweden]	Rejected. Term defined in Glossary
9-779	9	30	14	30	14	this section (9.4.2.1) should be adjusted to CMIP5 results, at least the first 2 paragraphs [ANNALISA CHERCHI, Italy]	Taken into account. Text modified accordingly, based on literature results
9-780	9	30	16	30	16	Salinity is not represented in Fig. 9.14 [Celeste Saulo, Argentina]	Taken into account. Figure modified accordingly
9-781	9	30	16	30	23	Why are these temperature and salinity important? [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified accordingly.
9-782	9	30	18	30	19	Is this warm bias at 200-3000m because the NADW is too warm? Is this because it forms in too warm locations? [Nathan Gillett, Canada]	noted: No literature could be found to provide an explanation that is not model dependent
9-783	9	30	18	30	22	In addition to the comments in the text, I notice that near Antarctica also the surface water is warm. The maximum in the NADW is large but in a very narrow area, and considering that the NADW occupies only a small portion of the Atlantic ocean, this may mean that, locally, it may be quite larger. The minimum of the cold area (or maximum cold bias) seems to me located not "near the surface at mid-latitudes of the NH" but in a very narrow zone between 200 and 300 m of depth at high latitudes (near 70°) of the NH. The presence of such so narrow areas of min and max at similar latitudes, but different depths, of the NH may suggest that the model does not simulate well the NADW formation. The smaller bias present near Antarctica suggests that the model has less difficulties in simulating the AADW formation. [Claudio Cassardo, Italy]	noted: No literature could be found to provide an explanation that is not model dependent
9-784	9	30	20	30	20	Omit the abbreviation NADW. [Farahnaz Khosrawi, Sweden]	Rejected. Term defined in Glossary
9-785	9	30	25			Figure 9.14 caption: 2004 should be 2005 [James Christian, Canada]	Taken into account. Caption modified.
9-786	9	30	33	30	34	"the last decade has seen important and substantial improvements in the development of global salinity observations, such as those from the ARGO network" surface salinity is now measured from space (starting Sept 2011, see aquarius.nasa.gov) [James Christian, Canada]	Taken into account. Text modified accordingly.
9-787	9	30	34	30	34	What is the ARGO network? [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified accordingly.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-788	9	30	35			Argo Steering => Argo Steering Team [Andreas Sterl, Netherlands]	Taken into account. Text modified accordingly.
9-789	9	30	36	30	36	I'm not sure that the evaluation of the global surface air temperature in section 9.4.1.1 was sufficient for proper evaluation of sea surface temperatures, whose errors can be fundamental to the overall simulation of global climate and its variability. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Text modified accordingly to better assess SST and figure provided
9-790	9	30	36			SSS - maybe add the recent capability to observe SSS from satellite [Tor Eldevik, Norway]	Taken into account. Text modified accordingly.
9-791	9	30	41			I had trouble finding much on SST simulation in 9.4.1.1 other than air temperature at 2m although I apologise if I missed it. While in some regions of the globe these two are synonymous, in other regions such as the North Atlantic they are not. I think a section on SST is needed to describe successes and failures of climate models. For example a very large cold bias occurs in the N Atlantic in many models (e.g. Keeley et al 2012, submitted to QJRMS). [Adam Scaife, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Text modified accordingly to better assess SST and figure provided
9-792	9	30	43	30	44	"The most systematic biases include a saline bias in the tropical Pacific (Delcroix et al., 2010) and Bay of Bengal" I think this should state eastern or western tropical Pacific since there is a very strong E-W salinity gradient. Also it is quite possible that the problem in the BoB does not originate in the ocean model since this is a runoff-dominated system (although some coarse-resolution models do have a lot of trouble simulating the circulation of the Northern Indian Ocean). [James Christian, Canada]	Taken into account. Text updated to clarify statement
9-793	9	30	46			FYI - Salinity article under revision. [Duane Waliser, USA]	Taken into account. Reference cited
9-794	9	30	47	30	47	The abbreviation ITCZ has not been introduced yet. [Farahnaz Khosrawi, Sweden]	Rejected. Term defined in Glossary
9-795	9	30	57	30	57	Write " (de Jong et al. 2009, Sloyan and Kamemkovicj, 2007). [Juan A. Blanco, Canada]	Editorial. Corrected
9-796	9	30	57	30	57	."..to some extent (de Jong et al., 2009) and (Sloyan and Kamenkovich, 2007) " should be "…..to some extent (de Jong et al., 2009;Sloyan and Kamenkovich, 2007)" [HASIBUR RAHAMAN, India]	Editorial. Corrected
9-797	9	30	58	31	4	Are the abbreviations SAMW, AAIW, SPMW, STMW really useful? As stated before there are definitely too many abbreviations introduced and used in this chapter. [Farahnaz Khosrawi, Sweden]	These are standard abbreviations for oceanographers and are useful for them.
9-798	9	31	5	31	6	"relative to the Gulf Stream and northwest Sargasso Sea" vague. What is the definition of the "northwest Sargasso Sea"? [James Christian, Canada]	Taken into account. Text clarified
9-799	9	31	8	31	8	"imply variation in ocean heat storage and advection" vague [James Christian, Canada]	Taken into account. Text clarified
9-800	9	31	19	31	19	Same as my previous comments. Is using the abbreviation SSH really useful? [Farahnaz Khosrawi, Sweden]	These are standard abbreviations for oceanographers and are useful for them.
9-801	9	31	19			How were these diagnosed? Is this based on wind and ocean denisty? [Nathan Gillett, Canada]	Taken into account. Caption modified for more clarity
9-802	9	31	45	31	45	Include "are" before nearly so that the text reads "are nearly linearly" [Farahnaz Khosrawi, Sweden]	Taken into account. Text clarified
9-803	9	31	45			Scale nearly linearly with what? [Nathan Gillett, Canada]	Taken into account. Text clarified
9-804	9	31	55			On average the CMIP3 models appear to warm more strongly thn the observations? Is this true for CMIP5? Why? [Nathan Gillett, Canada]	The CMIP3 models without volcanic forcings warm more strongly than observational estimates, but that is not the case for models that include volcinic forcings. Proper assessment of the CMIP5 models will require careful drift removal which will be resolved by the final draft.
9-805	9	31	58			Figure 9.17: need thicker legend lines and text: very difficult to relate to plot lines [Bruce Wielicki, USA]	Taken into account. Figure has been improved.
9-806	9	31				Fig 9.17. It would appear that the differences between modeled ocean heat uptake and measurements need to be discussed. It would seem that the derivative of ocean heat content with time (flux imbalance) varies over quite a range; suggest this be discussed. I see that the model results presented in the figure are from CMIP3,	Rejected. Lucarini et al. (2011) is an important paper but it is about SRES simulations in CMIP3, not model evaluation base on observatons.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						with intent to update. Lucarini and Ragone have called attn to major imbalances in preindustrial CMIP3 runs that on face would raise question with the ability of these models to yield an energy imbalance. Lucarini V, Ragone F (2011) Energetics of Climate Models: Net Energy Balance and Meridional Enthalpy Transport. Rev Geophys 49:RG1001. It would seem that the concerns raised in that paper need to be addressed here. If the figure is replaced by CMIP5 results, it would seem that an analysis such as that of Lucarini needs to be carried out on the new results before the modeled imbalance can be reported with confidence. [Stephen E Schwartz, USA]	
9-807	9	32	6	32	25	A figure illustrating the variability of the AMOC simulated by CMIP3 and CMIP5 models seems essential for the rest of the report. [Thierry Fichefet, Belgium]	Taken into account - we investigated the feasibility of including an assessment of AMOC variability in CMIP3 and CMIP5. based on the availability of published literature. The late availability of CMIP5 did not sllow this for the SOD but will be considered for the final version.
9-808	9	32	18			I am not sure that a sverdrup (Sv or million cubic meters/second) is an appropriate unit since it is not an SI unit (Sv would be a Sievert) and I, for one, had to look it up to have any idea of what this unit was. How about Gl/s? [Larry Thomason, United States of America]	Rejected. This is a standard unit for oceanographers.
9-809	9	32	23	32	23	Replace " a few years ago" with "in 2004" [Chris Roberts, Uk]	Taken into account. Section modified.
9-810	9	32	24	32	25	how long is "the observational record"? [ANNALISA CHERCHI, Italy]	Taken into account. Replaced by "The half-decacde long observational record"
9-811	9	32	24			what does the correlation mean? And what do the other EMICs say? If it is only one EMIC that show a correlation, it is possibly no tsignificant. [Andreas Sterl, Netherlands]	Taken into account. The correlation means that models having a strong overturning in the control climate tend to show the large AMOC reductions. A reference to the CMIP model simulations is included, and the sentence reworded as follows: The ability of models to simulate this important circulation feature is tied to the credibility of simulated AMOC weakening during the 21st century because, the strength of the weakening is correlated with initial AMOC strength (Gregory et al., 2005). Gregory, J. M., , and Coauthors, 2005: A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO2 concentration. Geophys. Res. Lett., 32, L12703, doi:10.1029/2005GL023209.
9-812	9	32	25	32	28	If this statement specifically relates to one specific EMIC only, then it is really more a placeholder for a particular reference than the general statement about climate models that it is dressed up like [Tor Eldevik, Norway]	Taken into account. The sentence is now reworded and the reference replaced by a CMIP reference. In contrast to the descriptive paper which is cited now, a physical mechanism is suggested in the EMIC paper.
9-813	9	32	28	32	29	"leads to western boundary currents that are too weak and diffuse" and in the wrong place. Low resolution models generally do not accurately represent the mean position of WBCs. This is discussed in Kwon et al 2010. [James Christian, Canada]	Taken into account. Study will be cited
9-814	9	32	29	32	29	Include "leads to" after hence so that it reads "leads to biases". [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-815	9	32	33	32	33	It has to be explained clearly that the horizontal resolution given in the brackets is just for zonal (west-east) direction, not for meridional direction. [Zhaomin Wang, UK]	Taken into account. Text clarified

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-816	9	32	33	32	35	Are these covariances likely to be realistic in the reanalyses? Or is this just a case of comparing one set of models with another? [Nathan Gillett, Canada]	Taken into account. Text updated to clarify statement
9-817	9	32	46	32	46	In a recently submitted paper, it has been demonstrated that the most important factor is eddy-induced thickness diffusivity used in eddy parameterization in IPCC AR4 models (T. Kuhlbrodt, R. Smith, Z. Wang, J. M. Gregory, The influence of eddy parameterizations on the strength of the Antarctic Circumpolar Current in coupled climate models, submitted to Ocean Modelling, in revision now) [Zhaomin Wang, UK]	Taken into account. Study will be cited
9-818	9	32	47	32	47	"the salinity gradient across the ACC down through the water column" vague [James Christian, Canada]	Taken into account. Re-phrased as "the change in salinity with depth across the ACC"
9-819	9	32	52	32	52	"Although the models generally capture a strong circumpolar circulation" I don't think this is true. The total transport of the ACC is something that the CMIP3 models generally did not capture well. Some overestimated it by more than 2X and others underestimated it by more than 2X. Only a handful of models were within +/- 25% of the observational estimate and all but one of these underestimated it (Russell et al 2006; their Table 1). [James Christian, Canada]	Taken into account. Text modified accordingly.
9-820	9	32	52	32	54	An evalution of simulated subpolar gyres in IPCC AR4 models are explicitly given in Wang and Meredith (2008) (Wang, Z., and M. P. Meredith (2008), Density-driven Southern Hemisphere subpolar gyres in coupled climate models, Geophys. Res. Lett., 35, L14608, doi:10.1029/2008GL034344.) [Zhaomin Wang, UK]	Taken into account. Study will be cited
9-821	9	33	3	33	3	"Sueyoshi and Yasuda (2009)" instead of "(Sueyoshi and Yasuda, 2009). [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-822	9	33	3	33	4	"For fifteen models out of twenty, the average radius for the mid-latitude bands and the phase speed of long baroclinic Rossby waves is (-->are) underestimated": this sentence does not give information, as firstly does not say what happens to the other models (overestimation?), and secondly the underestimation, in itself, is not a failure if it is not specified a term of comparison. [Claudio Cassardo, Italy]	Taken into account. Text clarified
9-823	9	33	8	33	8	"the last glacial ocean circulation" last glacial maximum? [James Christian, Canada]	editoral. Title revised
9-824	9	33	8			“last glacial ocean circulation”=> “ocean circulation at the last glacial maximum (LGM)”? [Andreas Sterl, Netherlands]	editorial; accepted
9-825	9	33	9	33	9	Resconstructions (e.g, Dokken, T. M., and E. Jansen (1999), Rapid changes in the mechanism of ocean convection during the last glacial period, Nature, 401, 458–461) also indicated that there was deep convection in Nordic Sea, driven by brine-rejection when sea ice forms. This feature has also been simulated in (Wang, Z. and L. A. Mysak (2006), Glacial abrupt climate changes and Dansgaard-Oeschger oscillations in a coupled climate model, Paleoceanography, 21, PA2001, doi:10.1029/2005PA001238) and in (Transient Simulation of Last Deglaciation with a New Mechanism for Bølling-Allerød Warming, Z. Liu et al. Sience 17 July 2009: 310-314. [DOI:10.1126/science.1171041]). . [Zhaomin Wang, UK]	Taken into account. The section has been revised in collaboration with chapter 5. Addition comments on brine-rjections are provided. Only more recent papers are considered.
9-826	9	33	15	33	17	meridional density' - should this be 'meridional density gradient'. Also why is this gradient a good criterion for comparing models? [Nathan Gillett, Canada]	editorial; accepted
9-827	9	33	21	33	21	"All models show increased salinity" relative to what? [James Christian, Canada]	noted
9-828	9	33	29	33	29	What is ODP? Abbreviation hasn't been introduced yet. [Farahnaz Khosrawi, Sweden]	editorial, ODP supressed
9-829	9	33	29	33	29	Parantheses are misplaced in the sentence. [Farahnaz Khosrawi, Sweden]	editorial, revised
9-830	9	33	30			Figure 9.18: legend uses circles: figure uses triangles [Bruce Wielicki, USA]	editorial, figure has been updated as well as the legend
9-831	9	33	40	33	58	what about CMIP5? [ANNALISA CHERCHI, Italy]	Taken into account. Text modified accordingly.
9-832	9	33	41	33	42	"zonal" is repeated both in text (line 41) and Fig. 9.19 legenda (pag. 144 line 4).	taken into account. "Zonal mean zonal wind" has a

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Later: "the reanalysis estimates lie within the range of model results": true, but the range of model results appears quite large in certain areas. For the SH midlatitudes, for instance, models range between 0.1 and more than 0.2 Nm-2. To decide whether this is a good performance or not, an index of dispersion, or standard deviations, of ERA40 reanalyses should be used. Finally, I suggest to change, in the sentence "At middle to low latitudes, the CMIP3 ... reanalysis" the part "At middle to low latitudes" with "At tropical latitudes" as, in the tropics (from 30°N to 30°S) the model range is less than 0.05 Nm-2, while in the other latitudes exceeds 0.05 Nm-2 up to more than 0.1 Nm-2 at middle latitudes of SH. [Claudio Cassardo, Italy]	precise meaning that requires the use of "zonal" twice. This section has been rewritten, taking into account the reviewers comments.
9-833	9	33	42	33	42	"At most latitudes, the reanalysis estimates lie within the range of model results." This sounds like you are setting the bar very low, but the skill here is actually quite good. [James Christian, Canada]	Taken into account. Section modified.
9-834	9	33	44	33	44	"near the equator this can occur through compensated zonal errors" compensating errors at different longitudes [James Christian, Canada]	Taken into account. Section modified.
9-835	9	33	44	33	44	Same as previous concern: In Chapter 2 the SST data used is HadSST3 and you are using HadSST1.1 in Figure 9.20…. [Celeste Saulo, Argentina]	Taken into account. Reference modified to be in line with Chap 2. Note that we use gridded data to compare with models. Further updates will be done for final version
9-836	9	33	50	33	50	"the surface heat flux balances the convergence of ocean heat transport". Isn't the usual convention to refer to divergence, which can be either positive or negative? [James Christian, Canada]	Taken into account. Section modified.
9-837	9	33	50	33	58	The statistical sample of the models includes 20 years, while that of the observations only 4 yers and two monts. Using a so short period of average for the observations, the data may not represent correctly the heat transport, being affected by some circulation or climatic modes. I think that this fact may be underlined also in the text. [Claudio Cassardo, Italy]	Taken into account. Section modified.
9-838	9	34	6	34	6	fig 9.20: which ones are results from CMIP3 and which are from CMIP5? not clear [ANNALISA CHERCHI, Italy]	Taken into account. Text updated and clarified
9-839	9	34	6			FYI - If figure and analysis is needed Tony Lee from JPL has produced an analysis from CMIP3 and is in the process of doing the same for CMIP5. ref data is quikscat. publication target is by IPCC deadline. tlee@jpl.nasa.gov. [Duane Waliser, USA]	Taken into account
9-840	9	34	15	34	15	The reference of Trenberth and Caron (2001) is missing in the reference list. [Farahnaz Khosrawi, Sweden]	Taken into account. Section modified.
9-841	9	34	23	34	32	The displacement of convection eastward of the Southern Pacific Convergence Zone, known as the double ITCZ, is one of the major error sources leading to the tropical Pacific biases. This should be mentioned too. It affects equatorial wind stress, thereby changes the upper ocean currents and thermocline structures, as demonstrated in a study by Zhang and Song (2010). [Guang Zhang, United States of America]	Rejected. ITCZ issues are described a few lines below
9-842	9	34	23	34	54	In this section only CMIP3 models are mentioned. What about CMIP5 models, as they are mentioned several times in other sections of this chapter? [Claudio Cassardo, Italy]	Taken into account. Text updated with available literature and data
9-843	9	34	24	34	24	A 10-year old paper is not a 'recent' review. [Robert Colman, Australia]	Taken into account. Text modified
9-844	9	34	27	34	29	"Many of the processes leading to these biases have, in principle, been identified, such as too strong trade winds; a too diffusive thermocline; insufficient penetration of solar radiation; and too weak tropical instability waves." What about horizontally isotropic mixing coefficients (see p. 10 line 8)? Also I think this badly understates the inability of these models to resove TIW's. These waves are essentially mesoscale features that are "too weak" in models of even higher resolution than the CMIP5 models. In CMIP1-CMIP3 models they may be entirely absent in many cases. These waves affect the mean state of the tropical ocean circulation in ways that require eddy-resolving resolution to fully represent (e.g. Jochum et al 2005 J. Climate, 18, 841; Jochum and Murtugudde 2006 JPO 36: 592). [James Christian, Canada]	Taken into account. Text modified
9-845	9	34	30	34	30	clumsy wording [Robert Colman, Australia]	Editorial. Corrected

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-846	9	34	30	34	30	Swap 'precisely' with 'to' [Zhaomin Wang, UK]	Editorial. Corrected
9-847	9	34	30	34	32	"processes leading to these biases … such as too strong trade winds, a too diffusive thermocline…"; what is listed are not processes [Tor Eldevik, Norway]	Taken into account. Text modified
9-848	9	34	35	34	35	Now finally the abbreviation ITCZ is introduced after it had already been used twice previously [Farahnaz Khosrawi, Sweden]	Rejected. Term defined in Glossary
9-849	9	34	36	34	36	include also the reference Bellucci et al.(2010). The reference is "Bellucci, A., S. Gualdi, A. Navarra, 2010: The Double-ITCZ Syndrome in Coupled General Circulation Models: The Role of Large-Scale Vertical Circulation Regimes. J. Climate, 23, 1127–1145" [ANNALISA CHERCHI, Italy]	Taken into account. Text modified
9-850	9	34	36			Could also cite the following on double ITCZ in eastern Pacific in CMIP3 models: de Szoeke, S.P. and S.P. Xie, 2008: The Tropical Eastern Pacific seasonal cycle: assessment of errors and mechanisms in IPCC AR4 coupled-ocean atmosphere general circulation models. J. Climate, 21, 2573–2590. and Bellucci, A., S. Gualdi and A. Navarra, 2010: The double-ITCZ syndrome in coupled general circulation models: the role of large-scale vertical circulation regimes. J. Climate, 23, 1127–1145. [Josephine Brown, Australia]	Taken into account. Text modified and moved to 9.4.1
9-851	9	34	37	34	37	Move "a" in these two sentences, so that it reads "….are a too strong seasonal cylce..." and "…a too weak meridional…" [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-852	9	34	37	34	39	A separate but possibly related error in the mean state of most models is the orientation of the SPCZ. In most models it is too zonal (i.e. tends to be too parallel to the equator) - see Brown JB et al. 2011 (J. Climate). [Brad Murphy, Australia]	Taken into account. Text modified and moved to 9.4.1
9-853	9	34	37			A little odd that the double ITCZ is treated in the ocean section. Maybe a pointer in the atmos section could be given to note its treatment below in the ocean section. ----- Very nice work by a postdoc of Yukari Takyabu recently showed that the double itcz seemed to arise from the insensitivity of convective parameterizations of envronmental moisture. ------- Hirota, Nagio, Yukari N. Takayabu, Masahiro Watanabe, Masahide Kimoto, 2011: Precipitation Reproducibility over Tropical Oceans and Its Relationship to the Double ITCZ Problem in CMIP3 and MIROC5 Climate Models. J. Climate, 24, 4859–4873. [Duane Waliser, USA]	Taken into account. The ITCZ discussion is moved to section 9.4.1
9-854	9	34	39	34	39	REGIONAL water vapour feedbacks (to contrast with global response which is not at issue here). Also clarify whar is meant by water vapour feedbacks -- is this top of atmosphere radiative feedback (as used elsewhere in the report)? If not don't use this term [Robert Colman, Australia]	Taken into account. Text modified
9-855	9	34	44	34	45	it would be better to include improvements in CMIP5 instead [ANNALISA CHERCHI, Italy]	Taken into account. Text updated with available literature and data
9-856	9	34	45	34	45	2007 not 2006 [Robert Colman, Australia]	Taken into account. Text modified
9-857	9	34	47	34	47	Please skip the abbreviation EUC. [Farahnaz Khosrawi, Sweden]	Rejected. This is a standard abreviation for oceanographers and helps keep the text short.
9-858	9	34	57	35	1	This section could refer to Fig 9.3. [Nathan Gillett, Canada]	Taken into account. Text modified
9-859	9	34				I suggest adding a reference to the work of J.J. Luo regarding the cold tongue bias and double ITCZ problems e.g. Luo 2002, JAS [Adam Scaife, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Study is cited
9-860	9	35	1	35	2	following refence examined the biases in the tropical Atlantic Ocean in current models and its possible reasons. [Zeng-Zhen HU, USA]	Taken into account. Study is cited
9-861	9	35	1	35	2	Huang, B., Z.-Z. Hu, and B. Jha, 2007: Evolution of model systematic errors in the tropical Atlantic basin from the NCEP coupled hindcasts. Clim. Dyn., 28 (7/8), 661-682, DOI: 10.1007/s00382-006-0223-8. [Zeng-Zhen HU, USA]	noted. ref associated to 9-860
9-862	9	35	6	35	6	Hard to know what this means. [Robert Colman, Australia]	Taken into account. What we meant is that Atlantic biases, in comparison to those in the Pacific, are more likely to involve model errors over the adjacent

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							continents, because of its small basin size. For example, model errors in simulating deep convective system over the Amazon can produce trade wind errors, which in turn contribute to the SST bias in the eastern equatorial Atlantic, as hypothesized by Richter and Xie (2008).
9-863	9	35	8	35	8	GCM? I guess you mean here AOGCM, aren't you? [Farahnaz Khosrawi, Sweden]	Taken into account. Text clarified. Here GCM studies include both AGCM and AOGCM studies.
9-864	9	35	9	35	9	previous to CMIP3? [ANNALISA CHERCHI, Italy]	Taken into account. Text clarified. Previous GCM studies referred here include both CMIP3 studies and those prior to CMIP3
9-865	9	35	10	35	10	The reasons caused the warm biases in the tropical Atlantic Ocean may be complicated. Our diagnostic and model sensitive experiments (Huang et al. 2007; Hu et al. 2008a, 2008b, 2011; Wahl et al. 2009) suggested that Cloud–radiation–SST interaction processes play an important role in model bias evolution as well as in anomalous climate events in the southeastern Atlantic Ocean. [Zeng-Zhen HU, USA]	Taken into account. Text modified. Yes. We agreed that the Atlantic bias problem is complex and the cause may vary from model to model. We now include a brief discussion on cloud-radiation-SST feedback and other processes in the revision and cite the relevant references as suggested.
9-866	9	35	10	35	10	Hu, Z.-Z., B. Huang, and K. Pegion, 2008a: Low-cloud errors over the southeastern Atlantic in the NCEP CFS and their association with lower-tropospheric stability and air-sea interaction. J. Geophys. Res. (Atmosphere), 113, D12114, doi: 10.1029/2007JD009514. [Zeng-Zhen HU, USA]	noted. ref associated to 9-865
9-867	9	35	10	35	10	Huang, B., Z.-Z. Hu, and B. Jha, 2007: Evolution of model systematic errors in the tropical Atlantic basin from the NCEP coupled hindcasts. Clim. Dyn., 28 (7/8), 661-682, DOI: 10.1007/s00382-006-0223-8. [Zeng-Zhen HU, USA]	noted. ref associated to 9-865
9-868	9	35	10	35	10	Hu, Z.-Z., B. Huang, Y.-T. Hou, W. Wang, F. Yang, C. Stan, and E. K. Schneider, 2011: Sensitivity of tropical climate to low-level clouds in the NCEP Climate Forecast System. Clim. Dyn., 36 (9-10), 1795-1811, DOI: 10.1007/s00382-010-0797-z. [Zeng-Zhen HU, USA]	noted. ref associated to 9-865
9-869	9	35	10	35	10	Hu, Z.-Z., B. Huang, and K. Pegion, 2008b: Leading patterns of tropical Atlantic variability in a coupled general circulation model. Clim. Dyn., 30 (7-8), 703-726, DOI: 10.1007/s00382-007-0318-x. [Zeng-Zhen HU, USA]	noted. ref associated to 9-865
9-870	9	35	10	35	10	Wahl, S., M. Latif, W. Park, and N. Keenlyside (2009): On the Tropical Atlantic SST warm bias in the Kiel Climate Model. Climate Dynamics, DOI:10.1007/s00382-009-0690-9. [Zeng-Zhen HU, USA]	noted. ref associated to 9-865
9-871	9	35	10	35	12	is 2005 recent? [ANNALISA CHERCHI, Italy]	Taken into account. Text modified
9-872	9	35	17			Section 9.4.2.5.3: can you comment here on the tendency for CMIP5 models to exhibit cold SST biases in the Arabian Sea region, since these have been shown to affect the development of the South Asian summer monsoon. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Text modified following available literature
9-873	9	35	27	35	34	How does this relate to the Indian Ocean Dipole discussed elsewhere? [Nathan Gillett, Canada]	Taken into account. The IOB mode in 9.4.2.5.3 Tropical Indian Ocean (page 35) moves to 9.5.3.4.2 Indian Ocean Dipole (page 9-50), into a subsection “Indian Ocean modes”
9-874	9	35	28	35	29	Skip the abbreviation IOB. [Farahnaz Khosrawi, Sweden]	Rejected. This is a standard abreviation and helps keep the text short.
9-875	9	35	34	35	34	Zheng or Zhang? I cannot find a reference for Zheng in the reference list. [Farahnaz Khosrawi, Sweden]	It is Zheng et al., listed in References (page 9-109, line 10)
9-876	9	35	36	35	36	the "summary" subsection should contains at least comments on the improvements in CMIP5 [ANNALISA	Taken into account. Text modified

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						CHERCHI, Italy]	
9-877	9	35	38	35	47	In case of insufficient evidence and insufficient agreement, again wouldn't it be more reasonable scientifically to stop the writing of the AR5 report there and avoid dubious projections ? [François GERVAIS, France]	Rejected. As shown by the vast body of literature, assessed in the several IPCC reports, there is sufficient scientific evidence that models are reliable enough to provide global projection of climate change. Projections are useful for making policy analysis.
9-878	9	35	39	35	39	"Du et al., 2009" => "Du et al., 2009; Huang and Shukla 2007a,b" [Zeng-Zhen HU, USA]	Taken into account. Text modified
9-879	9	35	39	35	39	Huang, B., and J. Shukla, 2007a: On the mechanisms for the interannual variability in the tropical Indian Ocean, Part I: The role of remote forcing from tropical Pacific. J. Climate, 20, 2917-2936. Huang, B., and J. Shukla, 2007b: On the mechanisms for the interannual variability in the tropical Indian Ocean, Part II: Regional processes. J. Climate, 20, 2937-2960. [Zeng-Zhen HU, USA]	Taken into account. Text modified
9-880	9	35	39	35	50	The section 9.4.3.6. would be much more effective summary if written as a table. [Juan A. Blanco, Canada]	Taken into account. The summary Fig. 9.43 is designed to present this summary as a table
9-881	9	35	40	35	40	"robust evidence and high agreement that SST is well simulated": with the exception of the Atlantic - more - and Pacific - less - tropical areas. [Claudio Cassardo, Italy]	Taken into account. Regional assessment is out of scope for this section. See below.
9-882	9	35	40	35	47	This long sentence reads awkward. [Guang Zhang, United States of America]	Taken into account. Section modified.
9-883	9	35	41	35	41	"limited evidence and medium agreement that SSS is not correctly simulated" I don't think the "not" is supposed to be here. [James Christian, Canada]	Rejected.The assessment made here is that the SSS is not well simulated. There is limited evidence and medium agreement for this assessment.
9-884	9	35	41	35	50	"essential processes" - which are? And to what extent does the preceding text docoment their simulation by CMIP3 models? [Tor Eldevik, Norway]	Taken into account. Text clarified
9-885	9	35	43			The statement that SST is "well-simulated," depends on the context. Even small biases less than 1 K in the tropics could have a significant on tropical storm climatology. A quantitative range, e.g less than 1 K, should be stated for what constitutes "well-simulated." [David Bader, USA]	Taken into account. Text modified. It is nevertheless very difficult to provide a quantitative range as this will depend on the region.
9-886	9	35	49			The large disagreement between sea ice simulations and observations are not reflected in the sea ice section. Important recent references are missing. Why are the model simulations only compared against satellite observations? There are good reconstructions available. The reconstruction of Kinnard et al (Nature, 2011) is only one example that is missing. The simulations of sea ice extent shown in Figure 9.24 clearly disagrees with the Kinnard et al. as well as with the Walsh & Chapman reconstructions in the first half of the century. The extent of the reconstructions before 1970 seem to be well above the standard deviation of the CMIP5 ensemble and are better represented by CMIP3. Thus, it seems that CMIP5 has only a warm bias with respect to CMIP3 but does not better represent the declining trend over the century. Another striking disagreement between models and observations is the trend of sea ice velocity and thickness.The findings of Rampal et al (JGR, 2011) show that the models do not even capture the seasonality of sea ice drift. The CMIP3 ensemble has underestimated the thinning trend by a factor of 4. Has this improved in CMIP5? Given the problems of representing correctly the dynamics, the models probably miss an important positive feedback process. These shortcomings and uncertainties should be communicated more truthfully. The section of sea ice dynamics (9-11) does not mention the importance of rheology at all and that almost all models include viscous-plastic or similar formulations. Coon et al (JGR, 2007) stress "that observations of stress and ice motion do not support the assumptions common to most ice dynamic models in use today, and new models based on different assumptions are needed". Thus, there is a strong risk of an ensemble failure in representing correctly the sea ice dynamics under transient climate changes if the rheology has just been tuned to a constant mean sea ice thickness. [Lars Kaleschke, Germany]	Taken into account. Reference is made to Ch.4 discussing the observationally based estimates of sea ice characteristics and their reliability. The figure 9.24 is updated by extending the observationally based time series and including more CMIP5 models. Text modified. Possible additions to the text depend on availability of published studies of ice velocity and thickness simulated by CMIP5 models.
9-887	9	35	52	37	23	Somewhere in this chapter - and this seems the best place - models should be evaluated for their surface albedo feedback, as in Boé et al (J Climate 22:4682–4695, 2009) and Qu and Hall (J Climate 19:2617, 2006). Chapter 7 has an extensive discussion of cloud, water vapor, and lapse rate feedbacks in both observations	Taken into accopunt -- covered in 9.7

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						and GCMs, but nowhere could I find in the IPCC report a similar discussion of the other most important climate system feedback - surface albedo feedback - in models and observations. It should be included in chapter 9. [Philip Mote, USA]	
9-888	9	35	53	35	54	Skip the abbreviation PMW. [Farahnaz Khosrawi, Sweden]	Accepted
9-889	9	35	54	36	2	The discussion of the quality of retrieval algorithms should be more consistent with the one of chapter 4, (page 4-7) [Hugues Goosse, Belgium]	Taken into account -- reference to Ch.4 is made
9-890	9	35	56	35	56	"as large as 1 million square kilometres": this number will be more informative is expressed also in percentage of the total extent in a given year or year interval. Furthermore, it is not clear if it is referred to the NH and SH separately, or to the total amount (I presume the total amount, but better to specify it). Finally, when it is said "from different algorithms" (and also line 9 pag. 36 " various observational sea ice extent estimates"), it refers to the algorithms of the two centers HadISST and NSIDC, mentioned in Fig. 9.22 (in this case, two is less than 'several' or 'various'), or includes also other algorithms/centers (in this case, please specify which ones in the text)? [Claudio Cassardo, Italy]	Taken into account -- text modified.
9-891	9	36	3	36	4	OK: therefore, it would be useful to evaluate models and multi-model means in terms of trends over the satellite era (1979 onwards), rather than only in terms of absolute values. [Jerome WEISS, France]	Noted. This is shown at Fig. 9.24
9-892	9	36	11	36	14	To correctly describe changes in Arctic minimum extent from CMIP3 to CMIP5, add the following: "Nevertheless, this improvement is partly due to more models with lower ice extents in the Beaufort Sea than observed." [Arne Melsom, Norway]	Noted, however it may be accepted only if there is a published study proving this.
9-893	9	36	14	36	14	fig. 9.22: why in CMIP5 the error limits are larger than in CMIP3? [ANNALISA CHERCHI, Italy]	Rejected -- the error limits are 1 std both for CMIP3 and CMIP5
9-894	9	36	24	36	24	"annual maxima and minima": in truth, Fig. 9.23 shows not the maxima or minima of sea ice extent, but the number of grid points in which the number of CMIP3/CMIP5 (or their difference) models simulating at least 15% of the area covered by sea ice. Fig. 9.22 shows the maxima/minima, but without the geographical subdivision. So I suggest to rephrase the sentence. [Claudio Cassardo, Italy]	Noted -- text modified for other reasons. The definition of extent as the area within the concentration >15% is in the text. And it is widely used in the literature.
9-895	9	36	24	36	25	"with few exceptions, particularly in the Northern North Atlantic in winter": very few exceptions, just a tenth of grid points, plus one or two in northern Pacific, but only in winter, while in the SH there are a few grid points both in winter and summer. In my opinion, I suggest to remove the geographical specification. The most important messages emerging from the figures are two: 1) CMIP5 model mean is lower than CMPI3 model mean; cumulating both hemispheres, and their difference is larger than 1 million square km (which can be considered as the observational accuracy); in the NH, CMIP5 model mean approximate better the data during the September means (Fig. 9.24); 2) the number of grid points in which the number of CMIP5 models simulating at least 15% of the area covered by sea ice outnumbers the number of grid points in which the number of CMIP3 models simulating at least 15% of the area covered by sea ice is larger than that in which the contrary occurs. These messages, in my opinion, does not emerge clearly from the actual text. [Claudio Cassardo, Italy]	Taken into account when revising the text.
9-896	9	36	38	36	40	The trend is the Southern Ocean is not at all discussed while it is an important question. The evaluation of the models regarding this point should also be included in Figure 9.43. [Hugues Goosse, Belgium]	Accepted.
9-897	9	36	40	36	40	The inability of current AOGCMs to simulate the observed increasing trend in Antarctic sea ice extent and the potential impact of this drawback on near-term and long-term projections should be addressed. [Thierry Fichefet, Belgium]	Accepted
9-898	9	36	40	36	42	What about trends in terms of sea ice thickness, from CMIP 3 to CMIP 5 ? Rampal et al. (JGR-C, 116, C00D07, 2011) have shown that CMIP3 simulations underestimated by a factor larger than 4 the thinning trend of Arctic sea ice observed over the period 1980-2008 by Kwok and Rothrock (GRL, 36, L15501, 2009). Are the CMIP5 simulations better in this respect ? [Jerome WEISS, France]	Taken into account - analysis of thickness trends will be added if available from published studies of CMIP5 ice thickness.
9-899	9	36	40			A useful / missing reference here is: Reference: Kwok, R. 2011. Observational assessment of Arctic Ocean	Noted, but seems to be helpful here only if a similar

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						sea ice motion, export, and thickness in CMIP3 climate simulations. Journal of Geophysical Research 116: 10.1029/2011JC007004. [Duane Waliser, USA]	paper on CMIP5 models appears.
9-900	9	36	50	36	58	This paragraph discusesses why sea ice in model results and observed sea ice differ. The main forcing that gives rise to changes in ice thickness (due to ridging) -winds- is mentioned, and a relevant paper is quoted. On the other hand, sea ice melting can to lowest order be attributed to ice-ocean interaction (heat flux from ocean to sea ice). This process is also listed, but no relevant study is quoted. I suggest adding the following on line 56 (after 'component.'): "Melsom et al. (2009) found that an improved description of heat transport by ocean currents leads to significantly enhanced results for sea ice." Reference: Melsom, A., V. Lien, and W.P. Budgell: Using the Regional Ocean Modeling System (ROMS) to improve the ocean circulation from a GCM 20th century simulation. Ocean Dynamics. 59, 969-981. [Arne Melsom, Norway]	Accepted - reference added
9-901	9	36	50	36	58	Are there no studies of e.g. albedo that can be discussed? [Gunilla Svensson, Sweden]	Taken into accopunt -- covered in 9.7
9-902	9	36	52	36	54	Another important factor is sea ice mechanics and kinematics, which play a strong role on average sea ice age, sea ice export, and so sea ice mass balance (Rampal et al. , JGR-C, 116, C00D07, 2011) [Jerome WEISS, France]	Noted. This is true, but the paragraph dicusses "external" problems of sea ice simulations.
9-903	9	36				Figure 9.22, Figure 9.24: Use a separate color where CMIP3 and CMIP5 ensembles overlap, so the full span of both ensembles is revealed. [Arne Melsom, Norway]	Accepted
9-904	9	37	2	37	2	natural variability' here should be 'internal variability in the models'. The internal variability in the models is very different in different models, while natural variability in the real world has only one realization, presumably caused by internal physical processes. Some internal variability may result from specific model designs (non-physical aspects). [Zhaomin Wang, UK]	Rejected. Natural variability here applies to the real-world sea ice.
9-905	9	37	4			nothing in this section on sensible and latent heat flux evaluation – e.g. see Blyth et al (2011, GMD) paper on benchmarking. Stress importance (and difficulty) of simulating carbon and energy fluxes accurately together. [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Revised text now mentions the role of the land-surface in water and energy partitioning and briefly covers offline validation studies.
9-906	9	37	6			permafrost is mentioned in the title here, but not in the text - see, e.g. Dankers et al, (2011, The Cryosphere) for an evaluation of simulated permafrost in the JULES model. [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account. The revised text now contains discussion of the implications of model projections for Permafrost, and points more clearly to Figure 9.25.
9-907	9	37	6			I find this sections very strange? As stated in the beginning of the chapter(9.1.3.2.3), the land model calculates the fluxes of heat, water and momentum between the land and the atmosphere. Why are evaluation of these not discussed here? It is also mentioned in the earlier section that the land-surface schemes ar emore straightforward to evaluate because they can be evaluated in stand-alone mode. I disagree with that statement, you do not know if you have done the land schemes right until you have evaluated their performance in coupled mode. That is something that is done within GLASS and some studies have come as a result. This should be discussed here. [Gunilla Svensson, Sweden]	Taken into account - see response to Comment 9-905.
9-908	9	37	6			This whole section contains far too much on DGV and carbon etc cycles considering that there is a whole chapter on that already which seems to contain evaluation as well. Or, move all evaluation to this chapter. Use the same principle as for clouds and aerosols. [Gunilla Svensson, Sweden]	Noted. The agreed demarcation between Chapter 6 and Chapter 9 concerning carbon cycle modelling is as follows: Chapter 6 covers offline land and ocean carbon cycle models, and the analysis of future carbon cycle feedbacks in ESMs; Chapter 9 focusses on the evaluation of the carbon cycle simulation (i.e. against observations) for the historical periods of the ESM simulations. However, we must also cover relevant offline evaluation where this realtes to the performance of climate models.
9-909	9	37	10	37	15	There isn't much on permafrost in this section. This is very important for carbon cycle modeling. [Annarita Mariotti, U.S.A.]	Taken into account - see response to Comment 9-906.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-910	9	37	11	37	13	Exactly this same sentence on masking of snow albedo by vegetation was in AR4 chapter 8 and attributed to Roesch (2006). An attribution is needed here again - the same if no new investigations have been undertaken, but reflecting differences in CMIP5 if possible. For point simulations in the SnowMIP2 intercomparison, Essery et al. (Bull. Amer. Met. Soc., 90, 1120 – 1135, 2009) did not find the same positive biases in albedo over snow-covered forests reported by Roesch (2006) in global models [Richard Essery, UK]	Taken into account. Text on snow and permafrost revised.
9-911	9	37	19			Why February? It's not a bad choice, but it is slightly arbitrary. Chapter 4 focuses on March-April average, which is important because of the strength of the snow-albedo feedback then. [Philip Mote, USA]	Noted. We could have chosen a different period, but we conclude that the choice of February demonstrates the general point clearly enough.
9-912	9	37	23	37	41	There are several inaccuracies in this section. It would require a more in-depth literature review. I have tried to note some of the issues hereafter, but there are more. [Sonia Seneviratne, Switzerland]	Noted. Unfortunately there is insufficient space to undertake a more in-depth literature review, but we have corrected the inaccuracies.
9-913	9	37	25	38	25	This section (land surface-atmosphere coupling) is less developed than the previous ones. However there are some recent papers (i.e from 2009 to 2011) referred. My question is why any of the results of the referred papers is included and/or adapted to this report, as is done in other subsections?. I consider that some evidences will help the reader to follow some of the sentences -regarding plausible impacts- that are included in this section. [Celeste Saulo, Argentina]	Taken into account. Text rewritten with additional references.
9-914	9	37	27	37	28	Also vice versa (wet anomaly can produce an excess of evapotranspiration, thus generating more precipitation which tends to maintain the anomaly). [Claudio Cassardo, Italy]	Noted.
9-915	9	37	27	37	28	Soil moisture-precipitation feedbacks are not necessarily positive, in some cases, enhanced precipitation can be induced by dry anomalies. See Seneviratne et al. (2010, Earth-Science Reviews) for a review of the literature on this topic, as well as Taylor et al. (2011, Nature Geoscience) for a recent article documenting a negative feedback. [Sonia Seneviratne, Switzerland]	Taken into account. Text rewritten.
9-916	9	37	27	37	35	Mention here that the soil moisture-precipitation feedback varies (even in sign) with timescale and spatial scale (see e.g. Taylor and Ellis, 2006 (GRL 33 (L03404). doi:10.1029/2005GL025252); Taylor and Lebel, 1998 (MWR 126:1597-1607)). [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Text rewritten.
9-917	9	37	31	37	33	Is such a bistable Sahel really seen in models? Give a reference. [Nathan Gillett, Canada]	Noted. Bistability of the climate-vegetation state in the Sahel has been seen in some models. However, this text has now been removed.
9-918	9	37	35	37	43	The land surfrace and hydrology discussion misses some key issues. Other aspects of land hydrology beyond soil moisture, snow and land-atmosphere flixes are also compared with observations data, eg: river flows (Falloon, Pete, Richard Betts, Andrew Wiltshire, Rutger Dankers, Camilla Mathison, Doug McNeall, Paul Bates, Mark Trigg, 2011: Validation of River Flows in HadGEM1 and HadCM3 with the TRIP River Flow Model. J. Hydrometeor, 12, 1157–1180.)doi: http://dx.doi.org/10.1175/2011JHM1388.1 Offline land surface models (some of which are also land surface schemes for climate models, eg: MOSES/JULES are increasingly benchmarked and compared with each other and with other hydrology models and observations, eg: the WaterMIP study. Also this section should mention the effects of CO2 physiological forcing on hydrology, via impacts on transpiration. I suggest the soil moisture section should be re-scoped and re-named to include some discussion of these important areas of the literature. [Richard Betts, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Our revision of this sub-section now focusses more on surface hydrology (including river runoff) and less on land-atmosphere coupling.
9-919	9	37	36	37	36	Fischer et al. (2007) investigated the role of land-atmosphere coupling for the occurrence of heat waves, not for the persistence of summer droughts. Please check publications led by Siegfried Schubert regarding the role of land-atmosphere feedbacks for drought persistence (e.g. Schubert et al. 2004, Science; Schubert et al. 2008, J. Climate) [Sonia Seneviratne, Switzerland]	Taken into account. Reference to Fischer et al. 2007 now moved to apply to extreme temperatures.
9-920	9	37	41	37	43	Claim that climate change is likely to increase semi-arid areas is unsubstantiated and speculative. The reference cited does not provide primary evidence for this [Stewart Franks, Australia]	Noted. However, this sentence has now been removed.
9-921	9	37	45	37	50	results from the ENSEMBLES EU: check for intercomparison with climate models including DGVMs [ANNALISA CHERCHI, Italy]	Noted.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-922	9	38	9	38	10	"...tropical ecosystems are thought to be phosphorus rather than nitrogen limited" - reference to the literature source is missing [Ladislav Metelka, Czech Republic]	Noted. See response to Comment 9-923.
9-923	9	38	10	38	11	Thhe statement "tropical systems are though to be phosphorus rather than nitrogen limited" is not accurate anymore in light of increasing body of knowledge. It is not accepted that the tropical ecosystems are also frecuently limited in nitrogen, depending on the soil type, but specially in secondary forests following distrubances and stand-replacing events. With the increase of extreme weather events under global warming conditions, it is expected that these stand-replacing events will be more common, and therefore the limitating role of nitrogen in tropical forest will increase. Read/cite Wei et a. (2012, Science of the Total Environment 416, 351-361), Davison (2007, Nature 2007;447:995–9), Lebauer and Treseder (2008, Ecology 2008;89:371–9) [Juan A. Blanco, Canada]	Noted. As for most scientific issues, there continues to be debate. Given the space constraints, in order to avoid a more extended discussion of this we have removed this sentence.
9-924	9	38	12			need a paragraph on evaluation of simulated carbon emissions from land-use as well as physical effects. This is a big uncertainty and models not well understood. I could provide some text for this – or you could approach Elena Shevliakova who is an expert on this [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Whilst it has turned-out to be difficult to diagnose land-use CO2 emissions from ESMs, we recognise that this is a new and significant source of spread amongst models. We make this point clearer in our revision.
9-925	9	38	13	38	13	Perhaps the Land-use change part could be expanded. Do models include desertification, or issues such as planting of crops for alternative food sources? [Graham Feingold, United States of America]	Noted. See response to Comment 9-924. The land-use scenarios used for the RCPs scenarios account for changes in land-use arising from deforestation, abandonment and agriculture. However, discussion of these issues is outside of the scope of this chapter.
9-926	9	38	13		25	Do (or can) models predict changes to land use based on other changes in clmate factors? It would seem like a natural outcome of climate change that land use would (must) change so their forcing may change as climate itself changes.It would be interesting to predict areas that may go fallow or become viable for agricultre (for example) and see if there are further climate impacts. [Larry Thomason, United States of America]	Noted. A land-use change scenario is associated with each RCP (consistent with the demands of land for food, energy etc. for each RCP), and these land-use scenarios are prescribed in the ESM runs. ESMs then interactively calculate the CO2 emissions associated with the prescribed land-use change. Those ESMs which include dynamic global vegetation models also attempt to simulate the impacts on "natural vegetation". However, discussion of these issues is outside of the scope of this chapter. See response to Comment 9-925.
9-927	9	38	17	38	19	What are the references for this sentence? [Sonia Seneviratne, Switzerland]	Taken into account. Reference for CO2 emissions from land-use now added, based-on the analyses presented in Chapter 6.
9-928	9	38	18			"20%" figure – cross-check with Ch6 for consistency [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account. See response to Comment 9-927.
9-929	9	38	28	38	54	This section should also mention issues related with the drought projections in the Amazon region. The IPCC SREX has assessed that there is low confidence regarding these changes (Seneviratne, Nicholls, et al. 2012; see Table 3.3) [Sonia Seneviratne, Switzerland]	Rejected - outside of the scope of the chapter. This chapter is about model evaluation against data rather than future projections.
9-930	9	38	30	38	38	Why? What is the reason for these differences between models and observations and that some models perform better than others? [Farahnaz Khosrawi, Sweden]	Accepted - text revised. In the models which have phase errors at the South Pole this suggests errors in the seasonal cycle of CO2 uptake in the Southern Ocean. This statement is therefore not relevant to this sub-section on the Terrestrial Carbon Cycle and has been removed.
9-931	9	38	31		39	Is soil carbon (lifetime) considered in the models (lignins and other decaying organic materia)? [Larry Thomason, United States of America]	Yes. Taken into account - text added to 9.4.4.3 to describe the nature of soil carbon models in DGVMs.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-932	9	38	37	38	38	The causes of these errors, due to the scarce vegetation (and generally land cover) of the SH, may be due to some ocean-carbon processes rather than to the terrestrial processes (for instance, as countereffects of the errors in the Southern Ocean circulation mentioned at page 32?), or some errors arising from the contribution of Amazon forest? [Claudio Cassardo, Italy]	Accepted - taken into account. This phrase has been removed. See response to comment 9-930.
9-933	9	38	42			Name the ESM. [Nathan Gillett, Canada]	Noted. However, this text has now been removed as the result did not actually pertain to an ESM.
9-934	9	38	45	38	47	Who made these conclusions? [Claudio Cassardo, Italy]	Accepted. Reference added.
9-935	9	38	49	38	49	the C emission by forest fires will be more informative is expressed also in percentage of the total extent in a given year or year interval. [Claudio Cassardo, Italy]	Accepted - text revised. % figure is now also given.
9-936	9	38	56			Section 9.4.5.2: I think it needs to be explicit here that some of the experiments cited are not climate model simulations but rather simulations with OBGC models similar to those used in climate models, but forced with historical observed meteorology (reanalysis). For the results of Friedrichs and Henson, these comparisons would not be meaningful if they were conducted with coupled models, because they involve comparisons for specific time periods or the timing of specific historical events. [James Christian, Canada]	Accepted - text modified. This section now makes the distinction between offline and online studies clearer.
9-937	9	39	4	39	6	Explain more what this sentence on 'similar error structures' is referring to. [Nathan Gillett, Canada]	Accepted - text modified. These "error structures" refer to similar patterns in the spatial structure of model errors, and in the timing of the seasonal cycle at given locations. This sentence has been rewritten to make this clearer.
9-938	9	39	16			Replace 'warming' with 'temeprature'. A trend in warming would be a second derivative of temperature. [Nathan Gillett, Canada]	Accepted - text modified.
9-939	9	39	29			Section 9.4.5.3: Shouldn't it be C4MIP (without the 4 as superscript?). [Farahnaz Khosrawi, Sweden]	Rejected. No the correct aronym should have a superscript (standing for the Coupled Climate Carbon Cycle Model Intercomparison Proejct).
9-940	9	39	29			Section 9.4.5.3: Some mention of the validity of the vegetation distributions simulated for the present-day by those CMIP5 ESMs that simulate the terrestrial carbon cycle would be useful. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Accepted. Most ESMs still use prescribed vegetation distributions, so it isn't possible in general to validate their vegetation distributions. However, a relevant sentence has been added to 9.4.4.3.
9-941	9	39	34			C4MIP used the A2 scenario (not A1B) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Accepted - text modified.
9-942	9	39	37	39	37	Is the sentence "...in which increases in atmospheric CO2 did not influence climate" correct? I think there should be "… in which climate-carbon cycle feedbacks did not influence climate“. [Ladislav Metelka, Czech Republic]	Rejected. The uncoupled runs exclude the effects of climate change on the carbon cycle, and they do this by disabling climate change due to all CO2 increases.
9-943	9	39	47			Ch.6. has adopted the phrase “compatible emission” rather than “permissible”. Suggest Ch.9. Does likewise. We felt “permissible” risked sounding too policy-prescriptive. WG1 as a whole should certainly be consistent [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Accepted - text modified such that "compatible emissions" is used rather than "permissible emissions".
9-944	9	39	54	39	54	This sentence is quite difficult to read with all these abbreviations. Are the abbreviations RCP and ESMs really necessary? [Farahnaz Khosrawi, Sweden]	Noted, and we have tried to avoid unnecssary aronyms in the the modified text. However, some short-hand terms such as "ESM" and "RCP" are required.
9-945	9	39	54	40	53	In all this discussion, it could be important to mention the uncertainty of the GCP data or fluxes, in order to have a term of comparison with which to evaluate the model performances. [Claudio Cassardo, Italy]	Accepted - text modified to include some estimate of uncertainty in the "observed" GCP global fluxes.
9-946	9	39	56	39	57	"each of these models also simulate the affects of land-use change on both the carbon cycle and the biophysical properties of the land-surface" and do so in a more physically consistent way than in earlier models: see Chapter 6 [James Christian, Canada]	Accepted - text modified to make the greater consistency clear.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-947	9	40	1	40	4	These text is repeated in the Figure caption, thus here it can be either omitted or shortened. [Farahnaz Khosrawi, Sweden]	Accepted - text and figures both modified in the SOD.
9-948	9	40	17			There is no dotted line in the figure. [Andreas Sterl, Netherlands]	Accepted - text and figures both modified in the SOD.
9-949	9	40	19	40	53	The description of figures 9.26 and 9.27 in the text does not match with what is shown on these figures. Please check. [Farahnaz Khosrawi, Sweden]	Accepted - text and figures both modified in the SOD.
9-950	9	40	21		22	What do you mean by land carbon storage? [Larry Thomason, United States of America]	Accepted - text modified. "Land carbon storage" means the carbon stored on land in soils and vegetation. We now make this clearer.
9-951	9	40	22	40	42	"GtC" appears a number of times in this passage, doesn't seem to occur anywhere else. Should use SI units. [James Christian, Canada]	Accepted - text modified. We now use "PgC" rather than "GtC" throughout.
9-952	9	40	24	40	27	Looking contemporary at the two figures, ocean part, it becomes evident that, even if all EMSs underestimate the GCP data, for some models the underestimate is decreasing in the more recent ten years of the period, when the modelled ocean flux is larger than the observed one. [Claudio Cassardo, Italy]	Noted. Figure has been replaced and focus is now on period 1986-2005.
9-953	9	40	29	40	38	What is happening in MIROC pre-1960 in Fig 9.26? Why are its carbon pools losing carbon? [Nathan Gillett, Canada]	Noted. It seems likely that prior to 1960 the MIROC model simulates a carbon loss from land-use change (especially deforestation) that is larger than the carbon sink in the undisturbed vegetation. However, we do not have space to go into a discussion of the behaviour of individual models.
9-954	9	40	30	40	30	"the observed meteorology" the historical atmospheric forcing [James Christian, Canada]	Accepted - text revised.
9-955	9	40	36	40	38	Does this sentence mean anything? It seems obvious to the point of tautology. Larger sinks, lower atm CO2. Smaller sinks, higher atm CO2. [James Christian, Canada]	Accepted - text removed in the SOD.
9-956	9	40	36	40	38	Yes, but what does this mean? What do we learn from this statement? [Andreas Sterl, Netherlands]	Accepted - taken into account. See response to comment 9-955.
9-957	9	40	37	40	37	I think it should be Figure 9.26 instead of Figure 9.27. [HASIBUR RAHAMAN, India]	Accepted - text revised.
9-958	9	40	40	40	40	in Figure 9.27, the IPSL model is mispelled as IPCL [Celeste Saulo, Argentina]	Accepted - text revised.
9-959	9	40	50	40	51	Is this really true? Based on Fig 9.27 I would say that relative to GCP the estimates for land are quite scattered but don't seem to have a bias (for the mean), and the observational estimates are not very well constrained. For the ocean the spread is small but there is a clear bias. Certainly the land models do not agree, but I don't have a lot of confidence in the observations. In any case, inferring the land sink from atmospheric concentration depends on the estimated ocean sink, and is determined in part by exactly how you weight the uncertainties in the ocean observations. [James Christian, Canada]	Accepted - text revised to make these nuiances clearer. The revision now also makes use of flux estimates from atmospheric inversions and the Takahashi dataset.
9-960	9	40	57			This section should really be on "Aerosols" rather than "Sulfur cycle". It is reductionist to talk about sulphate aerosols alone, especially in the context of dimming/brightening. [Olivier Boucher, France]	Accepted -- section title revised.
9-961	9	41	5	41	5	Write "North America" [Juan A. Blanco, Canada]	Editorial
9-962	9	41	7	41	7	This 24% reduction sounds high to me if it include Asian emissions. Reference please. [Olivier Boucher, France]	Accepted -- Value has been corrected to 15% and reference for percentage decline (Smith et al 2011) added.
9-963	9	41	7	41	9	I'm sorry but I don't understand how Figure 9.28 illustrates what is mentioned in this paragraph. [Celeste Saulo, Argentina]	Accepted -- the insertion of a refereence to figure 9.28 here was in error.
9-964	9	41	7	41	14	It might be useful for model validation and useful for intpreting model simulations to show timseries of global sulphate loading in CMIP5 models and CTMs. This would be helpful in interpreting aerosol-induced cooling in	Taken into account -- variation in historical AOT (Fig. 9.29) is supplied for purpose of Chapter 10 since this

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						chapter 10. [Nathan Gillett, Canada]	is particularly relevant for understanding aerosol cooling.
9-965	9	41	7	41	22	Figure 9.28 shows the error between the mean of the models and the observed data. These errors are (in absolute) larger than 0.2 units almost everywhere on land, and smaller on oceans with some exceptions in tropical and polar oceans. Figure 9.29 shows the values of some models and of the observed data, showing that they are comprised between 0.1 and 0.2 units (excluding volcanis effects). This value is of the same order of magnitude of the uncertainty of the model mean, that presumably is smaller that that of the individual models. Due to this large uncertainty,I think that these models can hardly be able to show a statistically significant trend (see also note 107). [Claudio Cassardo, Italy]	Rejected -- Uncertainty in the mean value of a time series does not logically imply that the trend is also uncertain.
9-966	9	41	7		14	Refs needed; one study or many; uncertainties; fig refers to total aod whereas text refers to sulfate. need to reconcile. [Stephen E Schwartz, USA]	Taken into account -- combined with comments 9-960, 9-967, and 9-976.
9-967	9	41	11	41	11	The figure is about total AOD not sulfate aerosols. The subsection title should reflect that as well. [Olivier Boucher, France]	Accepted -- subsection title revised.
9-968	9	41	12	41	13	"The reason is that ... less oxidant limited": this sentence is not clear to me (I am unfortunately not an expert of atmospheric chemistry). The reason for what? In the previous sentence the different effect of Asian and European trend variations is mentioned. This sentence mentions a southwards motion without specifying if it refers to European or Asian SO2. Please rephrase the two sentences to clarify better the process and the reason for the different effect. [Claudio Cassardo, Italy]	Accepted -- text revised
9-969	9	41	16	41	22	Honestly the analysis of Fig. 9.29 does not seem (to me) to support the text. First: in the text is mentioned a transition between the values before and after the 1980, while the figure shows data only after 1980, thus making impossible to compare the two periods. Second: the trends of all models, removing the effects of volcanic explosion, are relevantly different (actually, almost null :less than 0.01 unit/year) than that of the time series of global oceanic-mean AOT. [Claudio Cassardo, Italy]	Accepted -- The start time for the figure has been switched from 1980 to 1850. The data in the figure has been switched from the GACP retrieval of AOD to the MODIS assimilation-quality retrieval of AOD. The latter data set shows no trend and is fully consistent with the text.
9-970	9	41	16	41	22	Here sulfate and other aerosols are identified as the principle causes of the 'global dimming' between the 1950s and 1980s, and in IPCC AR4 their effects were hypothesized as interrupting the global warming trend during the 1950s and 1980. On page 9-31, only forcing form volcanic erruptions is mentioned. Certainly, volcanic erruptions emit sulfate aerosols, but human activities also emit very large amount of sulfate aerosols during the 1950s and 1980s. So, it is not clear what (sulfate aerosols from volcanic erruption or aerosols from human activities) caused the global (or northern hemisphere) temperature change during the 1950s and 1980. I think this issue has to be clarified over the fifth report. [Zhaomin Wang, UK]	Noted -- But the issue raised here is properly one for the D&A chapter, not the model evaluation chaper.
9-971	9	41	18	41	18	For a more recent, CMIP5, model, a useful reference here is Haywood et al. [2011]. [Nicolas Bellouin, United Kingdom]	Taken into account -- combimed with comment 9-972
9-972	9	41	18	41	24	It might be worth referring to Haywood, J. M., N. Bellouin, A. Jones, O. Boucher, M. Wild, and K. P. Shine, The roles of aerosol, water vapour and cloud in future global dimming/brightening, Journal of Geophysical Research, 116, D20203, doi:10.1029/2011JD016000, 2011. This paper also discusses past dimming in the HadGEM2 model. [Olivier Boucher, France]	Accepted -- this paper is now added and discussed in this subsection.
9-973	9	41	20	41	20	"Ruckstahl and Norris (2009)" instead of "(Ruckstahl and Norris, 2009)" [Farahnaz Khosrawi, Sweden]	Editorial
9-974	9	41	26			Figure 9.28: the relative error in this figure needs clarification: is 1.0 a 100% overestimate by the model? [Bruce Wielicki, USA]	Accepted -- caption has been clarified.
9-975	9	41	38			Section 9.4.6.2: At some point it should be mentioned that some CMIP5 models include prognostic models of the ocean biogenic DMS source (see also 9.1.3.3). [James Christian, Canada]	Accepted -- mention of prognostic DMS has been added to section 9.4.6.2 on FOD page 9-41 starting line 56.
9-976	9	41	40	42	5	Needs to be updated with CMIP5 results and extended to all aerosol species. [Olivier Boucher, France]	Taken into account -- literature on CMIP5 aerosols is still developing and there is still relatively little to assess on aerosols.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-977	9	41	44			This says that emissions uncertainties are the dominant source of uncertainty in sulphate burdens, but 9.4.1.1.4 seemed to say that the meteorology was most important. Cross-reference these sections and resolve the apparent discrepancy. [Nathan Gillett, Canada]	Rejected -- concluding sentence in paragraph clearly states that the major differences across the ensemble are due to processes/transport not emissions.
9-978	9	41	46	41	46	Please add "of " after 53.3 % [HASIBUR RAHAMAN, India]	Editorial
9-979	9	41	47	41	47	What is AerCom? [Farahnaz Khosrawi, Sweden]	Taken into account -- AeroCom is now expanded in the text in 9.4.1.1.4 where it first appears.
9-980	9	42	5			section 9.5: this section seems to me a bit too short with respect to local interannual and multi-decadal variability (irrespective of the climate process). I find Fig. 6 of Kravtsov and Spannagle (2008, cited in page 48 line 14) and Fig. 9 of Santer et al (2011) very illuminating and they would be nice contribution to the chapter. [Ramon de Elia, Canada]	taken into account -- text revised for second order draft.
9-981	9	42	5			Section 9.5: This section should refer to the recent assessment of the IPCC SREX report (chapter 3; Seneviratne, Nicholls, et al. 2012), and highlight modifications of assessments where necessary. [Sonia Seneviratne, Switzerland]	Accepted -- revisions will include better linkages to SREX
9-982	9	42	9	48	32	Whhile in the previous sections the CMIP5 models are references several times, in these sections there is not any reference to CMIP5 models (except in ENSO section). [Claudio Cassardo, Italy]	taken into account -- available CMIP5 results and published analyses will be assessed.
9-983	9	42	9	54	53	Simulating variability is much more difficuly, even when you have concealed or played down the true varaibility of your observations. But in the end it does not prove anythiing and it is irreversably handicapped by the basic unrealistic assumptions of the models [VINCENT GRAY, NEW ZEALAND]	Rejected -- no evidence provided to support statement that variability in observations has been concealed. It is important that this chapter evaluates models' ability to simulate variability
9-984	9	42	11			Replace 'natural variability' with 'internal variability'. Natural variability includes the response to natural forcings. See the glossary definition of 'climate variability'. [Nathan Gillett, Canada]	Accepted -- thanks
9-985	9	42	23	42	23	Now the abbreviation ENSO is introduced, after it had already been used previously in the text. [Farahnaz Khosrawi, Sweden]	Editorial -- abbreviations to be summarized in a separate section of the report.
9-986	9	42	27	42	27	Respond in what sense? Greenhouse gas changes as e.g. increased concentrations? [Farahnaz Khosrawi, Sweden]	Rejected -- comment is unclear. Text refers to modelss ability to simulate response to Earth's orbit and rotation which drive diurnal and seasonal cycles, not GHGs.
9-987	9	42	29	42	29	In this section? In the last two paragraphs? I think you mean section 9.5 in general, aren't you? [Farahnaz Khosrawi, Sweden]	Rejected -- 'section' refers to the main partitioning of each chapter, e.g. section 9.1, 9.2, etc. as defined explicitly on pg. 5, lines 25-30
9-988	9	42	38	42	42	not clear, better to rewrite [ANNALISA CHERCHI, Italy]	Taken into account, paragraph rewritten
9-989	9	42	44	42	44	Skip "(T)". [Farahnaz Khosrawi, Sweden]	Editorarial. Done
9-990	9	42	46	42	52	The explanation of model deficiency based on SST alone is not quite satisfactory. Land-breeze fronts propagate up to 500 km offshore and would have an important dynamical influence on the surface air temperature over coastal seas, quite apart from locally induced diurnal cycle assocated with SST variations within a day. Reference should be made to such studies as Teo et al. (2011) for maritime Southeast Asia; Rao and Fuelberg (2000) and Nichols et al, (1991) for Florida Peninsula in North America; Noonan and Smith (1986) for Cape York Peninsula in Australia to substantiate the manifestation of such diurnal processes over land and sea. It is doubtful whether such land-sea interactions are adequately captured in the coupled AOGCMs due coarse *horizontal* resolution and of coastlines. This may perhaps be one important reason why diurnal cycles over the sea is consistently under-estimated by models. Local SST variation alone (which is much smaller than land surface temperature variation) is not likely to drive diurnal changes in surface air temperature over the sea of comparable magnitude to that over the land, as the observations clearly show in Figure 9.30. The decrease in the total length of coastlines from the northern to southern hemisphere is also consistent with the fall in amplitude of the diurnal cycle over the sea in Figure 9.30. [REFERENCES: Nichols,	Taken into account. However it is not possible to fully consider this point. The section is on large scale circulation models and diurnal cycle over the open ocean away from coastal zone. Coastal zones and not fully resolved in the GCM we discuss. This is better amphazise in the the revised document. Over the open oceans, the diurnal cycle of near surface air temperature is tightly coupled to SST dirunal cycle, whose simulation is problematic in OAGCM due to the finite depth of the top model layer in OGCM. The difference in the land/ocean area ratios, not necissarily the coastal lines, between the Northern and Southern Hemisphere contributes to the

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						M. E., R. A. Pielke, and W. R. Cotton (1991), A two-dimensional numerical investigation of the interaction between sea breezes and deep convection over the Florida Peninsula, Mon. Weather Rev.,119, 298– 323; Noonan, J. A., and R. K. Smith (1986), Sea-breeze circulations over Cape York Peninsula and the generation of Gulf of Carpentaria cloud line disturbances, J. Atmos. Sci., 43, 1679–1693; Rao, P. A., and H. E. Fuelberg (2000), An investigation of convection behind the Cape Canaveral sea-breeze front, Mon. Weather Rev., 128, 3437– 3458; Teo, C. K., T. Y. Koh, C. F. Lo, B. C. Bhatt (2011), "Principal Component Analysis of observed and modelled diurnal rainfall in the Maritime Continent", Journal of Climate, 24(17), 4662-4675.] [Tieh-Yong Koh, Singapore]	difference in the diurnal amplitudes between the two hemisphere, as the land tends to have much larger diurnal variations due to the smaller heat inertial of land than that of water surfaces. We'll also improve the link between this section and the section on high resolution and regional models.
9-991	9	42	46	42	52	The referenced studies and Figure 9.30 does not support the very strong and general statement here: "coupled models capture the overall amplitude and phase of the diurnal cycle of surface air temperature (T) well over land". The figures only show the summer midlatitudes (up to 70N) and only for one model. The diurnal cycle in winter and at high latitudes, when we frequently have stably stratified boundary layers, is a totally different story. This needs to be modified. One study that evaluates the diurnal cycle of several near surface variables using flux stations for two versions of CAM is currently under review at J. of Climate and can be provided. A study of the CMIP5 models on these variables are underway. [Gunilla Svensson, Sweden]	Taken into account. Thank you for providing your reference. The revised figure consider CMIP5 simulations. In winter at high latitudes, nighttime Tmin is mainly controlled by LW radiation which depends mainly on water vapor and clouds, while daytime Tmax is also heavily influenced by LW radiation as well as shortwave radiation (and thus clouds). The PBL's mixing is less important in winter than in summer time. AOGCMs should overall simulate these radiative and PBL processes reasonably well, although detailed analyses are still needed.
9-992	9	42	48	42	50	The reduced amplitudes and phase shifts shown by the models over land and, especially, over oceans remembers me my first experiments performed with a land surface scheme. In my simulations, such kind of errors were due to an inadequate choice of the soil layers (with a depth too large) or to inadequate values of thermal conductivity or other soil parameters. If the former reason will be responsible of the disagreement over oceans, models in which the ocean component has a larger resolution should show a better agreement with the observations. [Claudio Cassardo, Italy]	Taken into account. Indeed we know that model SST diurnal cycle will be damped due to the thickness of the top model layer of the OGCM, as the model SST is the layer-averaged T. Since this SST serves as the lower boundary condition for near surface air temperature, the diurnal cycle of the surface air temperature is also damped. Unfortunatelly, the depth of the top ocean layer can't be too small for free-surface OGCMs, unlike for soil layers. Other consideration on the ocean-atmosphere coupling frequency is also an important factor to consider in the representation of the diurnal cycle over the ocean. New references have been included
9-993	9	42				Section 9.5.2.1: the discussion in this subsection generally overlooks dynamical underpinnings of model behaviour and presents the model performance as is. More emphasis should be given to interpreting model results based on dynamical understanding which ultimately determines whether climate projections are reliable or not. [Tieh-Yong Koh, Singapore]	Taken into account. Note that the limited space allocate to this section doe not allow lengthy discussions.
9-994	9	43	1	43	8	Why? What is the reason? [Farahnaz Khosrawi, Sweden]	Taken into account- Howeverr multiple reasons could be invoked. It is the results of complex processes and difficulties of the current generation of climate model to reproduce convection properly. Physical explanation are provided in the revised version when possible. Note that the limited space allocated to this secton doesn't allow lenghty discussions.
9-995	9	43	14	43	24	In my personal experience, also the surface soil moisture has proved to be a relevant parameter influencing convective precipitations in areas with weak advection. Surface soil moisture perhaps does not influence greatly the timing of the convection, but may relevantly influence its amount (see this study: C. Cassardo, G. P. Balsamo, C. Cacciamani, D. Cesari, T. Paccagnella and R. Pelosini (2002) "Impact of soil surface moisture initialization on rainfall in a limited area model: a case study of the 1995 South Ticino flash flood" Hydrol. Process., 16, pp. 1301–1317 available here: http://ccassardo.webuda.com/paperi/2002_impact_of_soil_surface_moisture_initialization.pdf) [Claudio Cassardo, Italy]	Taken into account – note that this is an assessment, not a review, and so not all literature on the subject can be cited.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-996	9	43	15	43	25	sesitivity analysis is a very important step in modelling. A more extended description of this modelling procedure is needed, and how it is used to improve the models, so the reader can understand better the modelling process [Juan A. Blanco, Canada]	Taken into account. Precision included. Note however that due to the limited space it is not possible to fully describe the sensitivity experiments. The reader has to refer to the original publications
9-997	9	43	16	43	16	What is meant with a "super-parameterization"? [Farahnaz Khosrawi, Sweden]	Editorial. The paragraph has been rewrtitten
9-998	9	43	19	43	21	The reference seems to be wrong and therefore needs to be replaced. Peterson et al. (2009) is not a modeling paper and it doesn't contain any 'dirunal cycle' in the text. [Daehyun Kim, USA]	Accepted. References have been revised
9-999	9	43	21	43	21	Suggest adding here "Convection-permitting models, which have been used for short simulations or limited areas, have been shown to give a significant improvement in the diurnal cycle of convection. In particular, explicit representation delays the onset of convection and gives a slower decay of convection in the afternoon (Hohenegger et al, 2008; Kendon et al, 2012)." [See comment 5 for references] [Elizabeth Kendon, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Note however that this section consider the assessement of global climate models and that it is not possible to site all the relevant literature.
9-1000	9	43	26	43	27	"other diurnal variations in have …": in what? [Hai Lin, Canada]	Typo Corrected
9-1001	9	43	27	43	27	TYPO: delete "in" [Ladislav Metelka, Czech Republic]	typo corrected see 9-1000
9-1002	9	43	27	43	27	delete "in" after "variations" [Celeste Saulo, Argentina]	type corrected see 9-1000 and 1001
9-1003	9	43	27	43	34	The manifestation and underlying dynamics of atmospheric tides are different from those of diurnal cycles in humidity, cloudiness and low-level winds. Thus mentioning the diurnal variations in surface atmospheric pressure and those in the other variables within the same paragraph can be misleading as though they evaluate the model in different ways. Atmospheric tides at the surface result mainly from semi-diurnal variations originating in the ozone layer in the stratosphere and to a lesser extent in the tropospheric water vapour content. The signatures of atmospheric tide is on a larger horizontal scale than those of humidity, cloudiness and low-level winds which originate from latent and sensible heat forcing from the surface and from low-level radiative heating/cooling. The latter has much more to do with the previous paragraph on temperature and precipitation diurnal cycles. Thus, it is suggested to relegate the evaluation of surface pressure tides to a new paragraph below the current and emphasis be given to the fact that the evaluation here is on a different set of processes in the models from the comparison of the other variables. E.g. the evaluation of atmospheric tides in the models would be more indicative of large-scale tropospheric water vapour distribution, which would be an important consequence of global warming. [Tieh-Yong Koh, Singapore]	Takein into account. We agree that the underlying phsycis for surface pressure tides are not the same as for the other variables. Due to limited space, we only point out this fact in the paragraph.. Also note that surface pressure tides also include a strong diurnal (24-hr) component over land, which is related to surface heating and thus temperature.
9-1004	9	43	27	43	34	There must be more studies to refer to than Dai and Trenberth 2004 (referenced three times in eight lines and once on page 42) [Gunilla Svensson, Sweden]	Accepted. Note however that the number of publication on this subject is limited. New results have been included
9-1005	9	43	38			Moreover these, and other ISV features, represent bridges between seasonal variability and weather, and account for many types of extreme events. [Duane Waliser, USA]	Taken into accoutn. Combined with the other comments on this section
9-1006	9	43	39	43	39	"Madden Julian" needs to be corrected as "Madden-Julian". [Daehyun Kim, USA]	Typo Corrected
9-1007	9	43	39	43	40	"... that appear to play an important role in mid-latitudes and the tropics respectively" too weak and imprecise a justification. [Robert Colman, Australia]	Taken into account. Combined with the other comments on this section
9-1008	9	43	40	43	40	"Madden Julian oscillation" should be "Madden-Julian Oscillation". Also there is no consistent choice of cases; both "Madden-Julian oscillation" and "Madden-Julian Oscillatin" are found in this chapter. [Masahiro Sugiyama, Japan]	typo Corrected see comment 1006
9-1009	9	43	42	44	15	Blocking is also discussed in Sections 2.6.6.3 and 14.2.11. Cross referencing needed [George Kiladis, USA]	taken into account, cross chapter issue
9-1010	9	43	45	43	49	An accurate simulation of atmospheric blocking is important in various timescale (e.g. medium- and long-range forecasts and climate projection). Numerical weather prediction (NWP) models in the past also underestimate the frequency of blocking. This should be mentioned here. Replace "climate models" with "climate and	Taken into account. This is an important point. Note that this is an assessment,not a review, and so not all literature on the sibject can be listed

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						numerical weather prediction models". In addition, the following paper, Matsueda (2009), revealed that state-of-the-art NWP models well simulate the frequency of blocking. This should be also mentioned here. Matsueda, M., 2009: Blocking Predictability in Operational Medium-Range Ensemble Forecasts, Scientific Online Letters on the Atmosphere, 5, 113-116 [Shoji Kusunoki, Japan]	
9-1011	9	43	46	43	46	missing references, Sillmann and Croci-Maspoli, GRL, 2009 have shown a significant relationship between atmospheric blocking and climate extremes in Europe, a more detailed study on the effect of atmospheric blocking on extreme cold temperatures in Europe can be found in Sillmann et al., Journal of Climate, 2011 [Jana Sillmann, Canada]	Taken into account. Note that this is an assessment,not a review, and so not all literature on the sibject can be listed
9-1012	9	43	46	43	49	Scaife et al (2011, GRL) show that realistic atmospheric blocking can be obtained with existing standard atmospheric resolutions, but higher horizontal resolution in the ocean is required to remove SST biases. The fact that climate models are able to accurately simulate blocking without higher atmosphere resolution is an important point and should be noted. [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Note that this is an assessment,not a review, and so not all literature on the sibject can be listed
9-1013	9	43	47	43	47	"atmospheric AOCGMs" is incorrect. "atmospheric GCM" is correct. [Shoji Kusunoki, Japan]	editorial, corrected
9-1014	9	43	50			Suggest adding something after this sentence like: "While other studies have shown that ocean surface temperature errors in the extratropics are a key component of the large blocking deficits in coupled ocean atmosphere models (Scaife et al., 2011)" Reference: Improved Atlantic Blocking in a Climate Model. A.A. Scaife, D. Copsey, C. Gordon, C. Harris, T. Hinton, S.J. Keeley, A. O'Neill, M. Roberts and K. Williams 2011,Geophys. Res. Lett., 38, 23, doi:10.1029/2011GL049573)" [Adam Scaife, United Kingdom of Great Britain & Northern Ireland]	taken into account. The section has been entirely revised
9-1015	9	44	4	44	4	Why does that let appear model skills better? Are these models then performing better or not? [Farahnaz Khosrawi, Sweden]	Taken into account. It was premature in the FOD to make definive statement withour knowing the new results. The assessment has been revisited at the light of new results
9-1016	9	44	10	44	15	improvements in CMIP5? [ANNALISA CHERCHI, Italy]	Taken into account.see 9-1015
9-1017	9	44	10	44	15	This paragraph needs to be clarified and linked to the previous discussion. Define 'circulation regimes' and clarify their importance and links with blocking. [Robert Colman, Australia]	Taken into account. Note that the space limit doesn't allow lengthy discussions
9-1018	9	44	11	44	13	The regime behavior being eddy driven was shown earlier by Luo et al. (2007a,b). Luo, D., and A.R. Lupo, 2007: Dynamics of eddy-driven low-frequency dipole modes. Part II: Free mode characteristics of NAO and diagnostic study. Journal of the Atmospheric Sciences, 64, 3 - 28. Luo, D., A.R. Lupo, and H. Wan 2007: Dynamics of eddy-driven low-frequency dipole modes. Part I: A simple model of North Atlantic Oscillations. Journal of the Atmospheric Sciences, 64, 29 - 51. [Anthony Lupo, USA]	Taken into account. Note that the NAO is not part of this section and is treated in section 5.3
9-1019	9	44	16	44	16	Wang et al. (2011) demonstrated that the Community Atmospheric Model version 3.1 (CAM 3.1) can successfully simulate Rossby Wave propagation over the West Pacific as a result of the interaction of the jet stream with the Tibetan Plateau. Imposing a warm anomaly over the plateau as a result of climate warming, they also show that a positive height amomaly exists in the model over the West Pacific a result of Rossby propagation. Wang, Y., X. Xu, A.R. Lupo, P. Li, and Z. Lin, 2011: The Remote Effect of the Tibetan Plateau on Downstream Flow in Early Summer. Journal of Geophysical Research Atmospheres, 116, D19108, doi:10.1029/2011JD015979. [Anthony Lupo, USA]	Taken into account. Reason of differences between models or of model biases have been included when possible. Note that this is an assessment,not a review, and so not all literature on the sibject can be listed
9-1020	9	44	17	44	17	"Madden Julian" needs to be corrected as "Madden-Julian". [Daehyun Kim, USA]	typo Corrected see comment 1006
9-1021	9	44	17	44	36	It is disappointing to see that this part doesn't have any results from CMIP5 simulations, even though in Figure 9.43 a conclusion on the capability of CMIP5 models to simulate the MJO is implicitly given. [Daehyun Kim, USA]	Taken into account. This was a place holder for CMIP5 results that were not available at the time mof the FOD. Resutls with CMIP5 simulations are now

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							included
9-1022	9	44	17	44	36	Will this be updated using results from the CMIP5 models? [David Randall, USA]	Taken into account. New results have been added, based on the work of the MJO pannel and CMIP5 simulations
9-1023	9	44	17			Section 9.5.2.2.2.: Why do the models have problems to simulate the MJO? [Farahnaz Khosrawi, Sweden]	Taken into account. Reasons of model defiencies are provided when possible. Note that given the space limit of the section it is not possible to provide lengthy discussion.
9-1024	9	44	18	44	18	"Madden Julian oscillation" should be "Madden-Julian Oscillation". Also there is no consistent choice of cases; both "Madden-Julian oscillation" and "Madden-Julian Oscillatin" are found in this chapter. [Masahiro Sugiyama, Japan]	editorial. Correction see 9-1006
9-1025	9	44	18	44	20	In this sentence, a reference for boreal winter MJO is needed. I suggest to cite the original Madden and Julian (1972) paper here. The reference coulde be added as "known as the Madden-Julian Oscillation (MJO, Madden and Julian 1972)". And the reference is given here: Madden, R. A. and P. R. Julian, 1972: Description of Global-Scale Circulation Cells in the Tropics with a 40-50 Day Period. Journal of the Atmospheric Sciences, 29, 1109-1123. [Daehyun Kim, USA]	Taken into account. This is indeed important and was not welll done in the FOD . We also cite Zhang 2005
9-1026	9	44	18	44	37	MJO has an influence on a broad range of climate and weather phenomena. Somehow, MJO-NAO link is strongly emphasized by citing Cassou (2008). It is a nice work, but the authors should mention impacts of MJO beyond those on Atlantic weather regimes. One might want to cite a work like Zhang (2005) or a more recent review. (Zhang, C., 2005: Madeen-Julian Oscillation. Reviews of Geophysics, 43, RG2003, doi:10.1029/2004RG000158.) [Masahiro Sugiyama, Japan]	Taken into account. Note that the link with the modes are not directly part of this section but of section 5.3. The section has been revisited
9-1027	9	44	18			Section 9.5.2.2.2: It may be better to add MJO's impact on North American Monsoon. [Jialin Lin, United States of America]	Taken into account. Note that all references cannot be added due to the limited space
9-1028	9	44	19	44	19	Please add reference for MJO [HASIBUR RAHAMAN, India]	taken into account. See 9-1025
9-1029	9	44	19			This section could use some additional attention that would describe why the MJO is important enough to call out and discuss here. It can have important multi-scale interactions with ENSO, dictate onsets and breaks (i.e. variability) of the monsoons, modulate TCs/extremes. ------ Moreover, there has been work to indicate that adequate vertical resolution, mean state, and especially sensitivity of the convection scheme to environmental moisture is important to simulate the MJO properly. -------Some really important gains have been made over the last decade. There is a monograph of new results in Lau and Waliser 2nd Edition ISV book, with notable chapters on MJO modeling by Sperber et al., theory by wang, and a number of others on interactions with monsoons, oceans etc. ------ I would be happy to help improve this section if needed or re-review. [Duane Waliser, USA]	taken into account. Precision has been included. Note that the interaction between MJO and other phenomena or modes is considered in the following sections (9.5.2.2.2 and 5.3). The section has been revised at the light of the new syntheses and results available.
9-1030	9	44	19			Seems as though some general references would be in order here for those less familiar with the MJO. -----(namesake) Madden, R. A., and P. R. Julian (1994), Observations of the 40-50-Day Tropical Oscillation - a Review, Monthly Weather Review, 122(5), 814-837. -------- (describes summer vs winter characteristics, impacts and predictability) Waliser, D. E. (2006), Intraseasonal Variability, in The Asian Monsoon, edited by B. Wang, p. 844 Springer, Heidelberg, Germany. ------(review specific to winter MJO) Zhang, C. (2005), The Madden Julian Oscillation, Reviews of Geophysics, 43, RG2003, doi:10.1029/2004RG000158. [Duane Waliser, USA]	Taken into account. Note that this is an assessment,not a review, and so not all literature on the sibject can be listed
9-1031	9	44	20	44	20	"Cassou (2008; Pan and Li, 2008)" need be changed by "Cassou (2008) and Pan and Li (2008)". [Daehyun Kim, USA]	editorial. Corrected
9-1032	9	44	20	44	23	The influence of the MJO on the extratropical weather regimes is an important aspect of the MJO, but not only one. It is well known that the MJO has large impacts on a wide variety of climate phenomena across different spatial and temporal scales. Some examples include the onsets and breaks of the Indian and Australian summer monsoons (e.g. Wheeler and McBride 2005), the formation of tropical cyclones (e.g. Maloney and Hartmann 2000; Bessafi and Wheeler 2006) and the evolution of some El Nino events (e.g. Takayabu et al.	taken into account. Precisions on the importance of considering the MJO have been added. Note that this section only adress the large scale features of MJO and not all the regional implications that are considered in chapter 14. Also this is an assessment,

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						1999; Bergman et al. 2001). References are given here: Wheeler, M. C. and J. L. McBride, 2005: Australian-Indonesian monsoon. Intraseasonal Variability in the Atmosphere-Ocean Climate System, W. K. M. Lau and D. E. Waliser, Eds., Springer, Heidelberg, Germany, 125-173. Maloney, E. D. and D. L. Hartmann, 2000: Modulation of hurricane activity in the Gulf of Mexico by the Madden-Julian oscillation. Science, 287, 2002-2004. Bessafi, M. and M. C. Wheeler, 2006: Modulation of South Indian Ocean Tropical Cyclones by the Madden-Julian Oscillation and Convectively Coupled Equatorial Waves. Monthly Weather Review, 134, 638-656. Takayabu, Y. N., T. Iguchi, M. Kachi, A. Shibata, and H. Kanzawa, 1999: Abrupt termination of the 1997-98 El Nino in response to a Madden-Julian oscillation. Nature, 402, 279-282. Bergman, J. W., H. H. Hendon, and K. M. Weickmann, 2001: Intraseasonal Air-Sea Interactions at the Onset of El Nino. Journal of Climate, 14, 1702-1719. [Daehyun Kim, USA]	not a review, and so not all literature on the sibject can be listed
9-1033	9	44	21	44	21	"Cassou (2008; Pan and Li, 2008) present…" should be "Cassou (2008) and Pan and Li, 2008 have presented…" [HASIBUR RAHAMAN, India]	typo see 1031
9-1034	9	44	26	44	28	References for the impacts of convection scheme on the simulation of the MJO are missing although convection schemes have stronger impacts than coupling with ocean. References about the effect of increasing vertical resolution are also missing. There is a typo: 'highligted' need to be replaced by 'highlighted'. Therefore, I suggest to modify the sentence as "Cumulus parameterizations, coupling with ocean, and vertical resolution were highlighted as important factors contributing to model deficiencies (Wang and Schlesinger 1999; Maloney and Hartmann 2001; Lin et al. 2008; Waliser et al. 1999; Bernie et al. 2008; Inness et al. 2001). References are given here: Wang, W. Q. and M. E. Schlesinger, 1999: The dependence on convection parameterization of the tropical intraseasonal oscillation simulated by the UIUC 11-layer atmospheric GCM. Journal of Climate, 12, 1423-1457. Maloney, E. D. and D. L. Hartmann, 2001: The sensitivity of intraseasonal variability in the NCAR CCM3 to changes in convective parameterization. Journal of Climate, 14, 2015-2034. Lin, J. L., M. I. Lee, D. Kim, I. S. Kang, and D. M. W. Frierson, 2008: The Impacts of Convective Parameterization and Moisture Triggering on AGCM-Simulated Convectively Coupled Equatorial Waves. Journal of Climate, 21, 883-909. Waliser, D. E., K. M. Lau, and J.-H. Kim, 1999: The influence of coupled sea surface temperatures on the Madden–Julian oscillation: A model perturbation experiment. J. Atmos. Sci., 56, 333–358. Inness, P. M., J. M. Slingo, S. J. Woolnough, R. B. Neale, and V. D. Pope, 2001: Organization of tropical convection in a GCM with varying vertical resolution; implications for the simulation of the Madden–Julian oscillation. Climate Dynamics, 17, 777–793. [Daehyun Kim, USA]	Taken into account. This is an important point. Thank you for the references. Note however that this iis an assessment, not a review, and so not all literature on the sibject can be listed
9-1035	9	44	26	44	28	The Bernie et al. reference is not the most representative one. The MJO is largely rooted in the atmosphere, although air-sea coupling will have important effects on it. Convection parameterization is the single most important factor responsible for the model deficiencies in MJO simulation, and Bernie et al. paper was more on diurnal air-sea coupling and its effect on model simulation. The papers by Liu et al. (2005) and Waliser et al. (1999) are better references. [Guang Zhang, United States of America]	Taken into account. Combined with comment 9-1034
9-1036	9	44	28	44	30	This sentence can be misleading, because here a boreal winter MJO and boreal summer ISO is compared with each other. Specifically, in Lin et al. (2006), Kim et al. (2009), and Xavier et al. (2010), boreal winter MJO is investigated, it is boreal summer ISO that is primarily examined in Sperber and Annamalai (2008). A comparison between two is not fair. Therefore, the findings of Sperber and Anamalai (2008) should be mentioned in a separate sentence or be removed. In my view, the simulation of boreal winter MJO, which is characterized by its eastward propagation, is still challenging. One of the reasons for the difficuly is the way of changing convection scheme to improve the simulation of the MJO degrades the mean state (Kim et al. 2011, reference is in the original document - Kim, D., A. H. Sobel, D. M. W. Frierson, E. D. Maloney, and I.-S. Kang, 2011: A Systematic Relationship between Intraseasonal Variability and Mean State Bias in AGCM Simulations. J. Climate, 24, 5506-5520.). [Daehyun Kim, USA]	Taken into account. The text has been revised and better highlight summer and winter conditions. New results are also included. Note however that this iis an assessment, not a review, and so not all literature on the sibject can be listed
9-1037	9	44	29	44	29	In the brackets I would add a reference to Lin et al 2009, who focus their analysis in South America. Intraseasonal Variability Associated with Summer Precipitation over South America Simulated by 14 IPCC AR4 Coupled GCMs Jia-Lin Lin, et al, 2009. Monthly Weather Review [Celeste Saulo, Argentina]	Taken into account. Not that it is not possible to consider the implications for all the regions. This is treated in chapter 14.
9-1038	9	44	30	44	36	This is inconsistent with the results of Lin et al. (2008a) , who evaluated the Boreal Summer Intraseasonal Oscillation (BSIO) in 14 IPCC AR4 coupled GCMs. They found that most models have difficulty in simulating	Taken into account. Reasons of model defiencies are provided when possible. Note that given the space

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						the eastward propagation of BSIO, but tend to simulate well the northward propagation of BSIO. This is also the case for MJO associated with North American Monsoon, which has been evaluated by Lin et al. (2008b) for the same 14 AR4 models. They found that most models simulate poor eastward propagation of MJO. Lin, J. L., K. M. Weickmann, G. N. Kiladis, B. E. Mapes, S. D. Schubert, M. J. Suarez, J. T. Bacmeister, and M.-I. Lee, 2008a: Subseasonal variability associated with Asian summer monsoon simulated by 14 IPCC AR4 coupled GCMs. J. Climate, 21, 4541-4567. Lin, J. L., B. E. Mapes, K. M. Weickmann, G. N. Kiladis, S. D. Schubert, M. J. Suarez, J. T. Bacmeister, and M.-I. Lee, 2008b: North American monsoon and convectively coupled equatorial waves simulated by IPCC AR4 coupled GCMs. J. Climate, 21, 2919-2937. [Jialin Lin, United States of America]	limit of the section it is not possible to provide lengthy discussion and list all the suggested references.
9-1039	9	44	32	44	36	I think the Figure 9.32 is not adequate as only one MJO plot, because it is not intuitive and not easy to understand for general readers, who are not familiar with the MJO. Adding figures of OLR patterns associated with PCs will be more helpful. [Daehyun Kim, USA]	Taken into account. The figure has been revised at the light of new results.
9-1040	9	44	40	44	40	TYPO: "Empirical" instead of "Empiricol" [Ladislav Metelka, Czech Republic]	typo. Corrected
9-1041	9	44	42	44	42	Skip the abbreviation PC. [Farahnaz Khosrawi, Sweden]	editorial. Done
9-1042	9	44	48	44	48	"Sampling variability": please specify better for unexpert readers which sampling and which variability is intended. [Claudio Cassardo, Italy]	taken into account. Text has been revised at the light of new restuls
9-1043	9	44	50	45	33	More detailed discussion of monsoons is contained in Chapter 14, and also mentioned in Chapter 10, cross references needed. [George Kiladis, USA]	taken into account. References to these chapters have been added
9-1044	9	44	50			For evaluation of CMIP3 model performance for Australian monsoon, could cite Colman, R.A., A.F. Moise and L.I. Hanson (2011), Tropical Australian climate and the Australian monsoon as simulated by 23 CMIP3 models, Journal of Geophysical Resarch, 116, D10116, doi:10.1029/2010JD015149. [Josephine Brown, Australia]	Taken into account. Note that this section is not supposed to provide an assessment of model results in all regions. Regional phenomena are part of chapter 14.
9-1045	9	44	52	44	52	change "The global" with "In a global view," [ANNALISA CHERCHI, Italy]	rejected. Se refer here to the definition of global monsoon as provided by Wang and Ding (2008)
9-1046	9	44	53	44	56	Separate the sentence into two sentences. "a models" needs to be correced as "a model" or "models". "ability to simulate the monsoon domain and its intensity" is not specific. So it is not easy to understand what does this mean. "by (Wang and Ding, 2008)" needs to be corrected as "by Wang and Ding (2008)". [Daehyun Kim, USA]	Taken into account. The paragraph has been revised.
9-1047	9	44	53	44	58	Should cite here the recent papers by Fan et al (2009;2010;2012) which introduce the concept of a multivariate South Asian Summer Monsoon (SASM) index and apply to CMIP3 historical simulations and future projections, defining an integrative monsoon index ("IMI") in terms of of multiple fields which they compare against other currently used indices, and explicity differentiating between dynamical and precipitation-related features of the SASM (since these two different measures can potentially diverge because of different relevant physics): Fan, F., Mann, M.E., Ammann, C.M., Understanding Changes in the Asian Summer Monsoon over the Past Millennium: Insights From a Long-Term Coupled Model Simulation, J. Climate, 22, 1736-1748, 2009; Fan, F., Mann, M.E., Lee, S., Evans, J.L., Observed and Modeled Changes in the South Asian Summer Monsoon Over the Historical Period, J. Climate 23, 5193-5205, 2010; Fan, F., Mann, M.E., Lee, S., Evans, J.L., Future Changes in the South Asian Summer Monsoon: An Analysis of the CMIP3 Multi-Model Projections, Journal of Climate (in press) [Michael Mann, USA]	taken into account and Fan et al 2010 added. Note that this an assessment, not a review, and so not all literature on the subject can be listed. Note that the millennium is discussed in chapter 5 and 10
9-1048	9	44	55	44	56	"…by (Wang and Ding, 2008)" --> "…by Wang and Ding (2008)" [Hai Lin, Canada]	editorial. Corrected
9-1049	9	45	4	45	4	insert "global" between "The" and "monsoon" [ANNALISA CHERCHI, Italy]	editorial. Corrected
9-1050	9	45	4	45	4	I think it will be nice to add (June-September) after monsoon precipitation [HASIBUR RAHAMAN, India]	editorial. Included
9-1051	9	45	4	45	14	Fan et al (2010) [Fan, F., Mann, M.E., Lee, S., Evans, J.L., Observed and Modeled Changes in the South	Taken into account. The paragraph has been revised

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Asian Summer Monsoon Over the Historical Period, J. Climate 23, 5193-5205, 2010] provide an alternative ranking of the skill of CMIP3 models in reproducing the observed South Asian Summer Monsoon (SASM) based on comparisons over the modern historical period of both mean and interannual variance in Monsoon-related precipitation. [Michael Mann, USA]	at the light of these results and of new results from CMIP5 simulations that were not available at the time of the FOD. Note that the regional phenomena are considered in chapter 14 and not in this section that concentrats on large scale features.
9-1052	9	45	5	45	5	I can not tell the grey-shaded contours in Fig. 9.33. Are you referring to the thin black line? [Celeste Saulo, Argentina]	taken into account. The figure has been revised at the light of new results.
9-1053	9	45	9	45	9	What is the reason for the underestimation? [Farahnaz Khosrawi, Sweden]	taken into account. Reasons for model failure are provided where possible. Note that the space limit doesn't allow for lenghty explanation.
9-1054	9	45	11			the phrase .."between the best and poorest CMIP3 models…," implies an absolute metric. More precise language is needed for the specific context of this paragraph. [David Bader, USA]	taken into account. The paragraph has been entirely rewrittent at the light of new results
9-1055	9	45	17	45	17	I think it will be nice to add (June-September) after Monsoon precipitation [HASIBUR RAHAMAN, India]	Accepted.
9-1056	9	45	17			Seems a central element is missing in this section that monsoon variability in part is made up by getting intraseasonal variations correctly (papers by Sperber et al., Waliser et al., Krishnamurti et al. (cola)) and possibly even diurnal cycle (e.g. webster et al. (jasmin) study). ------ Thus evaluation of the monsoon will necessarily have be be done and understood through multiple time scales. [Duane Waliser, USA]	Taken into account. The text has been revised accordingly.
9-1057	9	45	24	45	30	What is the reason for the underestimation of precipitation in the model simulations and why is a large spread in the results found? [Farahnaz Khosrawi, Sweden]	Taken into account. Reasons for model failure are provided where possible. Note that the space limit doesn't allow for lenghty explanation, and that the answer to this particular question is not straightforward and is at the art of several scientific studies.
9-1058	9	45	30	45	30	delete "in" after "except for" [Celeste Saulo, Argentina]	editorial.done
9-1059	9	45	41	45	41	few weeks? But this section is titled interannual to centennial, and blocking was covered earlier. [Robert Colman, Australia]	Taken into account. Scope of section now clearer and text revised accordingly
9-1060	9	45	41	45	41	Why the "a few weeks" time scale variability is in the "Interannual-to-Centennial Variability" section? [Daehyun Kim, USA]	Taken into account. Scope of section now clearer and text revised accordingly
9-1061	9	45	48	45	48	Temperature [Graham Feingold, United States of America]	Editorial - Typo corrected
9-1062	9	45	54	45	54	Misspelling in "forcings" [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-1063	9	45	58	46	1	I think volcanoes also contribute a significant fraction of forced preindustrial variability. I think Gabi Hegerl has published on this. [Nathan Gillett, Canada]	Taken into account. However we do not discuss here the response to individual forcing which is node in chapter 10
9-1064	9	45				Section 9.5.3 – a map showing the spread in magnitude of interannual and/or decadal variability across the pre-industrial control integrations would be valuable to enable comparison of the levels of variability. [ED HAWKINS, United Kingdom of Great Britain & Northern Ireland]	Taken into account. We included a panel showing zonal mean variance of CMIP5 model variability from Jones et al. 2012 .
9-1065	9	46	3	46	3	"consistent with an underestimation of the carbon fluxes by the interactive carbon cycle" vague. Consistent with an underestimation of the cumulative land and/or ocean uptake? [James Christian, Canada]	Taken into account. The section has been entirely revised
9-1066	9	46	3			I don't see that carbon fluxes are that relevant here. [Nathan Gillett, Canada]	Taken into account. Text corrected in the light of new results
9-1067	9	46	8			Figure 9.37: I can't make sense of this figure. The panels are not numbered and from the caption it appears that the top one is (b) and the lower is (a). It should be more explicitly stated that "long pre-industrial simulations" are control simulations with constant solar forcing and GHG concentrations. [James Christian,	Taken into account. Figure and legend were updated with new results

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Canada]	
9-1068	9	46	8			Figure 9.34: Need better clarification of anthropogenic + natural vs natural only. [Bruce Wielicki, USA]	Taken into account. Figure updated with new results and requested clarification included in the legend
9-1069	9	46	13	46	15	Solanki et al. (2004) do not attempt to reconstruct solar irradiance. Therefore, I am wondering why this reference is quoted here. [Raimund Muscheler, Sweden]	Taken into account. Figure update with new results and reference to the right publications.
9-1070	9	46	18	51	54	Sect.9.5.3: Discussion of modes v. teleconnections should be much more balanced - less on modes (which are only one part/aspect of interannual-decadal variability), and much more on teleconnections (of which there are many more important and studied examples that should be included). [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Taken into account. More is included in terms of teleconnections and a Contributing Author has been brought in.
9-1071	9	46	18	53	4	The connection to the relevant sections of Chapter 2 could be stronger. This is a somewhat different list and different order of discussion. For coherence, strive for the same list. [Philip Mote, USA]	Taken into account. More reference observations taken from Chap 2. List of phenomena and modes cannot be exactly aligned on Chap 2 as Chap 9 has also to address the needs for Chaps 10,11,12 13 and 14.
9-1072	9	46	18			Section 9.5.3.2: this is a very comprehensive and well-written section. The discussion of the biases in the eddy driven jets is somewhat buried due to the title. Suggest changing to 'Extratropical circulation and modes' [Julie Arblaster, Australia]	Taken into account - title revision has been considered.
9-1073	9	46	20	46	21	It might make more sense to put this section after 9.5.3.5 which first discusses teleconnections in a more general way. In any event, many of the modes discussed in the following pages are also defined in Chapter 2 and discussed in Chapter 14 as well. Cross referencing of the observational aspects of some of this material would save space here. [George Kiladis, USA]	Taken into account. More emphasis is put on teleconnections in 9.5.3. Coordination with Chaps 2 and 14 is improved.
9-1074	9	46	20	46	22	Either add NAO, NAM and SAM in parantheses to introduce this abbreviation or don't use them at all. [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified
9-1075	9	46	24	46	24	Please change "Gerber et al. (2008) confirmed the AR4" to "Gerber et al. (2008) confirmed in the AR4" [HASIBUR RAHAMAN, India]	Rejected. Incorrect suggestion
9-1076	9	46	32			The meaning of a 'broader annual cycle' is not clear to me. Should this be 'weaker annual cycles'? [Nathan Gillett, Canada]	Taken into account. Actually there are a few things wrong with the seasonal cycles. Text changed to: 'and had difficulty simulating the annual cycle of annular mode timescales.
9-1077	9	46	40	46	40	Space between "the" and "AR4" is missing. [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified
9-1078	9	46	40	46	40	theAR4 => the AR4 [Daehyun Kim, USA]	Taken into account. Text modified
9-1079	9	46	40	46	40	"theAR4" --> "AR4" [Hai Lin, Canada]	Taken into account. Text modified
9-1080	9	46	40	46	40	Please change "in theAR4" to "in the AR4" [HASIBUR RAHAMAN, India]	Taken into account. Text modified
9-1081	9	46	40	46	42	It would appear from this that "observed … variability in the NAO/NAM, in particular … strong trend" should be understand as part of anthropogenic climate change. Is this really so? If not, than climate models would in general not be expected to reproduce the specific variability and trend. [Tor Eldevik, Norway]	Taken into account. Text clarified: 'in particular variations as strong as the positive trend over the latter half of the 20th century'.
9-1082	9	46	40	46	54	Note that the NAO trend has weakened in recent years and is no longer significant compared to simulated internal variability - see figure 10.11. This should be mentioned. [Nathan Gillett, Canada]	Taken into account. Coordination has been made with Chap 10
9-1083	9	46	42			Gillett (2005) also showed this. [Nathan Gillett, Canada]	Taken into account. This is an assessment and not a review so there is limited place for extra references
9-1084	9	46	51	46	51	add "," before "model projections" [Hai Lin, Canada]	Taken into account. Text modified

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1085	9	46	56	46	58	I do not think that a different representation of the stratosphere would necessarily change the sign of the NAM response. For example, Karpechko and Manzini (2012) found that two versions of the same model, one with and another without a well-resolved stratosphere, both simulate a shift towards the positive NAM phase in response to doubling CO2, with a smaller (less positive) shift in the stratospheric model version, consistent with Scaife et al. (2011). It seems to be consistent with CMIP5 models which project SLP decreases in high-latitudes and increases in mid-latitudes (see Chapter 12, Section 12.4.4.1). The presence of stratosphere-resolving models in CMIP 5 does not change the sign of the trend. I do agree that the NAM response is quite uncertain, but still suggest being more careful with wording in this sentence. [Alexey Karpechko, Finland]	Taken into account. Text modified
9-1086	9	46	56	46	58	Suggested rephrasing: 'Furthermore, the representation of the stratosphere seems to significantly influence the NAM response to anthropogenic forcing (Scaife et al., 2011, Karpechko and Manzini, 2012) with some studies finding a negative NAM response in the stratosphere-resolving models (Morgenstern et al., 2010a).' Missing reference: Karpechko, A. Y. and E. Manzini (2012), Stratospheric influence on tropospheric climate change in the Northern Hemisphere, J. Geophys. Res., doi:10.1029/2011JD017036, in press. [Alexey Karpechko, Finland]	Taken into account. Text modified
9-1087	9	46	56			Morgenstern et al. is refering to Chemistry-Climate Models (CCMs) only. These are not climate models in a strict sense since they mostly are not coupled to an ocean. A CCM is an AGCM coupled with a chemistry model (so far mostly stratospheric chemistry and only a simplified tropospheric chemistry) and prescribed sea surface temperatures. [Martin Dameris, Germany]	Taken into account. Text modified to just just write 'Chemistry-Climate Models' in line 55
9-1088	9	47	5	47	10	I suggest to clarify the concept of skewness for unexpert readers, by adding an explicative sentence. "... has pronounced negative skewness (i.e. the majority of values are larger than the mean with respect to a normal distribution, even if few values can be very low)", or something like this. [Claudio Cassardo, Italy]	Rejected. Skewness is a standard statistical term for the level of reading of IPCC reports.
9-1089	9	47	12	47	13	What is the reason for this deficiencies? [Farahnaz Khosrawi, Sweden]	Noted There is no simple answer to this question so it was left out of the assessment. Text revised to make this point
9-1090	9	47	21	47	27	include references from Fogt et al., 2009; Vera and Silvestri 2009; and Silvestri and Vera (2009) this about interdecadal variability of SAM. The last reference is "Silvestri, Gabriel, Carolina Vera, 2009: Nonstationary Impacts of the Southern Annular Mode on Southern Hemisphere Climate. J. Climate, 22, 6142–6148." [ANNALISA CHERCHI, Italy]	Taken into account. Text updated with new ref
9-1091	9	47	21	47	27	Christian et al 2010 examined the correlation of ocean-atmosphere CO2 flux with the SAM index and found that the spatial structure generally agreed well with the results of an experiment where an ocean-only OBGC model was forced with reanalysis, i.e., the coupled model seems to do a generally good job of simulating the SAM effect on regional carbon fluxes. [James Christian, Canada]	Taken into account. Text updated with new ref
9-1092	9	47	21	47	27	This seems to be a rather brief treatment of the SAM. Could this be expanded? The SAM is important for future SH climate evolution, the Southern Ocean and carbon cycle, so validating its behaviour is worthwhile. [Nathan Gillett, Canada]	Taken into account, expansion limited to space available and existing or submitted literature
9-1093	9	47	21	47	27	Why? Are these model problems understood? [Farahnaz Khosrawi, Sweden]	The model deficiencies relating to this are not sufficiently well understood to be assessed here
9-1094	9	47	31			Section 9.5.3.3.1: The idea of the interior circulation as "density-driven" is not strictly true, and the strength of the AMOC depends on rates of diapycnal mixing in the main thermocline (Munk and Wunsch 1998, DSR I 45: 1977; Wunsch and Ferrari 2004 ARFM 36: 281) that are likely too high in the models. [James Christian, Canada]	Taken into account. To avoid ambiguity, 'density-driven component' is replaced by 'mid-ocean geostrophic component' throughout the section.
9-1095	9	47	48	47	48	change "AMOC variability that" with "that AMOC variability" [ANNALISA CHERCHI, Italy]	Taken into account. Text modified
9-1096	9	47	48	47	50	If the meridional coherence of the AMOC is to be discussed, the following recent results should be referenced as a caveat to the statement that "models have suggested AMOC variability that is specific to individual ocean gyres": The latitudinal coherence of internal MOC variability in HadCM3 , HadGEM1, and HadGEM2-ES is timescale dependent. On timescales >10 years, AMOC variability is dominated by basin-wide (north-to-south coherent) modes which account for 50-60% of total variance. On timescales less than ten years, dominant	Taken into account. This is also what Baehr et al 2009 showed: an AMOC weakening ca be identified across gyre boundaries. Reference added with the following sentence: Multidecadal AMOC weakening during the 21st

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						modes are dipoles with centres of action in the sub-polar and sub-tropical gyres. Changes to the AMOC streamfunction in response to GHG warming seem to project onto the multidecadal coherent mode of variability. See figure 5 in Roberts and Palmer (2012), accepted to Climate Dynamics, “Detectability of changes to the Atlantic Meridional Overturning Circulation in the Hadley Centre Climate Models”, DOI : 10.1007/s00382-012-1306-3. [Chris Roberts, Uk]	century appears to be coherent across gyres (Baehr et al., 2009; Roberts and Palmer, 2012). Reference: Roberts and Palmer (2012), Detectability of changes to the Atlantic Meridional Overturning Circulation in the Hadley Centre Climate Models , DOI : 10.1007/s00382-012-1306-3. accepted to Climate Dynamics.
9-1097	9	47	52			Section 9.5.3.3.2: The term "Atlantic Multidecadal Oscillation" is possibly misleading and is being superceded by "Atlantic Multidecadal Variability" (e.g., Vincze and Janosi 2011 Nonlin Proc Geophys 18: 469). A really good recent data set documenting AMV for the last 800 years is Winter et al 2011, Earth and Planetary Science Letters 308: 23. Vincze and Janosi do not accept that AMV is a periodic phenomenon, and the data of Winter et al also do not appear to support this hypothesis. [James Christian, Canada]	Noted. Both terminologies have been used in the literature published since the AR4 and continue to be used. The section title makes the equivalence clear. Chapter 14 on 'Climate Phenomena and their Relevance for Future Regional Climate Change' has a more in-depth description of the AMO/AMV phenomenon, and discusses the issue of nomenclature. The text here refers to the relevant section.
9-1098	9	47	53	47	57	Although the observed 20th century AMO cycle was indeed that of 60-70 years, the ice core evidence suggests a pre-20th century dominant cycle of about 20 years (Chylek et al 2011) which is also observed in the AOGCM simulations (HadCM3 and GFDL CM2.1) and supported by an conceptual ocean model (Frankcombe et al 2010). Frankcombe, L., A. von der Heydt, and H. Dijkstra, 2010: North Atlantic multidecadal variability: An investigation of dominant time scales, J. Clim., 23, 3626-3638, doi:10.1175/2010GCL13471.1. Chylek, P., C. Folland, H. Dijkstra, G. Lesins, and M. Dubey, 2011: Ice-core data evidence for a prominent near 20 year time-scale of the Atlantic Multidecadal Osciallation, Geophysical Research Letters, 38, L13704, doi:10.1029/2011GL047501 [Petr Chylek, USA]	Noted. The AMO concept originated from the apparent multidecadal fluctuations in the North Atlantic instrumental temperature record. North Atlantic variability at shorter, interdecadal timescales has long been proposed but is considered here to be separate from the AMO. The character and mechanisms of such variability are likely to be different to that of multidecadal fluctuations (see e.g. Dong & Sutton (2005), 'Mechanism of Interdecadal Thermohaline Circulation Variability in a Coupled Ocean–Atmosphere GCM', J. Climate, 18, 1117–1135.
9-1099	9	47	53	48	3	Given the length of the observed record, is it really possible to distinguish an observed period of 70 years from simulated timescales of 40-100 years? I doubt it. [Nathan Gillett, Canada]	Taken into account. The key point being made here is that models disagree with each other, rather than with the very uncertain estimate from a comparatively short observational record. The text has been modified to clarify this.
9-1100	9	48	12	48	12	"unforced simulation": results in previous paragraph and in fig9.35 are from unforced experiments? if yes, better to specify clearly [ANNALISA CHERCHI, Italy]	Taken into account. Text modified to clarify this
9-1101	9	48	12	48	13	Include "that" after "indicate". [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified
9-1102	9	48	12	48	21	Cite here Mann and Emanuel (2006) [Mann, M.E., Emanuel, K.A., Atlantic Hurricane Trends linked to Climate Change, Eos, 87, 24, 233-241, 2006] who argued that definitions of the AMO which involve a simple linear detrending of SST data suffer from a misallocation of forced and internal variability, and that the main changes in certain features sometimes attributed to the AMO, e.g. tropical Atlantic summer temperature trends over the past century, aremore likely radiatively forced, with a prominent role of aerosol forcing. [Michael Mann, USA]	Rejected. The definition of the AMO is established and discussed in Chap 2 so this comment should belong there. Chap 09 is about model evaluation.
9-1103	9	48	12	48	24	Otterå et al. (2010; Nature Geoscience, doi:10.1038/ngeo955) would be a useful and complementary additon here. They find as far as I understand it that volcanic erruptions lead change in AMO leads change in AMOC. [Tor Eldevik, Norway]	Taken into account. Reference cited.
9-1104	9	48	12	48	24	The entire paragraph should be improved since as it reads now it is quite difficult to follow. [Farahnaz Khosrawi, Sweden]	Taken into account. Some rewriting has been done to shorten sentences and improve sentence structure.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1105	9	48	12	48	24	Some recent observational results of AMOC variability since about 1980 (e.g., Longworth, H. R., Bryden, H. L. & Baringer, M. O. (2011), Historical variability in Atlantic meridional baroclinic transport at 26.5°N from boundary dynamic height observations, Deep Sea Research Part II Topical Studies in Oceanography, 58, 1754-1767) should be cited. The given reference found that the AMOC become weakened after about 1980, so this work does not support that the AMO is caused by the variability of the AMOC. [Zhaomin Wang, UK]	Rejected. This chapter concerns climate modeling. The work mentioned in this comment is relevant to Chap 3.
9-1106	9	48	19	48	19	Than what? [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified
9-1107	9	48	20	48	20	By what? [Farahnaz Khosrawi, Sweden]	Rejected. By anthropogenic tropospheric aerosols. It seems clear given the context that this is what is being referred to.
9-1108	9	48	36	48	36	Write also here Atlantic Meridional Mode (AMM) as it is done in the section title instead of only AMM. [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified
9-1109	9	48	37	48	37	What is the Atlantic Nino? [Farahnaz Khosrawi, Sweden]	Taken into account. Atlantic Nino refers to anomalous warm events similar to El Nino in the Pacific, which occur sporadically in the eastern equatorial Atlantic Ocean.
9-1110	9	48	40	48	40	suggested to add following reference about AMM: [Zeng-Zhen HU, USA]	Taken into account, expansion limited to space available
9-1111	9	48	40	48	40	Hu, Z.-Z. and B. Huang, 2006: Physical processes associated with tropical Atlantic SST meridional gradient. J. Climate, 19(21), 5500-5518. [Zeng-Zhen HU, USA]	ref associated to 9-1110
9-1112	9	48	42	48	42	What is EOF? [Farahnaz Khosrawi, Sweden]	Noted. EOF stands for Empirical Orthogonal Function analysis which is widely used in analyzing climate data. This will be in list of abbreviations.
9-1113	9	48	47	48	54	3 uses of 'severe' to describe Atlantic based errors, a word not used much elsewhere. This is an imprecise and not particularly useful word, which acts like an exclamation mark. I suggest instead a more quantitative description -- e.g. in line 53 say how much variance is overestimated by. [Robert Colman, Australia]	Taken into account. Text modified. We changed the sentence to "... one over-estimates the Atl-3 SST variance by a factor of nearly four, while the other underestimates it by 40%"
9-1114	9	48	57	48	57	suggested to add following reference before "Richter and Xie, 2008": [Zeng-Zhen HU, USA]	Taken into account. Text updated with new ref
9-1115	9	48	57	48	57	Huang, B., Z.-Z. Hu, and B. Jha, 2007: Evolution of model systematic errors in the tropical Atlantic basin from the NCEP coupled hindcasts. Clim. Dyn., 28 (7/8), 661-682, DOI: 10.1007/s00382-006-0223-8. [Zeng-Zhen HU, USA]	noted. ref associated to 9-1114
9-1116	9	49	4	50	5	Observed changes in ENSO also covered in Chapter 2, cross reference ahead to Chapter 10.3.3.2 needed [George Kiladis, USA]	Taken into account. More coordination and cross-reference with Chaps 2, 10 and 14
9-1117	9	49	12	49	13	There are additional references that fit into this statement. In Kim et al. (2008), the impacts of cumulus momentum transport on the simulation of ENSO in an AOGCM is examined, and a sensitivity of ENSO amplitude to a parameter in the convection scheme, which is same to what is tuned when people wants to improve the representation of the MJO, is investigated in Kim et al. (2011). References: Kim, D., J.-S. Kug, I.-S. Kang, F.-F. Jin, and A. Wittenberg, 2008: Tropical Pacific Impacts of Convective Momentum Transport in the SNU Coupled GCM. Clim. Dyn., 31, 213-226. Kim, D., Y.-S. Jang, D.-H. Kim, Y.-H. Kim, M. Watanabe, F.-F. Jin, and J.-S. Kug, 2011: ENSO Sensitivity to Cumulus Entrainment in a Coupled GCM. J. Geophys. Res., 116, D22112, doi:10.1029/2011JD016526 [Daehyun Kim, USA]	Taken into account, expansion limited to space available
9-1118	9	49	14	49	14	Concerning what? Which of the above mentioned processes? [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified to clarify this
9-1119	9	49	18	49	20	"The amplitude of El Nino in AOGCMs ranges from less than half to more than double the observed amplitude (AchutaRao and Sperber, 2006; Guilyardi, 2006; Guilyardi et al, 2009b; van Oldenborgh et al, 2005)..." See also Johnson, F., Westra, S., Sharma, A. & Pitman, A.J., 2011, "An Assessment of GCM Skill in Simulating	Taken into account, expansion limited to space available

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Persistence across Multiple Time Scales.", Journal of Climate, Vol. 24, DOI: 10.1175/2011JCLI3732.1. [Seth Westra, Australia]	
9-1120	9	49	29	49	33	see also DiLorenzo et al 2010. Central Pacific El Niño and decadal climate change in the North Pacific Ocean. Nature Geosci 3: 762-765 [James Christian, Canada]	Noted. The cited reference does not provide results about model evaluation of relevance to this section
9-1121	9	49	35			Figure 9.36: need a better way to show the difference between models and observations. thicker lines for obs? [Bruce Wielicki, USA]	Taken into account. Figure updated
9-1122	9	49	42	49	43	"….Adapted from AchutaRao and Sperber (2006) and Sperber " Year is missing in the reference [HASIBUR RAHAMAN, India]	Editorial. Corrected
9-1123	9	49	50	49	56	Some CMIP3 models have no ENSO at all - this should be mentioned. [Nathan Gillett, Canada]	Taken into account. Text modified
9-1124	9	50	8	50	24	What about CMIP5 models for IO? [Claudio Cassardo, Italy]	Taken into account. CMIP5 model assessments included as they become available in literature
9-1125	9	50	8	50	24	The IOD is defined in Chap. 2 Box 2.4 and a map is shown in Box 2.4 Fig. 2, although it is called the "DMI" there. This needs to be reconciled. [George Kiladis, USA]	Taken into account. Coordination has been made with Chap 2
9-1126	9	50	11	50	11	"...Hermes and Reason, 2005)..." => "...Hermes and Reason, 2005), as well as response to the Southern Annular Mode (Huang and Shukla, 2008)" [Zeng-Zhen HU, USA]	Taken into account. Considered in revision
9-1127	9	50	11	50	11	Huang, B., and J. Shukla, 2008: Interannual variability of the South Indian Ocean in observations and a coupled model. Ind. J, Mar. Sci., 38, 13-34. [Zeng-Zhen HU, USA]	Taken into account. Considered in revision
9-1128	9	50	26	50	39	The representation of the QBO in models with different top heights is discussed in Osprey et al. "Stratophseri variability in 20th century CMIP5 simulations of the Met Office climate model: High-top versus low-top" submitted to J. Climate [William Collins, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Comment added on this point.
9-1129	9	50	26	50	39	There are several more recent publications available! [Martin Dameris, Germany]	Noted.
9-1130	9	50	26	50	39	Do any of the CMIP5 models have a QBO? If so, name them. [Nathan Gillett, Canada]	Taken into account. Some of the CMIP5 models have a QBO, we are not sure about all the models but a comment is now included and examples are given
9-1131	9	50	26	50	39	This sub-session about QBO is in a wrong place, as 9.5.3.4 is about Indo-Pacific Modes [Hai Lin, Canada]	Taken into account. This is a fair point we will move it to a new 9.5.3.5 section
9-1132	9	50	26			Section 9.5.3.4.3: There should be mention here of the high-top models, not just those with different resolution. Models with a raised lid have been shown to represent the QBO far better than those with lower lids, even with relatively poor stratospheric resolution. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Model domain is now included.
9-1133	9	50	27	50	39	As all models have been categorized into CMIP3 or CMIP5, could you please specify also here which category of models is considered? [Claudio Cassardo, Italy]	Taken into account. Text clarified
9-1134	9	50	28	50	29	The statement "Some of these employ high vertical and horizontal resolution (Takahashi, 1999; Kawatani et al., 2011)" is wrong, as Takahashi used a T42 model, i.e. not high resolution. Takahashi 1999 did not use a GWD parameterization, but decreased on purpose the damping rate in the horizontal diffusion scheme to boost the resolved wave spectrum (which had other side effects). So only Kawatani has perfomed QBO simulations at high resolution without the need of parameterized wave driving. [Marco Giorgetta, Germany]	Taken into account. Takahashi 1999 used relatively high VERTICAL resolution of 500m compared to other models. We have added the word respectively to the reference list and added the point about diffusion though.
9-1135	9	50	30	50	31	... the need for such high resolution (Scaife et al., 2000; Giorgetta et al., 2002 and 2006; McLandress, 2002). ... Giorgetta et al. 2006 also discuss the sensitivity of the simulated QBO to changes in horizontal and vertical resolution and parameterized forcing by gravity waves. [Marco Giorgetta, Germany]	Taken into account. Reference cited.
9-1136	9	50	41	50	54	one of the few references on North Pacific variability is Furtado et al (2011). It should be included. The reference is: "Furtado, J.C., Di Lorenzo, E., Schneider, N. and Bond, N.A. 2011. North Pacific decadal	Taken into account. Reference cited.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						variability and climate change in the IPCC AR4 models. Journal of Climate 24: 3049-3067" [ANNALISA CHERCHI, Italy]	
9-1137	9	50	42	50	42	Again, the introduction of the abbreviation comes after it already had been used. [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-1138	9	50	42	50	44	"globally-averaged SST anomalies have been subtracted" I don't think this is part of the definition. Mantua et al 1997 only say that they normalized their data by subtracting the means for 1947-1995. [James Christian, Canada]	Taken into account. Mantua confirms that the global mean SST anomalies were subtracted in his PDO definition (Mantua, personal communication dated July 10, 2012)
9-1139	9	50	42	50	54	What are the reasons for the differences/model deficiencies? [Farahnaz Khosrawi, Sweden]	Taken into account. We don't have space to add this information; however, there are no established reasons for the model deficiencies in the peer-reviewed literature.
9-1140	9	50	42	53	4	What about CMIP5 models? [Claudio Cassardo, Italy]	Taken into account. There are no published studies yet for CMIP5.
9-1141	9	50	48	50	54	Annex I, Figure AI.16 The historical ensemble models in the figure show no evidence of the actual winter (DJF) PDO signal in Alaska and western North America over the past century, and show no climate variation of the proper amplitude (2-4C) and time scale (~60 year periodicity) for the region. The PDO is the dominant contributor to climate variations of Alaska over the past century, and the models fail to capture this process. This should be noted so the reader can assess the validity of regional reconstructions and projections. Refs. below. [Richard Keen, USA]	Taken into account. Since the PDO is an internal mode of climate variability rather than a forced response to external conditions, the chronologies of the PDO in nature and in historical climate model simulations do not have to coincide.
9-1142	9	50	48	50	54	Hartmann, B., and Wendler, G. 2005. The significance of the 1976 Pacific climate shift on the climatology of Alaska. Journal of Climate 18: 4824–4839. http://climate.gi.alaska.edu/researchprojects/hartmann%20and%20wendler%202005.pdf Wendler, G., and Shulski, M. 2009. A Century of Climate Change for Fairbanks, Alaska. Arctic 62 (3): 295-300. http://climate.gi.alaska.edu/papers/Arctic62-3-295.pdf [Richard Keen, USA]	Taken into account. Reference cited.
9-1143	9	50	48	50	54	Also cite Lapp et al. 2011 GCM projections for the Pacific Decadal Oscillation under greenhouse forcing for the early 21st century. International Journal of Climatology, DOI: 10.1002/joc.2364. This is a more recent and thorough analysis than Stoner et al. 2009. Lapp et al. 2011 is cited elsewhere in AR5 where the climate model simulation of the PDO is assessed (chp. 14). [David Sauchyn, Canada]	Taken into account. The Lapp et al. (2011) reference is not referenced in this chapter which is focused on past behavior of the PDO.
9-1144	9	50	50	50	52	see also Wang, M., J.E. Overland and N.A. Bond. 2010. Climate projections for selected large marine ecosystems. Journal of Marine Systems 79: 258-266, and Furtado JC, E Di Lorenzo, N Schneider, and NA Bond, 2011. North Pacific Decadal Variability and Climate Change in the IPCC AR4 Models. J Climate 24: 3049-3067 [James Christian, Canada]	Taken into account. Reference cited.
9-1145	9	50	57	50	57	"...(e.g., Richter and Xie, 2008)." => "...(e.g., Richter and Xie, 2008). Another possibility is that some Atlantic Ninos are triggered by the southern subtropical ocean-atmosphere variations in observations (e.g., Huang and Shukla, 2005), which may not be well represented in current climate models." [Zeng-Zhen HU, USA]	We don't understand the relevance of this point to the cited text.
9-1146	9	50	57	50	57	Huang, B., and J. Shukla, 2005: The ocean-atmospheric interactions in the tropical and subtropical Atlantic Ocean. J. Climate, 18, 1652-1672. [Zeng-Zhen HU, USA]	ref to 9-1145
9-1147	9	51	1	51	14	do we expect improvements in CMIP5? [ANNALISA CHERCHI, Italy]	Taken into account. No studies yet on CMIP5
9-1148	9	51	2	51	2	Skip abbreviation SCT. [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-1149	9	51	11	51	11	fix brackets. [Robert Colman, Australia]	Editorial. Corrected
9-1150	9	51	11	51	11	Please replace "….may not be fully accounted for. (Solomon and Zhang, 2006) suggest that the CMIP3.." by " ...may not be fully accounted but Solomon and Zhang, 2006, have suggested that the CMIP3.." [HASIBUR	Editorial. Corrected

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						RAHAMAN, India]	
9-1151	9	51	16	51	16	Teleconnections are discussed and defined in Chapter 14, and patterns shown in Chapter 2. Cross reference and reconciliation needed. [George Kiladis, USA]	Taken into account. More coordination and cross-reference with Chaps 2, 10 and 14 and section re-written
9-1152	9	51	16			Section 9.5.3.5: The section 9.5.3.5 'Teleconnections' misses an important teleconnection: ENSO-NAO. There is a number of model studies showing importance of a good stratospheric representation for simulation of the ENSO-NAO teleconnection. Relevant publications: (1) Bell, C. J., L. J. Gray, A. J. Charlton-Perez, M. M. Joshi, A. A. Scaife, 2009: Stratospheric Communication of El Niño Teleconnections to European Winter. J. Climate, 22, 4083–4096. (2) Cagnazzo, C., and E. Manzini, 2009: Impact of the stratosphere on the winter tropospheric teleconnections between ENSO and the North Atlantic and European Region. J. Climate, 22, 1223–1238. (3) Ineson, S., and A. A. Scaife, 2009: The role of the stratosphere in the European climate response to El Nino. Nature Geosci., 2, 32–36, doi:10.1038/ngeo381. [Alexey Karpechko, Finland]	Taken into account. Section re-written and cited references taken into account
9-1153	9	51	16			Section 9.5.3.5: Is there any reason why the teleconnection between the South Asian summer monsoon and SSTs, particularly ENSO, is not included here? This is an important aspect of monsoon variability that is often not well represented by models. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Section re-written to add more relevant teleconnections, to the point they are assessed in the literature and reflect the CMIP models performance
9-1154	9	51	16			Section 9.5.3.5 Teleconnections. I suggest adding a subheading for ENSO teleconnections to equatorial South America and to the horseshoe region in the warm pool, either separately or as a subsection "Tropical ENSO teleconnections". I suggest a statement along the lines of: "These teleconnections are among the most important in terms of precipitation impacts for land regions.Tthese moist teleconnection pathways, in which a baroclinic tropospheric warming signal propagates within convection zones (Chiang and Sobel 2002), involve mechanisms related to those at play in the precipitation response to global warming (Neelin et al. 2003). They provide challenging test statistics for model precipitation response on climate timescales at which earlier generation climate models did poorly (Neelin 2007)." [Results from B. Langenbruner indicating that the present generation seems to do reasonably by some measure should be available prior to the July 2012 deadline].I would suggest also adding ENSO teleconnections to Atlantic and Indian SST in this section with references by NC Lau, A. Sobel, H. Su, Alexander and others because these are reproduced to some extent in current models and test the moist dynamics interacting with tropospheric temperature and ocean atmosphere transfer. If the review authors decide to include such a paragraph I would be happy to contribute sentences and references on request. [J. David Neelin, United States]	Taken into account. Section re-written to add more relevant teleconnections, to the point they are assessed in the literature and reflect the CMIP models performance
9-1155	9	51	19	51	19	Please replace "another." by "another," [HASIBUR RAHAMAN, India]	Editorial. Corrected
9-1156	9	51	23	51	23	Please replace " As examples " by " As an example" [HASIBUR RAHAMAN, India]	Editorial. Corrected
9-1157	9	51	31	51	37	this subsection should be focused on PNA [ANNALISA CHERCHI, Italy]	This subsection is focused on the PNA. However, to understand it's temporal behavior, we have to discuss ENSO and the PDO.
9-1158	9	51	43	51	43	Vera and Silvestri (2009) instead of "(Vera and Silvestri, 2009)" [Farahnaz Khosrawi, Sweden]	Editorial. Corrected
9-1159	9	51	43	51	43	(Vera and Silvestri, 2009) ==> Vera and Silvestri (2009) [Daehyun Kim, USA]	Editorial. Corrected
9-1160	9	51	43	51	43	Please replace "(Vera and Silvestri, 2009) showed that.." to " Vera and Silvestri, 2009 showed that.." [HASIBUR RAHAMAN, India]	Editorial. Corrected
9-1161	9	51	47	51	54	Sect.9.5.3.5.3: The WAM is affected just as strongly by other SST patterns besides ENSO and the Gulf of Guinea, as numerous papers testify. The Atlantic dipole (eg. Lamb 1978, Vizy and Cook 2001), Mediterranean (eg. Rowell 2003, Jung et al. 2006) and Indian Ocean (eg. Palmer 1986, Rowell 2001) are all also very influential. Citation details are in comment #7. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Noted. The cited references do not provide results about model evaluation of relevance to this section and AR5. This is an assessment, not a review.
9-1162	9	51	47	51	54	Citation details for comment #6: Jung, T; Ferranti L; Tompkins AM, 2006: Response to the summer of 2003	refs of 9-1161

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Mediterranean SST anomalies over Europe and Africa. J. Climate, 19, 5439-5454. Rowell, D.P., 2001: Teleconnections between the Tropical Pacific and the Sahel. Q. J. R. Meteorol. Soc., 127, 1683-1706. Rowell, D.P., 2003: The Impact of Mediterranean SSTs on the Sahelian Rainfall Season. J. Climate, 16, 849-862. Lamb, P. J., 1978: Large-scale tropical Atlantic surface circulation patterns associated with sub-Saharan weather anomalies. Tellus, 30, 240–251. Vizy, E. K., and K. H. Cook, 2001: Mechanisms by which Gulf of Guinea and eastern North Atlantic sea surface temperature anom- alies can influence African rainfall. J. Climate, 14, 795–821. Palmer, T. N., 1986: Influence of Atlantic, Pacific and Indian Oceans on Sahel rainfall. Nature, 322, 251–253. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	
9-1163	9	51	47	51	54	Concerning the analysis of CMIP3 simulation over the monsoon regions, specifically over West AFrica, I would suggest to add 3 papers concerning the evaluation of the CMIP3 GCMs in terms of the link between West African Monsoon and ENSO and in terms of intra-seasonal variability. Joly et al. (2007) and Joly and Voldoire (2008) analyzed the variability of the West African monsoon precipitation, focusing on its correlation with tropical SSTs from inter-annual to multi-decadal scales. While, Ruti and Dell'Aquila (2010) showed that despite the fact that several models correctly reproduce the ENSO phenomena, many models are still unable to capture the teleconnection mechanism, which bridges the Pacific SST forcing to the West African intra-seasonal variability. REFERENCES 1) Joly M, Voldoire A, Douville H, Terray P,Royer J, 2007: African monsoon teleconnections with tropical SSTs: validation and evolution in a set of IPCC4 simulations. Climate Dynamics 29, 1-20; 2) Joly M, Voldoire A, 2008: Influence of ENSO on the West African monsoon: temporal aspects and atmospheric processes. J of Climate, DOI: 10.1175/2008JCLI2450.1; 3) Ruti PM and Dell'Aquila A., 2010 The twentieth century African easterly waves in reanalysis systems and IPCC simulations, from intra-seasonal to inter-annual variability Clim. Dyn. doi:10.1007/s00382-010-0894-z [Paolo Michele Ruti, Italy]	Taken into account. We believe we already cited the key reference of Joly et al. 2007. Text will be updated to take into account recent studies on CMIP5, space permitting
9-1164	9	52	1	52	12	This section overlaps with 9.5.3.2. I don't think this is needed as a separate section and the two could be merged. [Nathan Gillett, Canada]	Taken into account. We agree with this suggestion and have added this to 9.5.3.2 we do though think that this small extra section is needed as NAM and jet position are not necessarily synonymous
9-1165	9	52	1	52	12	Cross reference back to Chapter 1 needed. [George Kiladis, USA]	Taken into account. X-ref made
9-1166	9	52	1	52	12	The sub-session about jet position is in a wrong place, as it is not about "teleconnections" of 9.5.3.5 [Hai Lin, Canada]	Taken into account. Text re-organised and moved to 9.5.3.2
9-1167	9	52	1			Section 9.5.3.5.4: I do not understand why the sub-section 9.5.3.5.4 ‘Mid-latitude jet position’ is placed to the section ‘Teleconnection’. The link is not so obvious. Moreover some parts of this sub-section only repeat the section 9.5.3.2 ‘North Atlantic Oscillation and Annular Modes’. I suggest moving sub-section 9.5.3.5.4 to section 9.5.3.2 (for example making it a sub-section there), which would save some space. [Alexey Karpechko, Finland]	Taken into account. Text re-organised
9-1168	9	52	3		6	Broadening of Hadley cells and poleward shifts of mid-latitude jets have been found in response to other forcings than GHGs, particulaly those that influence the stratosphere (e.g. Haigh, Blackburn and Day, JClim, 2005; Simpson, Blackburn and Haigh, JClim, 2009) [Joanna Haigh, UK]	Taken into account. We have added a reference to Simpson et al 2009
9-1169	9	52	9	52	9	This line indicates "higher vertical resolution" is key. Is this the case, or is it the higher model top? If it is the vertical resolution, then is it the tropospheric or stratospheric vertical resolution? [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This has been clarified as model domain and the associated physical processes needed to represent the middle atmosphere
9-1170	9	52	10	52	10	"…limitations in resolution" - Does this refer to horizontal or further vertical resolution? [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This has been clarified as model domain and the associated physical processes needed to represent the middle atmosphere
9-1171	9	52	11	52	12	end of last sentence, the reference is missing [ANNALISA CHERCHI, Italy]	Taken into account. We have added two references to Kang et al 2010 and Fu and Lin 2011
9-1172	9	52	11			Morgenstern et al. is refering to CCMs only. [Martin Dameris, Germany]	Rejected. The results are similar so we don't see a problem with this. Does the reviewer has any specific

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							suggestion?
9-1173	9	52	14			you could mention that volcanic eruptions give an opportunity to evaluate C-cycle response to an anomaly as a test for models. (e.g. Jones et al (2001, GBC) were the first to do this – I think the Nor-ESM model has done this since) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Accepted - text revised to note that responses to volcanic eruptions provide a useful test.
9-1174	9	52	17	52	17	Please replace " Randerson et al (2009)" to " Randerson et al. (2009)" [HASIBUR RAHAMAN, India]	Editorial
9-1175	9	52	27	52	30	This reads strangely. It says the simulation of the interannual variability is moderately skillful, and then says that some models exhibit no skill. I think this should name models, and then this would make more sense. [Nathan Gillett, Canada]	Accepted - text revised to be clearer.
9-1176	9	52	32			Section 9.5.3.6.2: As in 9.4.5.2 above, I think it needs to be explicit here that some of the experiments cited are ocean-only (reanalysis forcing) simulations. Also there is nothing here about CO2 fluxes per se. Perhaps these have not been systematically assessed for fidelity with observations, but there has been some analysis of coupled models' internal variability e.g. by Christian et al (2010) and by Doney et al 2006 (J Clim 19: 3033), as well as the analysis of Thomas et al (2008) previously discussed on 39/14. [James Christian, Canada]	Accepted - this text has now been removed as it does not relate specifically to the performance of ESMs.
9-1177	9	52	32			sec. 9.5.3.6.2: not sure oceanic primary productivity is the best measure of flux here – is it really the main driver of air-sea flux of CO2? I suspect ocean mixing/circulation/temperature are more important. There must be some literature on interannual variability of ocean air-sea flux and its causes and modelling (Doney et al?) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Noted. Text now removed. See response to 9-1176.
9-1178	9	52	34	54	34	Skip abbreviation PP, it is not really useful. [Farahnaz Khosrawi, Sweden]	Noted. Text now removed. See response to 9-1176.
9-1179	9	52	38	52	38	Units for the numbers? [Farahnaz Khosrawi, Sweden]	Noted. Text now removed. See response to 9-1176.
9-1180	9	52	40	52	40	Please replace " Schneider et al (2008)" to "Schneider et al. (2008)" [HASIBUR RAHAMAN, India]	Noted. Text now removed. See response to 9-1176.
9-1181	9	52	40	52	41	" In analysis of a limited sample of biogeochemical GCMs, Schneider et al (2008) identify only one model …": this sentence appears generic and do not specifies exactly the details (how many models? Which model?). [Claudio Cassardo, Italy]	Noted. Text now removed. See response to 9-1176.
9-1182	9	52	46	53	4	the section 9.5.3.7 would be much more effective summary if written as a table. [Juan A. Blanco, Canada]	Taken into account. Text modified for increased readability.
9-1183	9	52	46	53	4	Better to show this (hard to read) set of conclusions in tabular form. [Robert Colman, Australia]	Taken into account. Text modified for increased readability.
9-1184	9	52	46			Suddenly there is a summary section (also appearind on page 54, 9.5.4.4) not appearing earlier in the chapter (even though there has been summaries at the end of many subsections without the heading) A common line in the chapter is perferable. [Gunilla Svensson, Sweden]	Taken into account. Chapter revised for consistency
9-1185	9	52	48	52	50	I am confident that this varied picture is rather due to model deficiencies than to the short reliable data records. So I would rather write that the variation is caused by model deficiencies, but that the lack of reliable long observational data sets may also play a role. [Farahnaz Khosrawi, Sweden]	Taken into account. Text modified to clarify this
9-1186	9	52	48	52	57	a lot of use of the term "mixed skill". I found it difficult to decide what this meant without some explicit definition. [Bruce Wielicki, USA]	Taken into account. Text clarified
9-1187	9	52	48	53	4	These conclusions seem mixed as they refer partially to CMIP3 models and partially to CMIP5 models. As CMIP5 models are improved with respect to CMIP3 ones, I think this fact should be evidenced. [Claudio Cassardo, Italy]	Taken into account. Text updated with new CMIP5 assessment
9-1188	9	52	49			The instrumental record might be less reliable and/or shorter than what one would ideally prefer, but putting this first in a list of causes for the models presenting "a varied picture" is misleading. There are much more obvious reasons for model deficiensies than this. Our understanding and concern about climate change is rooted in nature as we observe it, and I am quite wary of the model community's tendency to suggest model credibility by referring to the inadequateness of the instrumental record. [Tor Eldevik, Norway]	Taken into account. Text clarified

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1189	9	52				Paragraph 9.5.3.7 - In case of insufficient evidence and insufficient agreement, again wouldn't it be more reasonable scientifically to stop the writing of the AR5 report there and avoid dubious projections on the basis of models that do not agree with recent observations ? [François GERVAIS, France]	Rejected. As shown by the vast body of literature, assessed in the several IPCC reports, there is sufficient scientific evidence that models are reliable enough to provide global projection of climate change. Projections are useful for making policy analysis.
9-1190	9	53	1	53	4	The last sentence of this paragraph needs rephrasing. It is also not consistent with 9.5.3.5.1 [Hai Lin, Canada]	Taken into account. Text modified
9-1191	9	53	5			at the end of this section, need a summary statement about carbon-cycle simulations in GCMs. This is a big new area of CMIP5 models, so needs a clear statement on their evaluation. e.g. can you make a statement about the overall quality of carbon cycle simulation relative to, say, simulation of T or P? Are the biogeochemical parts of models comparable to, or weaker than, the well established physical components? [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Noted -- but insufficient results to allow a strong assessment statement regarding carbon cycle variability
9-1192	9	53	6	53	6	Chapter 2 should define a common set of extreme indicators that other chapters to use. Ideally, the extremes assessed here should be those assessed in the observation chapter, and in the following detection/attribution, and projection chapters. [Xuebin Zhang, Canada]	Taken into account - indices are chosen based on cross-chapter coordination.
9-1193	9	53	22	53	58	I assume there will be more assessment in the SOD. Right now, there is no description in the text what Fig. 9.38 tell us. [Xuebin Zhang, Canada]	Noted - what Fig. 9.38 (new 9.37) means is now described in text based on a submitted paper by Sillmann et al.
9-1194	9	53	24	53	29	Sterl et al. (A. Sterl, C. Severijns, G.J. van Oldenborgh, H. Dijkstra, W. Hazeleger, M. van den Broeke, G. Burgers, B. van den Hurk, P.J. van Leeuwen, and P. van Velthoven (2008) When can we expect extremely high surface temperatures? Geophys. Res. Lett., 35, (2008), L14703 doi: 10.1029/2008GL034071) show that hot temperature extremes tend to be overestimated in the ECHAM5/MPI-OM climate model in already warm areas and underestimated in colder areas as compared to values derived from ERA-40. [Andreas Sterl, Netherlands]	Noted - note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-1195	9	53	40	53	43	description of figure 9.83 is insufficient, updated version of this figure contains ERA40 and NCEP in one plot, thus leaves room for more text [Jana Sillmann, Canada]	Noted - the figure and description are updated.
9-1196	9	53	40	53	43	figure 9.83 shows that the model performance in simulating temperature extremes is better than in simulating precipitation extremes on a global scale [Jana Sillmann, Canada]	Taken into acount - text revised.
9-1197	9	53	40	53	43	updated error estimates for the CMIP5 models displayed in figure 9.38 agree with the results of Kharin et al. 2007 as summarized in line 28-29. Indices based on the minimum temperature (tnmin, tnmax, tnfd and tn20) from model simulations deviate from the reanalysis to a larger extent than indices based on the maximum temperature (txmax, txmin, txid, txsu). It is also important to note that there is a large discrepancy between the two reanalysis data sets in the minimum temperature (as will be shown in detail in Sillmann et al. 2012) [Jana Sillmann, Canada]	Noted - a more general description has been provided due to space limitation.
9-1198	9	53	41	53	41	"Geckler et al. (2008) instead of "(Geckler et al, 2008)" [Farahnaz Khosrawi, Sweden]	Editorial
9-1199	9	53	43	53	43	"varies with models and indices": also with the dataset used (the results seem different for ERA40 and NCEP). [Claudio Cassardo, Italy]	Taken into acount - text revised.
9-1200	9	53	45			Why is ERA-40 used instead of ERA-Interim? [Robert Pincus, USA]	Noted. Because ERA-Interim starts only from 1979, while the base period of analysis starts from 1961.
9-1201	9	53	46	53	46	I would add an example for the interpretation of positive and negative values as is Fig 9.9 [Celeste Saulo, Argentina]	Noted - we think it is now clear as the figure caption is revised.
9-1202	9	53	46			Please explain the portrait diagram. What is the metric used? [Andreas Sterl, Netherlands]	Taken into acount - the figure caption is revised to fully explain what is shown.
9-1203	9	53	57	53	57	reference Haylock et al. 2008 is not necessarry in that context. Indice displayed in figure 9.38 are calculated based on fclimdex as described in Sillmann and Roeckner, Climatic Change, 2008 [Jana Sillmann, Canada]	Taken into account - the figure caption is revised.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1204	9	53				section 9.5.4: consider also an example of influence a large-scale dynamical phenomena (like SAM) on extremes and changes (Menendez and Carril, 2012,Potential changes in extremes and links with the Southern Annular Mode as simulated by a multi-model ensemble.Climatic Change Volume 98, Numbers 3-4, 359-377, DOI: 10.1007/s10584-009-9735-7) [ANNALISA CHERCHI, Italy]	Noted - however, the suggested paper is on future projection and not including model evaluation, hence not cited.
9-1205	9	54	3	54	8	This paragraph is rather strong in it conclusion that precipitation extremes are realistically simulated. For instance, Lenderink (2010) (Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. Climate Research, 44(2-3), pp.151-166.) show rather large biases (commonly of 50%) in most regional climate model simulations over Europe under perfect boundary conditions. Considering that we are interested in the changes in precipitation extremes of order 10 to 20 % these biases substantial. [geert lenderink, The Netherlands]	Noted - text revised. Note that relative performance is discussed here and it doesn't mean anything like some models are 'sufficiently' realistic for some purposes.
9-1206	9	54	3	54	8	Russo, S., and A. Sterl (2012) (Global changes in seasonal means and extremes of precipitation from daily climate model data, J. Geophys. Res., t117, D01108, doi: 10.1029/2011JD016260) show that the median of daily maximum precip is well represented in the in ECHAM5/MPI-OM. Absolute differences wrt obs (GPCP) are largest in the tropics and poleward of 30 deg, , while relative differences are largest in the dry regions between 10 and 30 deg. [Andreas Sterl, Netherlands]	Noted - note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-1207	9	54	3	54	8	Part of the problem is a spatial scale mismatch. There is really no observational data for the kind of extreme precipitation that is at the spatial scale comparable with model simulations. So if a GCM simulated extremes are comparable with station observations, we know the model simulated values too high. But there is very little we can say if model simulated extremes at model grid is lower than that observed at stations. [Xuebin Zhang, Canada]	Taken into account - discussion added.
9-1208	9	54	3	54	26	somewhere the paper by Maraun et al., Clim. Dynam, 2012 (reference above) should be mentioned. They use a statistical model to evaluate the representation of the relationship between large scale atmospheric circulation and extreme precipitation in the UK in 14 different RCMs in a perfect boundary conditions setting. They find that strong extremes are underestimated and that in mountaineous regions the dependence on flow direction is misrepresented. The paper also lists several other relevant papers on validation. Similarly, Schindler, Maraun and Luterbacher (to be submitted in March) find that many RCMs have difficulties in representing the seasonality in precipitation extremes across the UK. [Douglas Maraun, Germany]	Noted - the paper is cited in Section 9.6.2 Regional Climate Downscaling.
9-1209	9	54	6	54	8	Related to the above comment, model deficiencies in precip intensity are not simply a resolution issue so this is a littel too simp,istic a discussion - in fact my study quoted in this chapter (Stephens et al. 10) show the drizzle bias existing in global cloud resolving models (7km resolution) and further Suzuki et al., JAS, 2011) indicate the root cause of the drizzle bias may be a combination of the the way preciptation is parameterizaed and a more advanced microphysics scheme does better in principle and the way preciptation is evaporated below cloud. These isses are common across scales. [Graeme Stephens, USA]	Noted - but couldn't adopted due to space limitation ; the suggested point is discussed in 9.5.2.1 and recapped in 9.7.1.1.
9-1210	9	54	10	54	16	There is evidence that very high resolution RCMs are needed to capture hourly precipitation extremes. For example, Wakazuki et al (2008) show that a 5km RCM is able to much better reproduce hourly precipitation extremes over Japan compared to a 20km resolution AGCM. Kendon et al (2012) show that a 1.5km RCM gives a much more realistic representation of heavy rain over the UK at the hourly timescale compared to a 12km RCM. In particular in the coarser resolution regional climate model heavy rain is too persistent and widespread, with these errors significantly reduced at 1.5km resolution. [See comment 5 for references] [Elizabeth Kendon, United Kingdom of Great Britain & Northern Ireland]	Noted - this part is deleted here due to limited relevance to evaluation of CMIP simulations. Kendon et al. paper is cited in 9.6.2 Regional Climate Downscaling.
9-1211	9	54	10	54	16	I am not sure what I can learn from this passage. I don’t think the fact that a regional model was able to reproduce something observed at one station tells me anything about the performance of that regional model, nor climate model performance in general. [Xuebin Zhang, Canada]	Noted - this part is deleted here due to limited relevance to evaluation of CMIP simulations.
9-1212	9	54	15	54	16	Further analysis for a number of stations in western Europe has shown that for higher temperatures models have difficulties to reproduce the observed intensificiation of hourly extreme precipitation with temperature due to a much too strong sensitivity of the modelled precipitation rates to decreases in relative humidity (Lenderink, G. & van Meijgaard, E., 2010. Linking increases in hourly precipitation extremes to atmospheric temperature and moisture changes. Environmental Research Letters, 5(2), p.025208). [geert lenderink, The Netherlands]	Noted - this part is deleted here due to limited relevance to evaluation of CMIP simulations. The paper is cited in 9.6.2 Regional Climate Downscaling.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1213	9	54	18	54	21	As per comment 1 - Min et al is dubious in that sucessive decadal increases in precip intensity cannot be related to increases in NH temperature as the first 25 years of this period did not exhibit increases in temperature. Either the convoluted data processing/interpolation is in error or else the 'observed' increases in precip intensity must be unrelated to temperature, and therefore not attributable to anthropogenic factors [Stewart Franks, Australia]	Noted - Although we do not think that Min et al is dubious because extreme precipitation doesn't depend only on temperature and thus their relationship doesn't have to hold decade by decade, text is revised to more exactly represent what Min et al have found.
9-1214	9	54	30	54	39	TC frequency is likely predictable, and improvements in its predictability occur with the inclusion of the SST assimilation in the coupled model improving the daily temperature, CAPE and vertical wind gradient (Alessandri et al, 2011). The reference is: "Alessandri, Andrea, Andrea Borrelli, Silvio Gualdi, Enrico Scoccimarro, Simona Masina, 2011: Tropical Cyclone Count Forecasting Using a Dynamical Seasonal Prediction System: Sensitivity to Improved Ocean Initialization. J. Climate, 24, 2963–2982." [ANNALISA CHERCHI, Italy]	Noted - the suggested paper is relevant, but the SST assimilation is not the central issue in this context where AGCMs with prescribed SST are evaluated. Also, note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-1215	9	54	41	54	44	A possibly relevant reference, which discusses droughts in several CMIP3 models, is: McCrary, R., and D. A. Randall, 2010: Great Plains Drought in Simulations of the Twentieth Century. J. Climate, 23, 2178-2196. [David Randall, USA]	Taken into account - citation added.
9-1216	9	54	48	54	48	I am not sure how to intepret "limited evidence" and "high agreement". It would be better to be more specific on this. For example, it might be useful to point out which evidence is here, and what do you mean high agreement. [Xuebin Zhang, Canada]	Noted - whole assessment is now revised and there is no 'limited evidence' & 'high agreemnt'.
9-1217	9	54	48	54	51	Same as three previous remarks. [François GERVAIS, France]	Rejected - here the agreement is high. Even if agreement is low, it is meaningful to assess it and recognize it.
9-1218	9	54	48	54	53	Assessment on extreme events is too thin, is this all you can do? There needs to be better fundation on this to build confidence for the relevant parts of the detection/attribution and projection chapters. [Xuebin Zhang, Canada]	Noted - more assessment based on new studies with CMIP5 analyses is added.
9-1219	9	54	50	54	51	as per comments 1 and 6 above. The text states the the models underestimate the observed increases in precip. The precip is not "observed" but derived through a convoluted process; the sucessive decadal increases in precip are not possible to link to NH temperature and this did not rise in the first 2 and a half decades. Either the interpolation/data filtering has produced a spurious trend, or if the trend is real it cannot be directly related to anthropogenic influence [Stewart Franks, Australia]	Noted - Although we do not think that Min et al is dubious because extreme precipitation doesn't depend only on temperature and thus their relationship doesn't have to hold decade by decade, text is revised to more exactly represent what Min et al have found.
9-1220	9	55	9	55	9	a.k.a? What is this abbreviation standing for? [Farahnaz Khosrawi, Sweden]	editorial. "a.k.a." stands for "also known as", shorthand in English.
9-1221	9	55	14	59	13	Regional simulation is so difficult that you tend to conceal, distort or even fabricate the actual observations. [VINCENT GRAY, NEW ZEALAND]	Rejected. There is a large body of peer-reviewed scientific literature on the subject of regional climate analysis and modelling.
9-1222	9	55	20	55	21	already specified in previous sections [ANNALISA CHERCHI, Italy]	Taken into account. Addressed duplications in the revision.
9-1223	9	55	23	55	35	This paragraph seems unnecessary, and could be removed without loss. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Noted. Evaluation of Antarctic P-E is, however, important background for the assessment in Chapter 13, and thus provides important cross-chapter information.
9-1224	9	55	24	55	24	first 'is' is redundant. [Zhaomin Wang, UK]	Editorial.
9-1225	9	55	38	55	38	Please replace " …Räisänen and Ylhäisi, 2011) " with "Räisänen and Ylhäisi, 2011)." [HASIBUR RAHAMAN, India]	editorial. Suggestion unclear.
9-1226	9	55	42	55	42	acronym for VOC [ANNALISA CHERCHI, Italy]	Noted. The acronym is not present in SOD.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1227	9	55	46	55	46	Figures 9.2 and 9.4? I think 9.3 and 9.4 are meant here. [Farahnaz Khosrawi, Sweden]	Taken into account. The correct figure references were 9.3 and 9.5.
9-1228	9	55	46	55	46	"...and precipitation is illustrated in Figures 9.2 and 9.4." I think its wronly quoted Figure 9.2 and 9.4 does not show the regional temperature and precipitation [HASIBUR RAHAMAN, India]	Taken into account. The correct figure references were 9.3 and 9.5.
9-1229	9	55	46	55	46	should refer to Figures 9.3 and 9.5 [Celeste Saulo, Argentina]	Taken into account. The correct figure references were 9.3 and 9.5.
9-1230	9	55	46			typo: Fig. 9.3 and Fig. 9.5 [Ramon de Elia, Canada]	Taken into account. The correct figure references were 9.3 and 9.5.
9-1231	9	55	46			Reference to Fig 9.2 should be to 9.3 [Brad Murphy, Australia]	Taken into account. The correct figure references were 9.3 and 9.5.
9-1232	9	55	54	55	54	What is the abbreviation CRU standing for? [Farahnaz Khosrawi, Sweden]	Noted. "CRU" stands for "Climate Research Unit", a commonly used and recognised acronym.
9-1233	9	55	54	55	54	Same comment as #46: CRU data: (is it HadCRUTT3v ? Is it HadCRUT4 as in chapter 2?). I suggest using the same reference data set for model verification [Celeste Saulo, Argentina]	Noted. We have clarified the use of the observational data.
9-1234	9	55	55			The term „metric“ should not be used. A metric is a distance measure with clearly defined properties, e.g., it is always equal to or larger than zero. This is definitely not the case for, e.g., daily minimum temperatures. Please use the term indices instead. [Douglas Maraun, Germany]	Noted. We aligned the assessment to the language used in the literature. "Metric" has a connotation in this contect that differs from "index".
9-1235	9	56	1			Bias correction methods have not been mentioned at all in this section (9.6.2). The only reference is to Christensen et al., 2008. As bias correction is routinely used, it should be discussed somewhere, either in 9.6.2.1 as a statistical downscaling method, in 9.6.2.2 as a postprocessing of RCMs, or in 9.6.3. Also limitations and potential problems should be discussed. E.g., Maraun (2012, submitted) find nonstationarities in temperature biases, where surface albedo changes in Winter and cloud cover and soil moisture changes in summer are important, as well as apparent nonstationarities in precipitation biases due to internal climate variability. [Douglas Maraun, Germany]	Accepted. Bias correction methods are now elaborated in the context of model skill. Application to address model errors, by means postprocessing of model outputs, for example in impact analysis is at the remit of WGII.
9-1236	9	56	1			Somewhere in the approach by Eden et al (Eden, Widmann, Grawe, Rast, 2012, J Climate, doi: 10.1175/JCLI-D-11-00254.1) should be mentioned. They nudge the ECHAM5 AGCM to reanalysis data and observed SST to calibrate the predicted precipitation to observed precipitation. Thus this approach is capable of directly downscaling from GCM grid box precipitation to observed precipitation. [Douglas Maraun, Germany]	Noted. The rereference was considered from the assessment point of view.
9-1237	9	56	3			This section (9.6.2.1) needs some work (specific comments will follow below). I am missing a classification of different statistical approaches (e.g., into Perfect Prog, Bias Correction Methods/Model Output Statistics and Weather Generators, as suggested by Maraun et al., 2010b). Also the references are a bit muddled. E.g., Vrac and Naveau are a very good example of an approach that predicts a whole distribution, it could therefore very well be listed in line 15. Please restructure this paragraph! [Douglas Maraun, Germany]	Taken into account. We considered the references from assessment point of view. Please note that the scope is one of assessment, not a review. Also, we note that there are different ways of classifying methods, and the discussion of these is not focal for the assessment.
9-1238	9	56	3			Sect.9.6.2.1: Need to mention that the relationships between large-scale and small-scale variables are often different after large anthropogenic forcing has been applied, so statistical downscaling can problematic (needs to be used with great caution) under such circumstances. (I suggest this inaddition to, or instead of, this caveat being noted at line 48.) [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Noted. Accepted. We have noted more explicitly on the stationarity hypothesis, when discussing "Relating Model Performance to Credibility of Model Applications".
9-1239	9	56	5	56	5	Is the abbreviation SD really useful or necessary? [Farahnaz Khosrawi, Sweden]	Noted. Its was used in Section 9.6 after defining it to save space. "SD" is commonly used abbreviation for "Statistical Downscaling".
9-1240	9	56	9			can two cites really cover the hole development for statistical method’s? [Frank Kreienkamp, Germany]	Noted. Other references were also cited. It is also important to take into account that the scope is an assessment and not a review, and the aim is to report on the main developments.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1241	9	56	12	56	12	first sentence not clear9 [ANNALISA CHERCHI, Italy]	Noted. Revised.
9-1242	9	56	12		24	there are methods covering not only one point and one element. Impact modelers need more: have a look at Kreienkamp et al 2011: Climate Signals on the Regional Scale Derived with a Statistical Method: Relevance of the Driving Model's Resolution. Atmosphere, 2, 129-145; doi:10.3390/atmos2020129 which also show a comparison between ESD and RCM results. [Frank Kreienkamp, Germany]	Noted. The reference provides an example of the importance of using multiple downscaling approaches whether statistical or dynamical and the possibility of using both approaches together. It is important to consider that this chapter addresses climate model evaluation. It includes also downscaling (in general) but the assessment of the all methods used in impact studies is out of scope here.
9-1243	9	56	12			„approaches now combine different approaches" This sentences is too unspecific! Please clarify what different approaches are meant. Also the references seem to be somewhat arbitrary. [Douglas Maraun, Germany]	Noted. The comment about the references is unclear.
9-1244	9	56	14	56	15	Most SDS do not predict raw values (only maybe the analogue method). Linear regression always predicts a mean value, which is a moment of the distribution, i.e., a statistical attribute. What several state-of-the-art methods actually do is predicting other statistics than just the mean, e.g., quantiles or a full parameterised distribution (e.g., Vrac and Naveau, Wat Resour. Res., 43:W07402, 2007 and Vrac and Naveau 2008, cited). [Douglas Maraun, Germany]	Accepted. The sentence was about the use for the projections in the future, and is removed.
9-1245	9	56	20			Fowler et al., 2007 and Maraun et al., 2010 are both review papers. Should they be cited here as actual studies? [Douglas Maraun, Germany]	Accepted. The two references were not cited among individual studies.
9-1246	9	56	26	56	39	Concerning recent developments of dynamical methods, recent EU and international projects (CIRCE-EU, CORDEX) have progressed in developing regional coupled models in order to better simulate climate variability and scenarios. In Artale et al. (2009) an atmosphere–ocean regional climate model for the Mediterranean basin, called the PROTHEUS system, has been assessed by using available observational datasets. Despite a persistent bias, the PROTHEUS system is able to capture the inter-annual variability of seasonal sea surface temperature (SST) and also the fine scale spatio-temporal evolution of observed SST anomalies, with spatial correlation as high as 0.7 during summer.While concerning scenarios, in Dell'Aquila et al. (2011) a more accurate description of orography produces in the regional coupled model a narrower identification of the effects of a warmer climate on intense precipitation events and on other key environmental indicators, such as the snow cover extension and the aridity index. An example of the impact of climate variability on river discharge is also evident for a medium/small-size catchment basin in Northern Italy, the Po river. REFERENCES 1) Vincenzo Artale, Sandro Calmanti, Adriana Carillo, Alessandro Dell'Aquila, Marine Herrmann, Giovanna Pisacane, Paolo M. Ruti, Gianmaria Sannino, Maria Vittoria Struglia, Filippo Giorgi, Xunqiang Bi, Jeremy S. Pal, Sara Rauscher, The PROTHEUS Group, 2009 An atmosphere-ocean regional climate model for the Mediterranean area: assessment of a present climate simulation Clim. Dyn. doi:10.1007/s00382-009-0691-8; 2) Dell'Aquila A., Calmanti S., Ruti P. M., Struglia M. V., Pisacane G., Carillo A., Sannino G., 2011 Impacts of seasonal cycle fluctuations in an A1B scenario over the Euro-Mediterranean. Climate Research Clim. Res. doi:10.3354/cr01037 [Paolo Michele Ruti, Italy]	Taken into account - the references are considered from the assessment point of view.
9-1247	9	56	28	56	39	Suggest adding at end of paragraph "Recent research is also seeing the first regional climate model simulations at convective permitting scales, which give a much better representation of the diurnal cycle of convection and hourly precipitation extremes (e.g. Hohenegger et al., 2008; Wakazuki et al., 2008; Kendon et al., 2012)" [See comment 5 for references] [Elizabeth Kendon, United Kingdom of Great Britain & Northern Ireland]	Accepted. Text and references complemented.
9-1248	9	56	38	56	39	The study by Somot et al. 2008 is irrelevant in the present context ("clilmate models with interactive ocean and sea ice"), and should be removed (or placed elsewhere). Furthermore, Smith et al 2011a is quoted twice. [Arne Melsom, Norway]	Editorial (Smith et al.). Taken into account (Somot et al.) that the model in question is a coupled atmosphere-ocean model, but sea ice is not relevant in the study. The text was revised.
9-1249	9	56	39			Important to the global energy and water balance, as well as society, is snow pack which often serves as a key reservoir of water. Waliser et al. showed that key uncertainties exist in the projection of snow pack (in the sierras) due to many remaining uncertainties in models: 1) when topography is not well resolved by coarse models, there snow pack is likely underestimated since tall, cold peaks do not exist in coarse models, 2)	Noted. Chapter 9 is about climate model evaluation. Projections are at the remit of Chapter 14 and discussion on how regional-scale climate information is used in impact assessment of Working Group II.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						single-layer snow models can overesimate the time it takes snow pack to melt under climate change conditions, and 3) the aerosol content of snow which is highly uncertain has a strong influence on seasonal and long-term melting. -------- Waliser, D., J. Kim, Y. Xue, Y. Chao, A. Eldering, R. Fovell, A. Hall, Q. Li, K. Liou, J. McWilliams, S. Kapnick, R. Vasic, F. D. Sale, and Y. Yu (2011), Simulating the Sierra Nevada snowpack: The impact of snow albedo and multi-layer snow physics, Climatic Change, 109 (Suppl 1):S95–S117, DOI 10.1007/s10584-011-0312-5. [Duane Waliser, USA]	
9-1250	9	56	41			Here the term "added value" is solely attributed to regional downscaling methodologies while it might be expected by some readers to come from high resolution AOGCM simulations, as well. The added value results from minimization of systematic errors in simulations of regional climate and its variability in response to model resolution increase (or empirically based adjustment of coarse resolution climate simulations). [Igor Shkolnik, Russian Federation]	Accepted. We have elaborated on the discussion of provision of regional-scale climate information, including AGCMs and variable-resolution GCMs in Section 9.6. High-resolution climate-length AOGCM runs for climate projections are still very little explored.
9-1251	9	56	43	59	13	the discussion only focuses on the view of climate modelers .. do not forget the needs of impact modelers .. is there much use of of plain results from rcm? Is bias correction an improvement or just a mathematical trick? [Frank Kreienkamp, Germany]	Noted. Chapter 9 is about climate model evaluation. Discussion on how regional-scale climate information is used in impact assessment is the remit of Working Group II.
9-1252	9	56	48	56	48	relationships [Graham Feingold, United States of America]	Editorial.
9-1253	9	56	52			it is correct that statistical connections may may not be stabile over time, but i have not seen the same argument in the discussion of parametrisations of AOGCM, ESM and RCM, please do not forget to add, please add also a extended discussion about the results from UKCIP 2009 experiment with a perturbet parametrisation experiment [Frank Kreienkamp, Germany]	Taken into account. Parameterisations in RCMs are much as in AOGCMs and based on physical principles, but do include tuning. We have elaborated on the discussion of model biases and on "Relating Model Performance to Credibility of Model Applications". The UKCP09 is focused on projections, which are not the focus of Chapter 9.
9-1254	9	56	54	56	54	remove "of" [ANNALISA CHERCHI, Italy]	Editorial
9-1255	9	56	54	56	55	delete"of" or "for" after evaluation method [Celeste Saulo, Argentina]	Editorial
9-1256	9	56	54	57	12	The discussion of process oriented validation is a bit muddled. First, papers by Lenderink and van Meijgaard and Maraun et al. are mentioned as papers that „evaluate physical processes“, only some lines later examples for „process oriented validation“ are cited (e.g., Sasaki and Kurihara). This paragraph needs restructuring. [Douglas Maraun, Germany]	Taken into account. Restructured.
9-1257	9	56	54	57	12	Somewhere the study by Schindler, Maraun and Luterbacher (to be submitted in March) may be mentioned. They find that many RCMs have difficulties in representing the seasonality in precipitation extremes across the UK. [Douglas Maraun, Germany]	Taken into account. We considered the paper, but opted for others as examples.
9-1258	9	56	54			typo: "method FOR" [Ramon de Elia, Canada]	Editorial.
9-1259	9	56	56	56	56	Please add reference to Ning et al 2012 [Ning, L., Mann, M.E., Crane, R., Wagener, T., Najjar, R.G. Jr., Probabilistic Projections of Anthropogenic Climate Change Impacts on Precipitation for the Mid-Atlantic Region of the United States, J. Climate (in revision)] which describes related application of the statistical downscaling procedure to the CMIP3 (A2) projections. The abstract is as follows: We use an empirical downscaling method based on Self-Organizing Maps (SOMs) to produce high-resolution, downscaled precipitation projections over the state of Pennsylvania in the Mid-Atlantic region of the U.S. for the future period 2046-2065. In order to examine the sensitivity of precipitation change to the water vapor increase brought by global warming, we test two approaches to downscaling: one uses the specific humidity in the downscaling algorithm and the other does not. Application of the downscaling procedure to the model projections shows changes in the relative occupancy, but not the fundamental nature, of the simulated synoptic circulation states. Both downscaling approaches predict increases in annual and winter precipitation, consistent in sign with the raw output from General Circulation Models (GCMs) but considerably smaller in magnitude. For summer precipitation, larger discrepancies are seen between the raw and downscaled GCM	Accepted. Cited as an example of methods' development, as Chapter 9 is limited to climate model and regional downscaling evaluation. Results of regional downscaling approaches applied to future projections may be found in Chapter 14 or in the Working Group II assessment report.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						projections, with a substantial dependence on which of the two downscaling approaches is used (downscaled precipitation changes employing specific humidity are smaller than those without it). Application of downscaling reduces the inter-GCM variation, suggesting that some of spread among models in the raw projected precipitation may result from differences in precipitation parameterization schemes rather than fundamentally different climate responses. Projected changes in the North Atlantic Oscillation (NAO) are found to be significantly related to changes in winter precipitation in the downscaled results but not for the raw GCM results, suggesting that the downscaling more effectively captures the influence of climate dynamics on projected changes in winter precipitation. [Michael Mann, USA]	
9-1260	9	56				further explanations dynamical downscaling methods is requried and regional analysis particularly for south Asia is missing. [Muhammad Amjad, Pakistan]	Noted. The first suggestion is unclear. The second was covered in light of assessed literature. A detailed region-to-region assessment is provided by WGII.
9-1261	9	57	1			Maraun et al., 2010b, is a review. Please cite the relvant references from the paper. E.g., Kendon, Rowell and Jones (2009), Clim. Dynam. 35:489-509; or Maraun, Osborn and Rust, Clim. Dynam. 2012, DOI 10.1007/s00382-011-1176-0 [Douglas Maraun, Germany]	Taken into account. The references in the section were considered from the assessment point of view and overall complemented.
9-1262	9	57	21			The recent review paper Feser, Rockel, von Storch, Winterfeldt and Zahn, B.A.M.S., 2011, on added value should be cited. [Douglas Maraun, Germany]	Accepted. Cited.
9-1263	9	57	31	57	31	What is the abbreviation MSLP meaning? [Farahnaz Khosrawi, Sweden]	Noted. "MSLP" stands for "mean sea level pressure"; normal nomenclature.
9-1264	9	57	32	57	32	insert "Indian" (right?) between "Whereas" and "monsoon" [ANNALISA CHERCHI, Italy]	Editorial.
9-1265	9	57	36	57	39	Figure 9,.40 is unclear wrt panel a). I could not tell the difference between the larger figure and the smaller one (the one to the bottom right) [Celeste Saulo, Argentina]	Taken into account. Figure caption revised.
9-1266	9	57	41			A study partially addressing this issue was performed in de Elia and Cote (2010) Climate and climate change sensitivity to model configuration in the Canadian RCM over North America. Meteorologische Zeitschrift 19:325-339. [Ramon de Elia, Canada]	Noted. The reference was considered from the assessment point of view.
9-1267	9	57	49	57	50	The discussion here should be extended with two studies. In Køltzow et al (2008) an extension of the Big Brother experimental concept was used to address the question of size of the integration domains for dynamical downscaling (or alternatively the use of spectral nudging, but that was not addressed explicitely).It was found that a properly applied RCM reproduces the climate features in the large-scale data forced through the boundaries provided these data are of high quality. When forced with data of lower quality (coarse resolution), the RCM significantly improved the climate statisitics over regions dominated by strong local surface forcing provided a sufficiently large integration domain for the RCM. Køltzow, M.A.Ø., T. Iversen. J. E. Haugen. (2008) Extended Big-Brother experiments: the role of lateral boundary data quality and size of integration domain in regional climate modelling. Tellus 60A,398-410. DOI: 10.1111/j.1600-0870.2008.00309.x [Trond Iversen, United Kingdom of Great Britain & Northern Ireland]	Accepted. The text was complemented.
9-1268	9	57	51	57	52	The aim of the second study (Køltzow et al, 2011) was to explore the relative importance of the quality of the oceanic surface forcing (SST and sea-ice) and the lateral boundary condition data as a function of the size of the integration domain. This was done by switching ocean-surface data and large-scale atmospheric data from different models when provided to the RCM. With Norway as the scope for the results, changes in the lateral boundary data and the size of the integration domain turned out to be more important than the quality of the ocean surface data. Køltzow, M.A.Ø., T. Iversen, J. E. Haugen (2011) The Importance of Lateral Boundaries, Surface Forcing and Choice of Domain Size for Dynamical Downscaling of Global Climate Simulations. Atmosphere 2011, 2, 67-95; doi:10.3390/atmos2020067. [Trond Iversen, United Kingdom of Great Britain & Northern Ireland]	Noted.
9-1269	9	57	51	58	12	The item is put together not too well, contains a number of heterostructural aspects and abstract technical details. A general impression is that the dynamical downscaling entails mostly side effects that are detrimental or useless every here and there. Desired is a clear message on the positive role of internal variability that is	Taken into account. The text may have sounded negative because the intention was to discuss the uncertainty of the RCM setup. The text was revised.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						generated due to nonlinearities in RCM physics and dynamics. Apparently, the internal variability in an RCM involves cumulative response of climate variations to mesoscale forcings. A bit more encouraging message would help maintaining logical flow and provide bridging to the next section. [Igor Shkolnik, Russian Federation]	
9-1270	9	57	51			Two distinct but related ideas are expressed here: 1- RCMs allow for internal variability and the smaller the domain the smaller the internal variability (all other things being equal) (Alexandru et al 2007 and many others), and 2- small domains do not allow for the full development of small scales, --in a way reducing the raison d'etre of RCMs (Leduc and laprise 2009). Both these issues have competing positive and negative effects and a compromise is usually sought. [Ramon de Elia, Canada]	Taken into account. Text revised.
9-1271	9	57	55	57	55	It may be worthwhile to explain the Big Brother experimental setup [Peter Guttorp, USA]	Rejected. The scope is assessment and excludes full elaboration of background information. The experimental set-up is well cited in the referenced literature.
9-1272	9	58	2			Internal variability has to do more with multiple solutions --as in global models-- than with deviations from the driving data. Sometimes the boundary driving data is not a solution in the center of the RCM domain. In other words, internal variability is not the cause of this difference, simply the RCM may not include the evolution in the low-resolution dataset as a solution. [Ramon de Elia, Canada]	Taken into account. The text was developed further.
9-1273	9	58	4			Another important effect of internal variability is to make tuning and perturbed-parameters experiments more costly, since longer runs are need to reach statistical significance. Separovic L, R de Elia, R Laprise, 2011: Impact of spectral nudging and domain size in studies of RCM response to parameter modification, Climate Dynamics, DOI 10.1007/s00382-011-1072-7. [Ramon de Elia, Canada]	Taken into account. Internal Variability discussion, although kept brief, was revised including citing the paper.
9-1274	9	58	9	58	9	What is the evidence for RCMs having more detailed representation of processes? Whilst this is the case when running at around 1km and convection becomes explicitly resolved, I'm unclear which processes have a more detailed representation at typical RCM resolutions (~25km) relative to the GCM (~100km) [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account. The text referred to the scale of the phenomena and the skillful scale issue as mentioned at the beginning of Section 9.6. More detailed is not necessarily explicitly resolved or more detailed parameterisations. However, in very-high resolution models, there now are few examples of models that run without convection parameterisation (9.6.2.2 of the SOD). Another example is of RCMs with lake modelling.
9-1275	9	58	17			Although I agree with the statement (e.g. similar conclusion in de Elia and Cote 2010), I have a difficult to see the origin of the qualification "medium agreement, limited evidence", even after reading Mastrandea et al (2010). My impression is that more information will be needed to make that statement "medium agreement, limited evidence" defensible. I would suggest to eliminate it even if it is part of the recommended language by IPCC. [Ramon de Elia, Canada]	Noted. We have developed the discussion of relevant issues and elaborated the background for the statement. The IPCC uncertainty language is an important feature of the Assessment, to provide transparent assessment results.
9-1276	9	58	21	59	13	Most of the examples of value added focus on a better simulation of the climatology. But to add value to projections the RCMs would have to give improved skill in the changes in climate. Moreover for this added value to be useful, high resolution projections would have to be robust to model uncertainty at the driving grid scale. And to be worth running an RCM, it would have to outperform statistical methods. I would advocate some more skepticism in this section. [Nathan Gillett, Canada]	Noted. Research on "added value" has to a large extent been pursued with evaluation of downscaling results against observations, as is reflected in the litterature. We have elaborated on the discussion of added-value and skill, while still not focusing on climate projection results, as they are outside the writ of Chapter 9-
9-1277	9	58	21			Section 9.6.3.2 : Research on the kind of information that RCMs are good --or unique-- at producing has been a constant in the RCM community. This created a movement that is usually associated to the "added value" studies (e.g. note the directness of this title: "Very high resolution regional climate model simulations over Greenland - identifying added value" Lucas-Picher et al. In press, J. Geophys, Res.). But the idea of added value should apply to any attempt to improve climate models in general; although this is done in part, difficulties to appreciate progress in GCMs are never considered evidence against the need of, for example,	Accepted. We have complemented Section 9.6 with more assessment of other means of providing regional-scale climate information, e.g. AGCMs.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						going to higher resolution. See section 9.7.3.3 where "added" and "lost" value is also found but the requirement for added value is not discussed. It seems to me that presenting a section on added value solely for RCMs is somehow misleading. [Ramon de Elia, Canada]	
9-1278	9	58	21			The section describes the added value that has been revealed mostly on spatial pattern basis. Another manifestation of the value added can be an increased accuracy in climate variability simulations by RCM. Since low resolution AOGCMs may underestimate the level of regional climate variability (see Sec. 9.5), RCMs serve to make up that deficiency by generating additional variability at higher resolution grids. Consequently, RCM application allows to overdetermine the range of regional uncertainties due to variability in AOGCM ensemble projection. This issue particularly implies that added value can exhibit potential not only in physical but also in probabilistic space (downscaling signal-to-noise estimates). It has been shown in (Shkolnik et al., 2012) that RCM at 25 km resolution reproduced the level of interannual variability of annual extreme temperature and precipitation indices closer to ground-based observations than driving GCM. Reference: Shkolnik I.M., Meleshko V.P., Efimov S.V., Stafeeva E.N., 2012: Changes in climate extremes over Siberia by the mid 21st century: ensemble projection using MGO RCM. Russ. Meteorol. Hydrol., 2 (in press). [Igor Shkolnik, Russian Federation]	Taken into account. We have cited the result as one of the examples of added-value located in simulated variability characteristics.
9-1279	9	58	23	58	23	It is not fully clear what exactly "useful additional information" means. A rigorous definition of added value would be appreciated. The study (di Luca et al., 2011) deals with the subject. Reference: di Luca, R. de Elia, R.Laprise, 2011: Potential for added value in precipitation simulated by high resolution nested Regional Climate Models and observations, Clim Dyn, doi:10.1007/s00382-011-1068-3. [Igor Shkolnik, Russian Federation]	Accepted. We have considered the reference in the revision of the section. We have also elaborated on the discussion of "added-value" and "skill", respectively.
9-1280	9	58	28	58	28	Suggested addition: "… and monthly mean precipitation variability (Shkolnik et al., 2007)". Reference: Shkolnik I.M., V.M.Kattsov, V.P.Meleshko, 2007: The MGO climate model for Siberia. Russian Meteorology and Hydrology, Volume 32, Number 6, 351-359, doi:10.3103/S1068373907060015 [Igor Shkolnik, Russian Federation]	Taken into account. We mentioned the MGO RCM studies alongside other new developments (SOD section 9.6.2.2).
9-1281	9	58	39			simulation *of* Atlantic… [Philip Mote, USA]	Editorial.
9-1282	9	58	58			A publication to add to this list: Di Luca A, R de Elia, R Laprise, 2011: Potential for added value in precipitation simulated by high-resolution nested Regional Climate Models and observations, Climate Dynamics, 10.1007/s00382-011-1068-3. And a submitted mansucript to Climate Dynamics: Di Luca, A. R de Elia, R Laprise, 2011: Potential for added value in RCM-simulated surface temperature. [Ramon de Elia, Canada]	Taken into account. We thank for the provision of this reference and have considered this reference.
9-1283	9	59	10			A point that can be misleading is whether RCMs add value to the simulation of climate or to the simulation of the climate change signal. I know of only one study about the latter (Di Luca A, R de Elia, R Laprise, 2011: Potential added value of RCM's downscaled climate change signal, submitted to Climate Dynamics). [Ramon de Elia, Canada]	Taken into account. We thank for the provision of this reference and have considered located available litterature.
9-1284	9	59	12			As in my comments of page 58 line 17, I fear that the qualification "high agreement, medium evidence" gives a sense of rigour that may be misleading. I suggest to remove it and accept the fact that we are not ready to formulate things this way in this aspect of climate modelling. [Ramon de Elia, Canada]	Noted. We have elaborated the background for the statement. The IPCC uncertainty language is an important feature of the Assessment, to provide transparent assessment results. Also, we deem that the today long legacy of downscaling research can support an IPCC uncertainty-language -framed statement.
9-1285	9	59	15	59	15	The source of all the model errors is the set of absurd assumptions upon which they are founded [VINCENT GRAY, NEW ZEALAND]	Rejected - no scientific basis for this comment. Models ARE to a very large extent based on physical principles.
9-1286	9	59	17	59	17	Omit "Approach to" in the section title, so that it readsLinking Process Understanding and Model Performance" [Farahnaz Khosrawi, Sweden]	Rejected - This subsection is a roadmap for the following subsections. But note that we are revising the structure of 9.7.1 - 9.7.3 anyway.
9-1287	9	59	26	59	33	There are several "Good Modelling Practices Guidelines" that are triying to uniformize the evaluation process	Rejected - We do not ourselves write a good-practice

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						in all type of models. They should be mentioned here as well as recommended for future developers of climate users: [1] Van Waveren, R. H., Groot, S., Scholten, H., Van Geer, F. C., Wösten, J. H. M., Koeze, R. D. and Noort J. J. (1999) Good Modelling Practice Handbook (STOWA Report 99-05), Utrecht, The Netherlands. [2] Council for Regulatory Environmental Modeling (CREM) (2009) Guidance on the Development, Evaluation, and Application of Environmental Models, US Environmental Protection Agency, Washington, DC. [3] Kimmins J.P., Blanco J.A., Seely B., Welham C., Scoullar K. 2010. Forecasting Forest Futures: A Hybrid Modelling Approach to the Assessment of Sustainability of Forest Ecosystems and their Values. Earthscan Ltd. London, UK. 281 pp. ISBN: 978-1-84407-922-3. [Juan A. Blanco, Canada]	guide on model evaluation or modelling in general; we are performing an assessment of the climate-model-evaluation literature. And as this chapter makes clear, there is not yet a scientifically sound uniform approach to model evaluation. Unless there is a specific approach listed in the cited literature that we should have adopted but do not, the proposal has no practical utility.
9-1288	9	59	26	59	33	I like this section, however, I think it would be nice to further explain that the numerical approximation both contains numerical methods and the representation problem in space and time. [Gunilla Svensson, Sweden]	Noted - Given space constraints, we prefer not to expand the paragraph.
9-1289	9	59	35	59	48	delete this para – no use to repeat (part of) the table of contents here. [Andreas Sterl, Netherlands]	Rejected - The special collection of subsections to follow requires explanation. These are given; hence the paragraph is much more than a mere "table of contents".
9-1290	9	59	39	59	39	Tranpose AMIP? What is this? [Farahnaz Khosrawi, Sweden]	Noted - the parantheses clearly announce that it will be explained.
9-1291	9	59	50	60	52	Most of the material in this section has already been covered elsewhere in the chapter. I would suggest deleting this section and merging any additional material from this section in elsewhere. [Nathan Gillett, Canada]	Taken into account - the section is reorganized.Materials in FOD section 9.7.1-3 are largely moved to earlier sections and SOD section 9.7.1 is devoted for a short synthesis for understanding performance of models.
9-1292	9	59	50			Section 9.7.2: The section on so-called “process-oriented evaluation” reflects the community’s use of this term but is fundamentally misleading. There is great confusion as to what this term means and what it should mean. Page 9-60, lines 1-4, for example, uses the terms “process”, “feedback”, “phenomenon”, and “mechanism,” none of which are explicitly defined. Perhaps this ambiguity can be resolved by clearly explaining what each of these terms means and how they differ from one another. I fear, though, that the issues are deeper. A process is a change in some quantity due to relationships among that quantity and perhaps other quantities. Chemical reactions are processes and ozone concentrations are not; autoconversion, collection, or even rain production are processes while rain rate is not. Process contribute to the rate of change of a quantity, so they would be included as sources or sinks in budget equations. This is how I interpret the dictionary definition, i.e. “a continuous ... series of changes taking place in a definite manner.” None of the examples here, in fact, focus on processes. Lines 44-52, for example, explore how the joint distribution of cloud properties, defined on certain subsets of those properties, appear in models and in observations. This tells us about the properties of the clouds themselves (which is of great value), but nothing about the processes (convection, sedimentation, etc.) that led to those distributions. The studies described in lines 28-42, similarly, grade models according to their ability to reproduce various precursors or factors or conditions (temperatures, gas concentrations, dynamical regimes) that affect ozone concentrations. This indeed provides more information than a simple assessment of concentrations but does not directly evaluate the equations that determine ozone concentration. One could argue that the closest thing our community has to process-level evaluation (at least for fast physical processes) is some combination of short-term forecasts (section 9.7.3.1) and/or running parameterizations off-line. Indeed, the latter approach is how physics parameterizations are often developed, and has been standard practice in the community for more than a decade (see the many papers published by participants in GCSS working groups). The kinds of evaluations described here are quite valuable because the focus attention on model performance in a more more finely delineated way than do the grosser metrics described in this chapter. But it is an error to call this “process evaluation.” We do the community a disservice by pretending that we are testing models with	Taken into account - as the section is reorganized, the term 'process-oriented evaluation' is no longer used.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						a stringency we can't live up to. [Robert Pincus, USA]	
9-1293	9	59		68		I think this section (9.7) needs rationalising and focussing. Much of it reads like a repeat of things already mentioned earlier in the chapter, which leads me wondering what it is for. After a short philosophical introduction I think you need to focus down on the principle model biases thought to be most important for global warming, discuss their possible impact and strategies for improvement. Section 9.7.3 seems out of place and superfluous. Any important points would surely be better off in section 9.3. 9.7.4 is more interesting, but elements of this are also repeated in other chapters. [Julia Hargreaves, Japan]	Taken into account - the section is reorganized.Materials in FOD section 9.7.1-3 are largely moved to earlier sections and SOD section 9.7.1 is devoted for a short synthesis for understanding performance of models. FOD section 9.7.4 is also reorganized to become SOD 9.7.2 and now better coordinated with other chapters.
9-1294	9	60	7			It would be more accurate to say that Transpose AMIP tests the behavior of parameterizations when subject to realistic large-scale states. (The initialization is not the crucial point here.) [Robert Pincus, USA]	Taken into account - This section has been removed and its content moved to section 9.2.2.5. A sentence on and references to work using nudging methods has been added to highligh that intial value techbniques are just one method to achieve more realistic large-scale states.
9-1295	9	60	14			It would be useful to point out that the one center which uses the same model (roughly) for weather forecasting and climate projection (the UK Met Office) shows quite good skill in both areas. [Robert Pincus, USA]	Noted - however text moved and substantially revised and shortened
9-1296	9	60	21	60	26	hard to understand why a relationship of storm track motion and "the first spring storm with strong southerly winds over Japan" is a relevant metric. [Bruce Wielicki, USA]	Noted. It was an example of model's performance of a large-scale field such as storm-track activity being a relevant metric of a regional phenomena. As the section is reorganized, this paragraph has been deleted.
9-1297	9	60	28	60	42	This paragraph does not fit well here. Are results derived from CCMs important for further discussions (i.e. in Chapter 10-12)? [Martin Dameris, Germany]	Taken into account. As the section is reorganized, this paragraph has been deleted.
9-1298	9	60	28	60	42	Again a very long section describing CCMVal results… [Gunilla Svensson, Sweden]	Taken into account. As the section is reorganized, this paragraph has been deleted.
9-1299	9	60	48	60	48	Please replace " (Williams and Brooks, 2008) have.." with " Williams and Brooks, 2008 have.." [HASIBUR RAHAMAN, India]	Editorial
9-1300	9	60	48	60	50	A discussion of the time-independence of cloud regime errors should be deferred to section 9.7.3.1. [Robert Pincus, USA]	Taken into account. As the section is reorganized, this paragraph is moved to and merged with the atmosphere section.
9-1301	9	60	48			typo, parenthesis in the wrong place. [Ramon de Elia, Canada]	Editorial
9-1302	9	60	54	64	3	The material in this section has mostly been covered elsewhere in the chapter. Many of the same ideas have already been introduced and much of the same literature cited. 9.7.3.1 is covered in 9.2.2.5. 9.7.3.2 overlaps with 9.4.1.5. 9.7.3.3 contains new material. 9.7.3.4 overlaps with 9.6.2. 9.7.3.5 overlaps with 9.2.2.7. I woudl advocate removing this section, and folding in any material not already covered to other sections of the chapter. If retained, this section could be condensed, cross-referenced to other sections, and made to have a distinct message. [Nathan Gillett, Canada]	Taken into account - the section is reorganized.Materials in FOD section 9.7.1-3 are largely moved to earlier sections and SOD section 9.7.1 is devoted for a short synthesis for understanding performance of models.
9-1303	9	60	56	60	56	What is AMIP? [Farahnaz Khosrawi, Sweden]	Taken into account - This section has been removed and its content moved to section 9.2.2.5. Acronymn no longer used
9-1304	9	60	56	61	30	It is useful to have a section on process oriented evaluation, including "targeted experiments". In this, experiments such as Transpose AMIP are worth mentioning, but half page on it in this assessment seems excessive, considering there are essentially no results -- it is mostly future work. [Robert Colman, Australia]	Taken into account - This section has been removed and its content moved to section 9.2.2.5
9-1305	9	60		60		Section 9.7.2: Massonnet et al. (The Cryosphere, 5, 687-699, 2011) have recently proposed a set of metrics to	Noted - as the section is reorganized, the suggested

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						evaluate model skills in terms of sea ice physics. [Jerome WEISS, France]	paper is considerd in the sea ice section.
9-1306	9	61	16	61	30	This experimental design has been used to diagnose a difference between two versions of a same model in Kim et al. (2012). In Kim et al. (2012), a version of the GISS GCM which exhibit relatively poor representation of the MJO is initialized with restart files from a relatively better version. They demostrated that this experimental design can be used to diagnose the inter-model difference. Reference: Kim, D., A. H. Sobel, A. D. Del Genio, Y. Chen, S. J. Camargo, M.-S. Yao, M. Kelley, and L. Nazarenko: Tropical Intraseasonal Variability Simulated in the NASA GISS General Circulation Model, J. Climate, in press. [Daehyun Kim, USA]	Noted - This section has been removed and its content moved to section 9.2.2.5. T
9-1307	9	61	16		30	I like this paragraph in that it outlines tests and then points out how such tests can be used to test various parameterizations. It would be helpful to me to see a concrete example of the testing-improvement-testing sequence for parameterizations in at least one specific case. This area is well outside my area of expertise so some more detailed examples would help me understand. [Larry Thomason, United States of America]	Taken into account - New literature will be assessed to see if such an example emerges before the submission deadline.
9-1308	9	61	27			More elaboration on the Transpose Amip methodolgy would be helpful. [David Bader, USA]	Taken into account - This section has been removed and its content moved to section 9.2.2.5. Slightly more elaboration on the method itself has been provided.
9-1309	9	61	30			An additional advantage of the transpose-AMIP framework is the ability to take advantage of specialized observation periods (e.g. field programs specific satellite records) and programmatic and infrastructure resources (e.g. targeted analyses), such as Transpose-AMIP does with the YOTC activity (e.g. Moncreiff et al. 2012; Waliser et al. 2012). -------Moncrieff, M.W., D. E. Waliser, M. Miller, M. A. Shapiro, G. R. Asrar, J. Caughey: 2012: Multiscale Convective Organization and the YOTC Virtual Global Field Campaign, Bull. Am. Met. Soc., In Press. --------Waliser, D. E., M. Moncrieff, D. Burrridge, A. Fink, D Gochis, B. N. Goswami, B Guan, P Harr, J Heming, H.-H. Hsu, C Jakob, M. Janiga, R. Johnson, S Jones, P. Knippertz, J Marengo, H Nguyen, M Pope, Y Serra, C Thorncroft, M Wheeler, R. Wood, and S. Yuter, 20112 The "Year" of Tropical Convection (May 2008 to April 2010): Climate Variability and Weather Highlights, Bull. Am. Met. Soc., In Press. [Duane Waliser, USA]	Taken into account - This section has been removed and its content moved to section 9.2.2.5. A sentence highlghting this advantage has been added.
9-1310	9	61	34	61	34	Last time the abbreviation LGM was used was so long ago that it would be worth to write it out once again. However, as previously stated I would recommend not to use this abbreviation at all. [Farahnaz Khosrawi, Sweden]	Editorial. Considered. Part of the material has move to section 9.4.1 and 9.4.2
9-1311	9	61	38	61	38	"atmosphere alone models", is there a way to formulate that better? [Farahnaz Khosrawi, Sweden]	Editorial
9-1312	9	61	41	61	42	Ohgaito and AbeOuchi (2009) instead of "(Ohgaito and AbeOuchi, 20009)". [Farahnaz Khosrawi, Sweden]	Editorial Corrected
9-1313	9	61	48	61	48	What is the abbreviation AMOC standing for? [Farahnaz Khosrawi, Sweden]	Editorial, part of the text moved to 9.4.2
9-1314	9	61	51			A correct simulations of the temperature and salinity fields is' should be 'Correct simulations of the temperature and salinity fields are' [Larry Thomason, United States of America]	Editorial grammar corrected
9-1315	9	61	55	61	57	This does not seem to be model evaluation… [Gunilla Svensson, Sweden]	rejected. This is part of the understanding of model spread and of implication of model biases on simulated climate change.
9-1316	9	62	4	62	4	title reads strangely. [Robert Colman, Australia]	As the section is reorganized, this title has disappered.
9-1317	9	62	4	63	8	By definition, a formulation of the 'sub-grid scale' parameterization should depend on the size of grid cell. We know it but don't know how to do it properly. This point is not discussed and need to be added in this section. [Daehyun Kim, USA]	Noted - but could not adopted due to space limitation.
9-1318	9	62	4			Section 9.7.3.3: I think this section should contain a link to RCMs. Part of the improvements of having high-resolution in GCMs should be already present in RCMs --not of course the large scale features, but small scales--, and hence a comment about this will be welcome. To some practitioners, RCMs are a look into the future resolution of GCMs. [Ramon de Elia, Canada]	Taken into account - the section is reorganized and the resolutional issue is synthetically but briefly discussed here covering both AOGCMs and RCMs based on materials appeared in earlier sections.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1319	9	62	4			Section 9.7.3.3: You could include mention of the step change in the ability of atmospheric models to simulate precipitation distribution and intensity that comes from resolving, rather than parametrising, convection. Clearly this is not yet possible for climate simulations but such studies over limited domains are providing useful information on systematic biases and their link to parametrisations. It is also shows that there are horizontal resolution thresholds across which benefits are realised which justify the related costs, while between these there may not be. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Noted - however, the suggested point is rather indirect to understanding performance of current generation models and not much can be mentioned here. Note that convection resolving simulations are discussed in Chapter 7.
9-1320	9	62	4			Section 9.7.3.3: the benefits and sensitivities related to changing vertical resolution are rather different from those surrounding horizontal resolution and should be discussed separately. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Noted - it is briefly mentioned in 9.1.3.3
9-1321	9	62	11	62	17	It would help to describe these results more abstractly and to provide a synthesis rather than a literature review. [Robert Pincus, USA]	Noted.
9-1322	9	62	21	62	21	In simulations with higher resolution. Rephrase the sentence so that its meaning becomes clearer. [Farahnaz Khosrawi, Sweden]	Noted. As the section is reorganized, the sentence has disappered.
9-1323	9	62	24	62	55	I think these two sections should change place [Gunilla Svensson, Sweden]	Taken into account. As the section is reorganized, the two paragraphs no longer exist in the original form.
9-1324	9	62	24			seems to be a highly relevant reference ---- Jochum, M., R. Murtugudde, R. Ferrari, and P. Malanotte-Rizzoli (2004), The impact of horizontal resolution on the heat budget of the mixed layer in ocean general circulation models, Journal of Climate, 18, 841-851.------ they showed that the chronic cold tongue bias was greatly reduced with higher resolution ocean model so that horiztonal eddies were better represented that mixed warm water under the ITCZ southward into the cold tongue. [Duane Waliser, USA]	Noted - but it is not a post-AR4 literature. Also note that this is an assessment, not a review, and so not all literature on the subject can be cited.
9-1325	9	62	30	62	30	FAQ 8.1. page 8.62; line 30. After the word "forcing".So, it is necessary to stress that the concept of forcing includes: that we are looking at a difference of the radiation (i.e., the e.m. field) between two time of observation) and 2: that we includes in the consequences of this modification of opacity, ,not only the role of direct absorption due to this change of the opacity of this item (i.e. CO 2), but also, the change in the water vapor, with its particular properties of change of phase, and so, in humidity, specific and relative, and the consequences on opacity , and then, on the e.m. field. So we see the importance of this concept of " forcing" [Robert DAUTRAY, France]	Forwarded to Chapter 8.
9-1326	9	62	33			McClean, et al 2011 (already in refernce list) demonstarted the importance of an a fully-coupled high resolution ocean-atmosphere model for the correct simulation of Agulhas eddies as opposed to high resolution ocean simulations alone. [David Bader, USA]	Noted - however, we have decided to focus more on understanding performance of models used in later chapters. As the suggested work is on higher resolution simulation than any CMIP5 model, not much can be mentioned here.
9-1327	9	62	35	62	37	I was surprised to find such a restricted discussion of the potential role of tropical-cyclone mixing given the wealth of recent research in this area, and the potential important of this as a missing process in climate models, as well as lack of attribution of seminal studies in this area that predate the single study (Vincent et al, submitted) cited. This topic deserves a more comprehensive treatment. see e.g. Sriver and Huber (2007) [Sriver, R. L., and M. Huber (2007), Observational evidence for an ocean heat pump induced by tropical cyclones, Nature, 447, 577–580, doi:10.1038/nature05785.Sriver et al (2010] , Korty et al (2008) [Korty, R. L, K. A. Emanuel, and J. R. Scott (2008), Tropical cyclone induced upper-ocean mixing and climate: Application to equable climates, J. Clim., 21, 638–654, doi:10.1175/2007JCLI659.1], and Sriver et al (2010) [Sriver, R. L., M. Goes, M. E. Mann, and K. Keller (2010), Climate response to tropical cyclone-induced ocean mixing in an Earth system model of intermediate complexity, J. Geophys. Res., 115, C10042, doi:10.1029/2010JC006106]. [Michael Mann, USA]	Noted - As the section is reorganized, this paragraph has been deleted. Also note that this topic was mentioned here only in the context of resolution dependence and, thus, what can be cited here was limited.
9-1328	9	63	1	63	1	… improvement of resolution (both horizontal and vertical) is ... [Claudio Cassardo, Italy]	Noted - As the section is reorganized, this sentence has disappered.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1329	9	63	6	63	8	Since we do not exactly know why a model give a ENSO it is not quite clear that it is the uncertainties in parameterisations that change the amplitude, that is a speculation. [Gunilla Svensson, Sweden]	Noted - As the section is reorganized, this paragraph has been deleted. the intention was that, since parameterizations mean anything but resolved processes, what cannot be improved by improving resolution should be due to uncertainties in parameterizations.
9-1330	9	63	7	63	8	Or possibly, that parameterizations formulated for low resolution models are not suitable for high resolution models? [Elisa Manzini, Germany]	Noted - As the section is reorganized, this paragraph has been deleted. the intention was that the suggested case is coverd by the idea of "uncertainties in parameterizations".
9-1331	9	63	10	63	10	Write in the section title "Regional Climate Models (RCMs)" instead of just "RCMs". [Farahnaz Khosrawi, Sweden]	Editorial.
9-1332	9	63	10			Section 9.7.3.4: You could mention here the potential benefits of using RCMs (forced with reanalyses at the boundaries) alongside global models to identify local versus remote forcing of systematic biases. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. We have elaborated on this application of RCMs while revising Section 9.6 of the chapter.
9-1333	9	63	10			Sect.9.7.3.4: Another use of RCMs that fits nicely with the theme of sect.9.7 is their use as an experimental tool to understand the key processes and uncertainties of global models, by using their regional character to adjust boundary conditions etc. Examples are Rowell and Jones (2006) and Bielle et al. (2009) though there are also many others. Rowell, D.P. and Jones, R.G., 2006: Causes and uncertainty of future summer drying over Europe. Climate Dynamics, 27, 281-299. Bielli, S., Douville, H. & Pohl, B. (2009) Understanding the West African monsoon variability and its remote effects: an illustration of the grid point nudging methodology. Climate Dynamics, 35. 159–174. [David Rowell, United Kingdom of Great Britain & Northern Ireland]	Taken into account. We have elaborated on the investigation of processes by means of RCMs, while revising Section 9.6 of the chapter.
9-1334	9	63	12	63	18	Why is not CORDEX mentioned here since it is mentioned explicitly earlier, a publication would be nice. Why is not ARCMIP mentioned, there are several papers published on several aspects of RCMs for the Arctic within this e.g. Tjernström, Michael, Joseph Sedlar, Matthew D. Shupe, 2008: How Well Do Regional Climate Models Reproduce Radiation and Clouds in the Arctic? An Evaluation of ARCMIP Simulations. J. Appl. Meteor. Climatol., 47, 2405–2422. Tjernström, M., M. Žagar, G. Svensson, J Cassano, S. Pfeifer, A. Rinke, K. Wyser, K. Dethloff, C. Jones and T. Semmler, 2005: Modeling the Arctic Boundary Layer: An evaluation of six ARCMIP regional-scale models with data from the SHEBA project. Boundary-Layer Meteorology, 117, 337-381. Wyser, K.; Jones, C.; Du, P.; Girard, E.; Willén, U.; Cassano, J.; Christensen, J.H.; Curry J.A.; Dethloff, K.;Haugen J.E.; Jacob, D.; Koltzow, M.; Laprise, R.; Lynch, A.; Pfeifer, S.; Rinke, A.; Serreze, M.; Shaw, M.J.; Tjernström, M.; Zagar, M., 2008: An Evaluation of Arctic Cloud and Radiation processes during the SHEBA year: Simulation results from 8 Arctic Regional Climate Models, Clim. Dyn. 30, 203-223, doi:10.1007/s00382-007-0286-1 Inoue, J., Liu, J., Pinto, J.O., Curry, J.A., 2006: Intercomparison of Arctic Regional Climate Models: Modeling Clouds and Radiation for SHEBA in May 1998, J. Clim. 19, 4167-4178 Rinke, A.; Dethloff, K.; Cassano, J.; Christensen, J.H.; Curry J.A.; Du, P.; Girard, E.; Haugen, J.E; Jacob, D.; Jones, C.G.; Koltzow, M.; Laprise, R.; Lynch, A.H.; Pfeifer, S.; Serreze, M.C.; Shaw, M.J.; Tjernström, M.; Wyser, K.; Zagar, M., 2006: Evaluation of an ensemble of Arctic regional climate models: spatiotemporal fields during the SHEBA year, Clim. Dyn. 26, 459-472, doi:10.1007/s00382–005-0095-3 [Gunilla Svensson, Sweden]	Accepted. We elaborated on the assessment of coordinated RCM experiments that have been published on since the literature cutoff date for the AR4. Specifically on CORDEX, a few papers have since the preparation of the FOD become available.
9-1335	9	63	21	63	21	"which and may"? Something went wrong in this sentence. [Farahnaz Khosrawi, Sweden]	Editorial.
9-1336	9	63	25		28	please give a cite for the good performance of the multi model approach [Frank Kreienkamp, Germany]	Taken into account. We have elaborated on the results of coordinated RCM experiments in Section 9.6 of the chapter.
9-1337	9	63	30	63	30	Add PPE in paranthesis, thus "(PPE)" to the title. [Farahnaz Khosrawi, Sweden]	Accepted.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1338	9	63	35	63	35	What are EMICs? What is the abbreviation standing for? [Farahnaz Khosrawi, Sweden]	Taken into account. EMICs are defined as Earth-system Models of Intermediate Complexity and introduced in Section 9.1 (pg 9-6 FOD)
9-1339	9	63	47	63	48	The following paper should be added here: Loutre et al., 2011, Evaluating climate model performance with various parameter sets using observations over the last centuries, Climate of the Past, 7, 511-526, doi:10.5194/cp-7-511-2011. [Thierry Fichefet, Belgium]	Accepted. The reference will be added.
9-1340	9	63	50			Yokohata et al 2011 demonstrates a clear and substantial difference in performance between the two (CMIP3-based) MMEs and 4 single-model PPEs tested. Hargreaves et al 2011 further strengthened this result for a single MME and PPE, using paleoclimate simulations. Your assessment of the literature here seems somewhat partial and favours the use of PPEs more strongly than the literature supports (there's also a Klocke and Pincus paper, and the earlier Yokohata et al 2010, that point to limitations of PPEs). I'm not sure of the basis for your unreferenced claim "there is evidence that the experimental design of a PPE can be controlled to more closely mimic the multi-model case". It surely needs a citation, and chapter 12 p12 appears to directly contradict the claim. [James Annan, Japan]	Taken into account. This material is now discussed in Section 9.2.2.7.
9-1341	9	63	51	63	51	CFMIP? Has the abbreviation used before? What is it standing for? [Farahnaz Khosrawi, Sweden]	Noted. CFMIP stands for Cloud Feedback Model Intercomparison Project but is no longer used in the chapter.
9-1342	9	63	55	63	55	What is MME standing for? [Farahnaz Khosrawi, Sweden]	Noted - MME stands for Multi-Model Ensemble and was defined earlier in the chapter on page 9-18 in Section 9.2.2.7.
9-1343	9	63	56	64	3	… there is evidence …: where? Can you refer some paper for this sentence? [Claudio Cassardo, Italy]	Taken into account. This sentence refers to two missing references. They are: Collins et al. (2010, DOI 10.1007/s00382-010-0808-0) and Yokohata et al. (2011, DOI 10.1007/s00382-011-1203-1). These were added in the revision and moved into Section 9.2.2.7.
9-1344	9	64	1	64	3	It is not clear how the comparison bewteen the error structure of MMEs and PPEs may lead to a better characterization of model uncertainty. [Daehyun Kim, USA]	Taken into account. This sentence refers to two missing references that address the characterization of uncertainty specifically. They are: Collins et al. (2010, DOI 10.1007/s00382-010-0808-0) and Yokohata et al. (2011, DOI 10.1007/s00382-011-1203-1). These were added in the revision and moved to Section 9.2.2.7.
9-1345	9	64	5	68	43	A figure that presents schematically which feedback processes that has been evaluated in this chapter should be included in Section 9.7.4 Climate Sensitivity and Climate Feedbacks. You may consider to move Chapter 1 Figure 1-2 to this section, make a new figure or include a reference to Figure 1-2. [Øyvind Christophersen, Norway]	Noted - Will consider referring to Chapter 1 or to Chapter 7.
9-1346	9	64	5			Section 9.7.4: This material seems inappropriate in a chapter on model evaluation. It may make sense to include section 9.7.4.1, which examines the robustness of different ways of estimating climate sensitivity, since this is consistent with trying to understand how model results may best be interpreted. But the models can not be evaluated on the basis of their climate sensitivity, so it is unclear why this chapter should include discussion of how various processes affect climate sensitivity. The link to model evaluation should be made more explicit or the (relevant and interesting) material moved to another chapter. [Robert Pincus, USA]	Taken into account. Note that 9.7.1 justifies comprehensively why assessing model spread in climate sensitivity and feedbacks is crucial for understanding model spread in projections. Chapter 7, in particular, does not do the global account of spread in cloud feedbacks. Chapter 9 SOD also shows how the spread in climate sensitivity arises from the various feedbacks, thus offering a proximate explanation.
9-1347	9	64	10	64	15	There is a large discrepancy between the satellite observations and the behavior of the IPCC climate models	Taken into account (Section 9.4.1 and Chapter 12).

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						on how the Earth loses energy as the surface temperature changes. The following studies imply that the modeled climate sensitivity to CO2 is largely overestimated by the IPCC models. Singer, S.F., 2011. Lack of consistency between modeled and observed temperature trends. Energy & Environment 22 (4), 375–406. The second is: Spencer, R.W., Braswell, W.D., 2011. On the misdiagnosis of surface temperature feedbacks from variations in earth's radiant energy balance. Remote Sensing 3, 1603–1613. [James Wanliss, USA]	
9-1348	9	64	10	64	53	This section is confusing. I think the text would benefit from breaking to a new section in the middle of line 25. Then, a statement that climate sensitivity is diagnosed using two different methods. Explain from the beginning very clearly that the new way is to use a quadrupling but the definition is for a doubling. In the equation it is confusing that alpha is used and not 1/lambda as usually is used. [Gunilla Svensson, Sweden]	Accepted - section reworded for clarity.
9-1349	9	64	17	64	28	What has exactly done? I cannot follow. It is clear that this is quite good, but it should be clarified. Non eperts don't know that. [Farahnaz Khosrawi, Sweden]	Taken into account. Please consult published literature.
9-1350	9	64	18	64	18	There? Which report? AR4 or AR5? [Farahnaz Khosrawi, Sweden]	Accepted - text revised
9-1351	9	64	20	64	20	Fixed what? [Farahnaz Khosrawi, Sweden]	Accepted - text revised
9-1352	9	64	24			the phrase …"the agreement was within 10% or even less…" could possibly be reworded. Should it be the agreement was within 90% or greater, or possibly the disagreement was within 10% or less? [David Bader, USA]	Accepted - text revised
9-1353	9	64	26	64	26	Gregory et al. (2004) instead of (Gregory et al, 2004). [Farahnaz Khosrawi, Sweden]	Editorial
9-1354	9	64	30	64	30	There is a mistake in the formula rendering in pdf. [Claudio Cassardo, Italy]	Editorial
9-1355	9	64	30	64	30	something wrong with the equation [James Christian, Canada]	Editorial
9-1356	9	64	30	64	30	Fix equation. [Robert Colman, Australia]	Editorial
9-1357	9	64	30	64	30	Something went wrong with the formatting of the formula. On my prinntout it was completely unreadable. [Farahnaz Khosrawi, Sweden]	Editorial
9-1358	9	64	30	64	30	Is giving the formula and discussing it really helpful for the reader? [Farahnaz Khosrawi, Sweden]	Taken into account - logic unchanged
9-1359	9	64	30	64	30	Equation (9.1) is not appropriately written. [Daehyun Kim, USA]	Editorial
9-1360	9	64	30	64	30	needs to be improved. [Zhaomin Wang, UK]	Editorial
9-1361	9	64	34	64	36	This isn't quite correct since alpha is the inverse of the slope of the line [Graham Feingold, United States of America]	Taken into account - alpha is the inverse climate sensitivity parameter, but has been revised for clarity.
9-1362	9	64	34			Why is the "CO2 concentration … quadruplet" for CMIP5. This does not follow, nor appears motivated, from the changes in protocol desctibed. [Tor Eldevik, Norway]	Taken into account; however, there is no account available in the published literature explaining why quadrupling instead of doubling CO2 was applied.
9-1363	9	64	36	64	36	Where is this formula coming from? References? [Farahnaz Khosrawi, Sweden]	Please consult literature cited.
9-1364	9	64	37	64	38	The logarithmic dependence is questionable if the greenhouse effect is almost saturated. It would be valid only in the absorption band edges. Nearer the absorption peaks, a few meters thick air layer just above ground or sea is already opaque to the earth radiation. There is no longer, therefore, dependence on CO2 concentration. The approximation was proposed at the time of the Arrhenius's papers when spectrometers had limited available spectral range in the infrared and very poor performances. It has now to be reconsidered under the light of modern available data. [François GERVAIS, France]	Taken into account (see 9-1367)
9-1365	9	64	38	64	45	This sentence is heavy going. Break up and re-express. [Robert Colman, Australia]	Accepted - text revised for clarity.
9-1366	9	64	38	64	45	This sentence is quite long and should be split into two sentences. [Farahnaz Khosrawi, Sweden]	Accepted - text revised for clarity.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1367	9	64	38		47	The language seems ambiguous at best and perhaps erroneous. One might read it to state that there are two counterexamples to the logarithmic dependence of forcing on CO2. I looked at both papers (Gregory, Li) if those papers stated that, and did not find that they did. So it is not clear what these papers are being held up as counter examples to. Revision should be more specific to the point being made. [Stephen E Schwartz, USA]	Accepted - text revised to make clear that it is the logarithmic dependence of the temperature response to CO2 that is meant.
9-1368	9	64	40	64	43	this text is hard to follow [Graham Feingold, United States of America]	Accepted - text revised for clarity.
9-1369	9	64	45	64	47	The reference should be given without brackets. [Farahnaz Khosrawi, Sweden]	Editorial
9-1370	9	64	49	64	49	There is a problem in the text. [Claudio Cassardo, Italy]	Editorial
9-1371	9	64	49	64	49	Solve the problem causing this error message in the text. [Farahnaz Khosrawi, Sweden]	Editorial
9-1372	9	64	49	64	53	The first method is decribed extensively while the second is just mentioned? [Farahnaz Khosrawi, Sweden]	Taken into account. Both methods are described to the detail required. No change.
9-1373	9	64	51	64	52	A full discussion of these differences and their sources is given in Colman, R.A. and B.J. McAvaney, 2011: On tropospheric adjustment to forcing and climate feedbacks. Climate Dynamics, 36, 1649-1658, doi: 10.1007/s00382-011-1067-4. [Robert Colman, Australia]	Accepted - reference added.
9-1374	9	64	51	64	53	That is not clear. Wasn't it said that before that CO2 was quadrupling? Are both methods using quadrupling? [Farahnaz Khosrawi, Sweden]	Taken into account. Please consult glossary - this paragraph is about radiative forcing.
9-1375	9	64	51	64	61	It did not become clear what the difference between the two methods is. [Farahnaz Khosrawi, Sweden]	Taken into account. Please consult published literature.
9-1376	9	64	57	64	57	What do you mean with "1% compound per year"? [Farahnaz Khosrawi, Sweden]	Taken into account. Word 'compound' deleted for clarity
9-1377	9	65	5	65	18	The paragraph labels models as" CMIP5" models or "AR4" models. More correctly, the CMIP5 models should be compared to CMIP3 models. [David Bader, USA]	Taken into account - will ensure consistency wrt. CMIP3 vs. CMIP5; or AR4 vs. AR5.
9-1378	9	65	5	65	18	Where is the spread coming from? If it is not caused by the method than it must be due to the model performance/set-up? [Farahnaz Khosrawi, Sweden]	Taken into account - to be dealt with in later subsections.
9-1379	9	65	11	65	11	"from 2.1 K to 4.6 K" Which method? [Elisa Manzini, Germany]	Accepted - text revised for clarity.
9-1380	9	65	11	65	18	It may be noteworthy that the spread in climate sensitivity has not increased since AR4 despite the introduction of ESMs. This may be because the spread in sensitivity between the physical models has reduced or, possibly more likely, the effect of the earth system components on the spread in climate sensitivity is small compared with traditional factors (e.g. clouds). [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Add sentence concerning inclusion of ESMs (however, in the RCPs, because of prescribed concentrations, there is no reason why the ESMs should behave differently).
9-1381	9	65	21	65	21	table 9.2: to be expanded including all possible models, right? [ANNALISA CHERCHI, Italy]	Noted - will be expanded to the extent possible.
9-1382	9	65	21	65	26	CO2 quadrupling by two? Contradiction? As it is written now it is quite confusing and misleading. [Farahnaz Khosrawi, Sweden]	Accepted - text revised for clarity.
9-1383	9	65	21	65	26	Please rewrite the table caption, it is also confusing. It does not make sense that the table heading includes Radiative forcing with the unit Wm-2, that's not what is in the table, right? The unit on the other columns is K. I am totally confused, climate sensitivity should have K/Wm-2 [Gunilla Svensson, Sweden]	Caption revised for clarity. Concerning definitions, please consult glossary. All units are used correctly, but table now explicitly and separately lists both climate sensitivity and climate sensitivity parameter.
9-1384	9	65	22	65	22	Hanson et al. (2005) instead of (Hanson et al., 2005). [Farahnaz Khosrawi, Sweden]	Editorial
9-1385	9	65	26	65	27	As the spread between models is larger than 1 Wm-2 or 1K, I suggest to put only one decimal digit after the point: I am not sure the second one is significant (may be neither the first one…). [Claudio Cassardo, Italy]	Accepted.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1386	9	65	26	65	27	Recommend presenting sensitivities in units K/(Wm-2) instead of (preferably) or in addition to (for historical continuity) units of K per CO2 doubling. Use of systematic units always advances the science; we report pressure in N m-2 or pascal in lieu of Torr because densitiy of mercury needs to be specified, etc. The problem is much worse when the unit of measurement, here temperature response per CO2 doubling is based on a quantity, CO2 doubling forcing, that is itself uncertain. Here fractional standard deviation of data in col 4 is 23% vs 28% or 33%, resp, for the sensitivity normalized by the forcings in cols 2 or 3 resp., indicative of some compensation between forcing and sensitivity, perhaps indicating a concurrence among models that is better than is actually the case. [Stephen E Schwartz, USA]	Accepted - added climate sensitivity parameter as a column. However, kept climate sensitivty to ensure continuity with previous reports. Concerning compensation of forcing and sensitivity, not published literature seems to be available on that.
9-1387	9	65	26	65	27	Table would be much more valuable if extended to all AR5 models [Stephen E Schwartz, USA]	Taken into account - has been expanded to the extent possible.
9-1388	9	65	29			Section 9.7.4.2: this section seems out of place, as Section 9.7.4.3 more clearly follows on from section 9.7.4.1. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Noted. Section structure and text revised.
9-1389	9	65				Table 9.2 - It appears that practically with the same forcing inputs but only changing the methodology to treat them, model HadGEM2-ES gives an Equilibrium Climate Sensitivity 120 % higher than model INM-CM4. This raises severe questions about the global model projections. Model INM-CM4 would comply with an objective to limit the warming to 2°, not the other, and only based on different methodologies ? [François GERVAIS, France]	Noted - this spread in climate sensitivity among models is why we comprehensively assess climate sensitivity and feedbacks in this section.
9-1390	9	66	1	66	1	The abbreviation has already been used several times and now finally it is written what it means. That should have been done much earlier. [Farahnaz Khosrawi, Sweden]	Rejected -- EMIC is defined in the executive summary and section 9.1.1 and there is a whole section devoted to the topic (with acronym expansion) in 9.1.2.3.
9-1391	9	66	12	66	12	What are the abbreviations ECS, TCR and MME standing for? [Farahnaz Khosrawi, Sweden]	Accepted -- TCR and MME acronyms expanded and explained. ECS is expanded in the previous paragraph.
9-1392	9	66	13	66	18	The LOVECLIM model also includes an ice-sheet component. This EMIC, whose results are discussed in several chapters, should be included in Table 9.3. [Thierry Fichefet, Belgium]	Accepted -- model added to Table 9.3
9-1393	9	66	16	66	16	It is not clear what is meant here. [Farahnaz Khosrawi, Sweden]	Accepted -- Results of EMIC/AOGCM comparison have been clarified.
9-1394	9	66	23	66	57	This section should make connections with the extensive discussion of the same topic in Chapter 7. [David Randall, USA]	Accepted -- The connections as discussed during the LA3 with Chapter 7 CLAs have been implemented.
9-1395	9	66	27	66	27	This should be the variations in water vapour and lapse rate combined. Individually their inter-model spread can be fairly large. [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Accepted -- These two feedbacks are now combined in this sentence.
9-1396	9	66	30	66	30	Write CO2 doubling instead of 2 x CO2. [Farahnaz Khosrawi, Sweden]	Editorial -- Accepted
9-1397	9	66	30	66	30	What is meant with "to 1% CO2 yr-1"? An increase? [Farahnaz Khosrawi, Sweden]	Editorial -- Accepted and clarified through expansion of 1%CO2 yr^-1 into words.
9-1398	9	66	32	66	35	The points made here for the MIROC and HADsm ensembles are also visible in an ensemble of ECHAM 5 (doi:10.1175/2011JCLI4193.1). [Robert Pincus, USA]	Accepted -- Reference to ECHAM5 ensemble added.
9-1399	9	66	35	66	36	Shiogama et al. (2011) suggested that SW middle-level cloud feedback mainly causes the low climate sensitivity of MIROC5. Shiogama, H., et al. (2011) Physics Parameter Uncertainty and Observational Constraints of Climate Feedback: An Ensemble of Coupled Atmosphere Ocean GCM without Flux Corrections. Climate Dynamics, submitted [Hideo Shiogama, Japan]	Accepted -- This additional information regarding the origin of the mid cloud feedbacks in MIROC5 has been added.
9-1400	9	66	38	66	38	What are radiative kernel techniques? [Farahnaz Khosrawi, Sweden]	Taken into account -- combined with comment 9-1401

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1401	9	66	38			Many readers will need at least a little explanation as to what the “radiative kernel technique” is. It is also worth noting that cloud feedbacks are traditionally estimated from residuals in this method, making it harder to interpret the rough agreement among models. The next draft will certainly want to note progress on cloud radiative kernels (doi:10.1175/JCLI-D-11-00248.1, doi:10.1175/JCLI-D-11-00249.1; also forthcoming week by B. Sanderson and K. Shell). [Robert Pincus, USA]	Accepted -- Radiation kernels are described in more detail, and recommended references with updates to method have been included.
9-1402	9	66	44	66	44	Dessler, A.E. (2010) Science, 330, 1523-1527 is a better reference for this. Need to add a note that interannual/climate change analogues are not necessarily close. [Robert Colman, Australia]	Accepted -- Reference updated.
9-1403	9	66	48	66	48	cloud entrainment rate also shown to be important in Sexton et al (2011a). [David Sexton, UK]	Acceoted -- Results from PPEs included.
9-1404	9	66	50	66	53	the jump from shallow clouds to cold clouds is very abrupt. [Graham Feingold, United States of America]	Accepted -- Introductory sentence for paragraph added to explain transition to other cloud types beside low clouds.
9-1405	9	66	53	66	55	Comment on the line "Analyses of the tendencies … (Ogura et al., 2008)." The accuracy and completeness of the text might increase by changing the expression as follows; "The budget terms of the cloud condensate tendency equation in GCMs are analysed when multiple models are subjected to a CO2 increase. The results show that inter-model differences in the cloud response are attibutable to processes such as condensation-evaporation and ice sedimentation (Ogura et al. 2008a,b).". Reference to the above papers is as follows; Ogura, T., S. Emori, M. J. Webb, Y. Tsushima, T. Yokohata, A. Abe-Ouchi, and M. Kimoto (2008a): Towards understanding cloud response in atmospheric GCMs: the use of tendency diagnostics. J. Meteor. Soc. Japan, 86(1), 69-79. Ogura, T., M. J. Webb, A. Bodas-Salcedo, K. D. Williams, T. Yokohata, and D. R. Wilson (2008b): Comparison of cloud response to CO2 doubling in two GCMs. SOLA, 4, pp29-32, doi:10.2151/sola.2008-008. [Tomoo Ogura, Japan]	Accepted -- Sentence altered and additional reference added.
9-1406	9	66	56			There is quite some progress on model representation of the ice content of clouds that both fall speed and entrainment are directly relevant to. A study by Li et al JGR submitted (can send to relevant LAs) shows very nicely via a Taylor diagram how models compare to observations (including some idea of observational uncertainty) and the marked progress that has occurred from CMIP3 to CMIP5 with mldels tuning towards the observations. I think more concrete examples like this are needed in this chapter as I fiind the assessments given too broad brushed. I should note that model assesment of cloud properties is really skimmed over in this chapter and not really addressed in chapter 7 so it sort of falls between cracks so I think some sort of coordination between ch7 and ch 9 is needed to determine where/how much of this is given. [Graeme Stephens, USA]	Taken into account
9-1407	9	66	57	66	57	Effect of ice fall out speed is also now discussed in Sexton et al (2011a). [David Sexton, UK]	Acceoted -- Results from PPEs included.
9-1408	9	67	1	67	57	Again, a missed opportunity to discuss the role of surface albedo feedback in models and observations. Several papers by Alex Hall and colleagues are relevant to the problem. [Philip Mote, USA]	Accepted -- References to and new section on surface albedo analyses added.
9-1409	9	67	2			Section 9.7.4.3: It has been suggested that more coordinated parameter tuning associated with climate sensitivity and aerosol forcing be taken (see p.823 of Tanaka and Raddatz, 2011, Climatic Change Letters, http://www.springerlink.com/content/k160023102g83v4v/). Does this contribute to the discussion here? [Katsumasa Tanaka, Switzerland]	Taken into account -- The proposal to jointly tune climate sensitivity and aerosol forcing is not relevant.
9-1410	9	67	4	67	4	which models? [ANNALISA CHERCHI, Italy]	Taken into account -- The range quoted is the range cited in AR4, and a reference to discussion of the range in the AR4 WG1 report has been added. Enumeration of models is unnecessary.
9-1411	9	67	4	67	11	"another mechanism" unclear: If the models used in calculating the sensitivity and the historical runs are the same, why this "another mechanism" is not acting in the sensitivity runs? The full paragraph is cryptic. Please	Accepted -- Sentence rephrased so that it is clear that the mechanism could appear in the all-forcing

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						explain better. [Elisa Manzini, Germany]	historical runs but non in the CO2-only sensitivity runs.
9-1412	9	67	4	67	11	Is this still true in CMIP5? [David Randall, USA]	Accepted -- Based upon Forster (2012), the evidence for compensation between sensitivity and forcing required to reproduce the 20th century TCR is much weaker. The Forster paper is cited here.
9-1413	9	67	4	67	11	this paragraph argues for a potential increase in range of sensitivity but then quotes the same range of 2.1 to 4.4C: typo? [Bruce Wielicki, USA]	Taken into account - citation to Huybers was not appropriate (it concerns compensation among feedbacks, not between equilibrium climate sensitivity and aerosol forcing) and has been removed.
9-1414	9	67	4	67	22	this section seems inconsistent with a recent paper by Hansen et al. in Atmos Chem and Physics 2011 "Earth's energy imbalance and implications". in this paper they discuss the relationship between the time scale of climate model response to forcing, ocean mixing (claiming that most GCMs have excessive vertical mixing and longer response timescales), aerosol forcing, and global net energy imbalance. [Bruce Wielicki, USA]	Accepted -- Discussion of Hansen and importance of uncertainties in ocean mixing processes have been added.
9-1415	9	67	9	67	11	The logic is unclear here. Why would a major expansion of climate sensitivity go to 2.1-4.4C, when the conclusion just a little earlier was that the current range from the models is 2.1-4.6C?. Indeed this point seems an important one, with unclear implications for sensitivity. I suggest this section be expanded and clarified. [Robert Colman, Australia]	Taken into account -- combined with comment 9-1411
9-1416	9	67	11	67	11	Makes no sense as written – unless "to" should be "from", the numbers need replacing. [William Ingram, UK]	Taken into account -- combined with comment 9-1411
9-1417	9	67	11	67	11	Same range as above. I would have expected that you give here a minor and major expansion range. [Farahnaz Khosrawi, Sweden]	Taken into account -- combined with comment 9-1411
9-1418	9	67	13	67	22	This section is severely underdone, considering the central importance of water vapour and lapse rate changes in determining climate sensitivity, and consequently the central importance of assessing our confidence in those feedbacks for our confidence in overall projections of climate change. Furthermore, focusing on SST's throughout this paragraph and the coincident impact on regional OLR misses the main point about the global scale nature of the feedback, and the fact that SST 'fluctuations' are not necessarily correlated with changes to the overlying atmosphere (due to circulation changes etc), and that these analogues must be treated carefully. How much these feedbacks conrtibute to sensitivity is not discussed. Lapse rate feedbacks are not even mentioned, despite the title. This section needs a complete rewrite, emphasising: (1). The magnitude of the impact on climate sensitivity of water vapour/lapse rate feedback (Bony et al 2006, Randall et al, 2007, Soden and Held, 2006) (2). The offsetting relationship between water vapour/lapse rate feedbacks, with its implication for climate sensitivity (Soden and Held, Colman, 2003). (3) The basis of these feedbacks in quasi-unchanging relative humidity, our improved understanding of the robustness of these (Ingram, 2010, 2011, Sherwood et al, 2010), and the observational support on seasonal, inter annual and climate change timescales (Chung et al, 2010, Dessler and Wong, 2009; Minschwaner et al, 2006, Soden ,….), and, yes, regionally too (Chung et al 2010). (4). Our knowledge of importance of upper tropospheric humidity in particular, and our means of detecting this and testing models, including satellite simulations. [Robert Colman, Australia]	Accepted -- the section has been extended with new literature on offsetting between feedbacks formulated in terms of lapse rate and specific humidity
9-1419	9	67	13	67	22	Section 9.7.4.3.1: "Role of humidity and lapse rate feedbacks in climate sensitivity" in "Climate Sensitivity and Climate Feedbacks" could mention the link between the water vapour feedback on climate change & control climate established by Ingram (2010), if only for the theoretical advantage of such a link, as no practical implementation has been achieved. Text could be like "One direct link between an important component of climate change and control climate has been established: most of the water vapour feedback on climate change is attributable to a component which depends only on control climate (Ingram, 2010). This component is not directly observable, however, and does not seem to account for a large fraction of the uncertainty in climate sensitivity (Ingram, 2012)." The references are Ingram, W., 2010: A very simple model for the water	Taken into account -- combined with 9-1418

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						vapour feedback on climate change. Q. J. R. Meteorol. Soc., 136, 30-40. available at http://onlinelibrary.wiley.com/doi/10.1002/qj.546/abstract and W. J. Ingram, 2012: A new way of quantifying GCM water vapour feedback. Clim. Dyn. 2012, 10.1007/s00382-012-1294-3 (in press) - proofs available for IPCC review purposes as pub/user/ingram/pS/Ingram2012newway.pdf by anonymous ftp into ftp.atm.ox.ac.uk. (This comment also applies to 9.2.2.2 & is repeated there.) [William Ingram, UK]	
9-1420	9	67	15	67	15	§ 9.7.4.1, PAGE 9.67; LINE 15 Moreover, the theorem of fluctuation-dissipation, [Andrew Majda, Rafail Abramov, Marcus Grote: Information theory and stochastics for multiscale non linear systems; American Mathematical Society; 2005; chapter 2: "Roughly speaking, this theorem state s that for systems in statistical equilibrium, the averaged mean response to small external perturbation can be calculated through the knowledge of suitable correlation functions of the unperturbed statistical system"] useful for a physical system far of equilibrium [Derrida and al; CRAS; série physique xxx] permit us to simplify the computation the responses of our climatic climate to different forcings using a single operator [Andrew Majda..PNAS] [Robert DAUTRAY, France]	Noted
9-1421	9	67	19			"… to global scales (e.g. Allan, 2009)." (to help back up the preceding sentence with a reference) [Allan, R. P. (2009), Examination of Relationships between Clear-Sky Longwave Radiation and Aspects of the Atmospheric Hydrological Cycle in Climate Models, Reanalyses, and Observations, J. Climate, 22, 3127-3145] [Richard Allan, UK]	Accepted -- Reference added.
9-1422	9	67	22			I. Held and a co-author have recent work on combining the water vapor and lapse rate feedback to show that the overall contribution to uncertainty is much less than might be assumed from magnitudes assessed as if the feedbacks were independent. This seems worth citing here. [J. David Neelin, United States]	Taken into account -- combined with 9-1418
9-1423	9	67	31	67	31	The effects of long-term processes, such as ice sheet melting, on climate sensitivity should be mentioned here. See, e.g., Goelzer et al., 2011, Impact of Greenland and Antarctic ice sheet interactions on climate sensitivity, Clim. Dyn., 37, 1005-1010, doi:1007/s00382-010-0885-0. [Thierry Fichefet, Belgium]	Taken into account -- Timescales of interest (centennial) clarified.
9-1424	9	67	33	47	57	Why is this material in a chapter on model evaluation? It may be worth discussing formal parameter estimation as a model evaluation method, although these methods are designed to produce the best model behavior subject to structural errors, and not the most realistic parameter values. [Robert Pincus, USA]	Taken into account -- This material spans multiple components and processes and hence needs to be included here rather than in earlier, more specialized chapters. Chapter 9 is also focused on connecting model diversity/uncertainty to the underlying model formulation, hence this material is approprriate in the context of explaining the sensitivity range of the CMIP5 MME.
9-1425	9	67	36		39	Clarify this sentence [Robert Colman, Australia]	Accepted -- sentence revised.
9-1426	9	67	38	67	38	What exactly? It sounds for me like an uncomplete sentence. [Farahnaz Khosrawi, Sweden]	Taken into account -- combined with comment 9-1425.
9-1427	9	67	41	67	45	There is still debate on this matter (see Roe and Armour, 2011; Roe and Baker, 2011) and the "non-linear" arguments in Zaliapin and Ghil 2010 do not invalidate the issue brought up in RB07 since the non-linear terms related to df/dT appear relatively small in the modern climate, nor is there evidence of substantial bifurcation structure given the relatively small magnitude of climate change we are talking about (as opposed to say, a runaway greenhouse or snowball Earth bifurcation). There is also a typo in this last sentence. [Chris Colose, United States]	Accepted -- References to latest Roe and Baker papers added and ongoing debate noted.
9-1428	9	67	42	67	44	In fairness to the older literature, and perhaps to strengthen the point, the strong dependence of sensitivity to feedbacks through the feedback factor "discovered" by Roe and Baker was noted and well understood by Hansen et al (1984 AGU monograph) and Schlesinger (1986, Clim Dynamics) [Stephen E Schwartz, USA]	Accepted -- Sentence rephrased so that Roe and Baker's findings are described as a rediscovery.
9-1429	9	67	43	67	45	This sentence does not make sense. Also clarify why these are 'artefacts' and what the possible implications are of this. [Robert Colman, Australia]	Editorial -- sentence rewritten.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1430	9	67	43	67	45	Nice balanced discussion about Roe and Baker. In Sexton et al (2011b) we actually narrow climate sensitivity range relative to our previous study, Murphy et al (2004). Sanderson (in review) also narrows his constraint in his submitted work. So they both contradict Roe and Baker too and maybe should be added. [David Sexton, UK]	Accepted -- Reference added.
9-1431	9	67	44	67	44	artefacts ==> artifacts [Daehyun Kim, USA]	Editorial -- corrected as suggested
9-1432	9	67	44			“are” missing before “artifacts” [Andreas Sterl, Netherlands]	Editorial -- corrected as suggested
9-1433	9	67	47	67	49	The papers cited here show that cloud entrainment parameters are those with the largest direct impact on climate sensitivity. That does not make them the most important source of uncertainty - errors of formulation and truncation are certainly larger. [Robert Pincus, USA]	Accepted -- Sentence changed so that parametric uncertainty is placed in context relative to structural uncertainty, etc.
9-1434	9	67	47	67	57	According to this para, equilibrium climate sensitivity is determined from experiments with instantaneously doubling of CO2 in a mode with a slab ocean. However, the paragraph covering lines 32-47 on page 9-64 explains a different method, involving the quadrupling of CO2 an no slab ocean. The two paragraphs should be reconciled. [Andreas Sterl, Netherlands]	Taken into account -- The diversity of methods for estimating climate sensitivity discussed at beginning of 9.7.4 is not in conflict with this discussion.
9-1435	9	67	49			Suggest adding "Covey et al. 2012 identify an convective adjustment time scale parameter and entrainment as leading parameters in the contribution to uncertainty while Neelin et al. 2010 find a parameter governing convective onset not only to be a leading source of uncertainty but to have a highly nonlinear parameter dependence of a form likely to produce bias in multi-model ensemble means." [J. David Neelin, United States]	Taken into account -- Relevant paper is not Covey et al 2012 (which deals with uncertainty in mean climate) but Lucas et al 2012 (which deals with clmate sensitivity), and the latter paper will not be submitted in time for citation in the AR5 report according to the authors.
9-1436	9	67	50	67	50	Apparently an inherent feature...'. This is a misunderstanding of these papers, in my opinion. The 'rapid response' of results reveal not that the there are issues from the experiments, but rather that according to the definition of 'feedback' (i.e. radiation changes related to global surface temperature perturbations) these adjustments from clouds are not feedbacks, and therefore should instead be understood as part of the forcing. This is analogous to cloud adjustments consifered part of the aerosol forcing, and would equally well be the case under less idealised experiments (e.g. from an RCP experiment), just that they cannot be readily diagnosed from such an experiment. [Robert Colman, Australia]	Accepted -- Reference to differences between radiative and adjusted forcing discussed in chapter 8 added.
9-1437	9	67	53	67	57	A cross-reference to chapter 7 could be useful here. [Olivier Boucher, France]	Accepted -- Reference to Chapter 7 added.
9-1438	9	68	4	68	15	Estimates of the radiative feedback alone using HadCRUT3 global temperature anomalies, assuming a feedback parameter (λ) of 2 Watts per sq. meter per deg, do not match data. The radiative variations CERES measures look nothing like what the radiative feedback should look like. You can put in any feedback parameter you want (the IPCC models range from 0.91 to 1.87), and you will come to the same conclusion. [James Wanliss, USA]	Noted
9-1439	9	68	7	68	7	Please replace "(Edwards et al., 2007) summarized.." with "Edwards et al., 2007 summarized…" [HASIBUR RAHAMAN, India]	Editorial -- corrected as suggested
9-1440	9	68	8	68	8	Please replace "….done by (Annan and Hargreaves, 2006)." with "….done by Annan and Hargreaves, 2006." [HASIBUR RAHAMAN, India]	Editorial -- corrected as suggested
9-1441	9	68	9		10	While I do expect this to be evident for any single model ensemble, and also for a large enough PMIP ensemble, this is not a general result, as it was within a single model ensemble (MIROC3.2). Perhaps the "In ensemble simulations with the MIROC model coupled to a slab ocean model" could preface this sentence rather than the one later in the paragraph? [Julia Hargreaves, Japan]	Accepted -- Preface to sentence added.
9-1442	9	68	11	68	11	Reference to Figure 9.33 - is this correct? [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Accepted -- Figure reference removed.
9-1443	9	68	13	68	14	"the LGM cooling and the warming induced by a doubling of CO2 are not symmetrical": why they must be symmetrical? Usually warming and cooling during ice ages are not symmetrical. What I am missing to understand? [Claudio Cassardo, Italy]	Taken into account -- Under assumptiions of climate linearity to small perturbations, symmetrical but opposite-signed forcing should induce symmetrical but

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							opposite-signed global surface temperature response.Language has been clarified to reflect departure from linear response to forcing.
9-1444	9	68	17			Figure 9.42: in this comparison: which is model and which is observation? not obvious [Bruce Wielicki, USA]	Taken into account - Figure 9.42 has been suppressed.
9-1445	9	68	24	68	25	The sentence would be easier to read if the references would be put at the end of the sentence. [Farahnaz Khosrawi, Sweden]	Rejected -- The sentence discusses a sequence of data sets with different authors and their rendering in a graph -- moving references to the end would make it more difficult to associate specific data with the source papers.
9-1446	9	68	28	68	28	"is that model tend ": which model? MIROC? [Claudio Cassardo, Italy]	Accepted -- The set of models has been clarified to the collection of CMIP3 models.
9-1447	9	68	28	68	28	"models" [Graham Feingold, United States of America]	Editorial -- corrected as suggested
9-1448	9	68	31	68	31	use' should be 'used'. [Zhaomin Wang, UK]	Editorial -- corrected as suggested
9-1449	9	68	37	68	37	What is meant here? Please clarify. [Farahnaz Khosrawi, Sweden]	Noted -- in the context of the previous sentences regarding the differential between land and ocean cooling, the meaning of this sentence does not require further explication.
9-1450	9	68	39	68	43	Thus which is the overall conclusion of section 9.7? [Claudio Cassardo, Italy]	Accepted -- Section restructured to clarify conclusions and summary statement added.
9-1451	9	68	49	72	25	The only comparisons worth making are the comparisons with a range of future climate behaviour. So far they have always been less successful than conventional weather forecasts. [VINCENT GRAY, NEW ZEALAND]	Rejected - Comparisons against observations of the recent past are an essential part of model evaluation, because otherwise one would need to wait for decades before any assessment can be made. The past 20+ years of observations have, however, been compared to the range of projections from the previous IPCC assessments in Chapter 1. Verifying weather forecasts is conceptually much simpler, because thousands of case studies are available.
9-1452	9	68	49			A lot in the text is based on CMIP3 models and should be referenced here also. [Gunilla Svensson, Sweden]	For the FOD, CMIP5 results were not available. This has been changed for the SOD.
9-1453	9	68	50	68	54	Here the reference to Karpechko et al. (2010) is missing. I see no reason to leave it out. This study was one of the first to assess model's ability to simulate trends in addition to ability to simulate mean state. Missing reference: Karpechko, A. Yu., Gillett, N. P., Hassler, B., Rosenlof, K. H., and Rozanov, E.: Quantitative assessment of Southern Hemisphere ozone in chemistry-climate model simulations, Atmos. Chem. Phys., 10, 1385-1400, 2010 [Alexey Karpechko, Finland]	The text has been revised and no specific references are listed, since this cannot be a complete list as it is a summary of the chapter. Therefore, we couldn't take into account this comment here.
9-1454	9	69	2	69	2	What is the abbreviation ECVs standing for? [Farahnaz Khosrawi, Sweden]	Taken into account. This abbreviation has been removed.
9-1455	9	69	4	69	4	"Figure 9.8.1" --> "Figure 9.43" (?). [Claudio Cassardo, Italy]	Editorial.
9-1456	9	69	4	69	4	Figure 9.8.1? I guess the numbers for the section and the figure got mixed up.I guess you mean figure 9.3. [Farahnaz Khosrawi, Sweden]	Editorial.
9-1457	9	69	4	69	4	Figure should be labelled as Figure 9.43 instead 9.8.1 [Irina Mahlstein, Switzerland]	Editorial.
9-1458	9	69	4	69	4	Reference to Figure 9.8.1 is incorrect, should be 9.43 [Gill Martin, United Kingdom of Great Britain & Northern	Editorial.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Ireland]	
9-1459	9	69	4	69	4	There is no Figure 9.8.1 quoted in the text. [HASIBUR RAHAMAN, India]	Editorial.
9-1460	9	69	4	69	4	Replace "Figure 9.8.1" with "Figure 9.43" [Celeste Saulo, Argentina]	Editorial.
9-1461	9	69	4	69	23	change "9.8.1" with "9.43" [ANNALISA CHERCHI, Italy]	Editorial.
9-1462	9	69	4			Figure 9.43? [Larry Thomason, United States of America]	Editorial.
9-1463	9	69	10	69	11	" The level of agreement is high … if there is only one study": conceptually, with only one study, is impossible to have agreements: by definition, there is agreement between two things, not only one. [Claudio Cassardo, Italy]	Taken into account. Text revised. In the revised version, only the confidence is shown.
9-1464	9	69	16	69	16	The abbreviation SAT is not really useful. [Farahnaz Khosrawi, Sweden]	Taken into account. Abbreviation changed.
9-1465	9	69	17	69	18	skip "so is placed in the upper right corned of the figure". [Farahnaz Khosrawi, Sweden]	Accepted.
9-1466	9	69	22	69	22	same as comment #63 [Celeste Saulo, Argentina]	Editorial.
9-1467	9	69	22			typo in number of Figure 9.8.1 [Ramon de Elia, Canada]	Editorial.
9-1468	9	69	26			I found this figure a bit hard to interpret, since it really is trying to show variables in three dimensions. I would advocate condensing the 'evidence' and 'agreement' dimensions into a single 'confidence' dimension, and then using the y dimension to show how well the simulations perform 'badly -> mixed -> well'. The 'evidence' and 'confidence' for each variable could be listed in the caption. The figure also has variablility, climatology and trends all on the same figure. At the risk of reintroducing complexity, these could be colour coded - but at least this colour would not be a dimension on the figure as it is now. [Nathan Gillett, Canada]	Taken into account. This is a good suggestion and we have revised the figure accordingly.
9-1469	9	69	42			NAO should stand for North Atlantic Oscillation and NAM for Northern Annual Mode (Section 9.5.3.7) [Christof Appenzeller, Switzerland]	Accepted. Changed as suggested.
9-1470	9	70	4	70	4	include "(DA)" after "attribution". To use the acronym in the subsequent paragraphs [ANNALISA CHERCHI, Italy]	Noted. The section has been revised
9-1471	9	70	7			“Biases in magnitude or forcing fingerprint are less important” is a strong general statement. [Robert Pincus, USA]	Taken into account. The section has been revised
9-1472	9	70	12	70	12	spell out D&A and DA [Graham Feingold, United States of America]	accepted. See comment 9-1470
9-1473	9	70	12	70	12	D&A? [Farahnaz Khosrawi, Sweden]	Editorial. See 9-1470
9-1474	9	70	12			D&A what, the some for DA in line 20, please use the same acronym and add a introduction [Frank Kreienkamp, Germany]	Editorial. See 9-1470
9-1475	9	70	15	70	15	Seems that surface is double in this sentence. Skip the first one. [Farahnaz Khosrawi, Sweden]	editorial. It should read surface variables such as
9-1476	9	70	15	70	15	8th word on this line should be "such" [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	editorial. Done
9-1477	9	70	19	70	19	"cautions that care…." : Please rephrase. [Farahnaz Khosrawi, Sweden]	taken into account. The paragraph has been updated
9-1478	9	70	20	70	20	DA --> D&A (?) [Claudio Cassardo, Italy]	editorial. See 1470
9-1479	9	70	20	70	20	DA? [Farahnaz Khosrawi, Sweden]	editorial. Same as 1478
9-1480	9	70	22			A citation should support this assertion. [Robert Pincus, USA]	taken into account. The section has been revised
9-1481	9	70	25	70	25	Please clarify: do you mean extreme precipitation events? [Farahnaz Khosrawi, Sweden]	Taken into account. The section has been revised

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1482	9	70	25	70	37	This is needlessly detailed. [Robert Pincus, USA]	Taken into account. The section has been revised
9-1483	9	70	27			I think this section is novel and well-written, and it fits in well. There is some overlap with chapter 12, so the relevant section of that chapter should be cross-referenced. [Nathan Gillett, Canada]	Taken into account. Cross-refence added.
9-1484	9	70	29	70	38	This is a good paragraph. But it should also refer to the uncertainties that derive from this assessment. Specifically, it should be added that the model range only provides a lower bound of the full uncertainty (see also the previous comment). [Gregor Betz, Germany]	Taken into account. Text revised.
9-1485	9	70	30	70	32	The statement "Certainly the ability to realistically simulate the response to historical changes in climate forcing (between contemporary and paleo, or transient changes over the 20th century) provides some reassurance that projections of future change are credible" is a fairly weak statement of confidence ("some reassurance", "credible") but even this statement is undermined by the correlations noted between forcing and sensitivity. Certainly a more objective assessment of the models would be to have them all to the same experiment; i.e., report the response to the same forcing. [Stephen E Schwartz, USA]	Taken into account. The sentence quoted was indeed weak and has been revised. Concerning response to the same forcing: As far as GHG concentrations are concerned, this has been done im CMIP5.
9-1486	9	70	33	70	35	The logic behind this statement is that simulations in closer agreement with observations are likely to arise from more faithful models. The sentiment is common but should be hedged, since it is the entire simulation (including forcings and boundary conditions) and not just the model that is evaluated. [Robert Pincus, USA]	Taken into account. Text revised.
9-1487	9	70	34	70	34	The ability to reproduce processes is key to our confidence in the ability of GCMs to simulate climate change as it's assessing the physical basis of the model. This sentence should therefore be made stronger and I suggest inserting "significantly" after "interrelationships". [Keith Williams, United Kingdom of Great Britain & Northern Ireland]	Taken into account. The role of processes is highlighted in the revised version: "Ultimately, our confidence in model projections is built upon the demonstration of how well models represent a wide range of processes on various spatial and temporal scales."
9-1488	9	70	40	70	48	Concerning weighted model means, you should 1) refer to the problems of weighting mentioned above (e.g. sensistive to metric chosen), you should 2) make explicit that one looses all information about the uncertainty we face and that weighted means tend to overstate our understanding, and you should 3) repeat the cautious stance adopted in the IPCC Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections. [Gregor Betz, Germany]	Rejected. One does not loose all information about the information about the uncertainty if model projections are weighted. However, the text has been revised and caveats of weighting are discussed.
9-1489	9	70	40	70	48	Mention that the models in CMIP5 include a variety of different levels of complexity, but have been lumped together for much of this chapter. This will add to the inhomogeneity of the results. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account by adding the following: "These models differ in complexity as discussed in Section 9.1, contributing to the inhomogeneity of the results."
9-1490	9	70	45			It is worth noting here that model weighting carries risks, too, in that weighting by measure that are not related to the model's true skill can make forecasts much worse (doi:http://dx.doi.org/10.1175/2010JCLI3594.1) [Robert Pincus, USA]	Taken into account. The revised version discusses the risks of model weighting in more detail.
9-1491	9	70	52			A study that uses objective evaluation to produce weights that are then used within projections is I. G. Watterson and P. H. Whetton (2011) Distributions of decadal means of temperature and precipitation change under global warming. JGR, 116, D07101, doi10.1029/2010JD014502 [Ian Watterson, Australia]	Taken into account. Reference added.
9-1492	9	71	2			You might also refer to Section 12.2.2, where the critical assumptions of Bayesian methods are mentioned. Ideally, these assumptions (unwarrented prior probabilities) would also be discussed here. [Gregor Betz, Germany]	Taken into account. Cross-refence added.
9-1493	9	71	2			At end of line add a sentence "Perturbed physics ensembles have shown that multi-model ensemble means for variables such as precipitation tend to underestimate amplitude due to the nonlinearity of the parameter dependence (Neelin et al. 2010). [Alternately, this could be used on page 9-64, line 1] [J. David Neelin, United States]	Taken into account. More discussion on PPE has been added.
9-1494	9	71	10	71	10	C4MIP? Is the 4 in this abbreviation correct as superscript? [Farahnaz Khosrawi, Sweden]	Taken into account. The acronym C4MIP is defined in the chapter: Coupled Climate Carbon Cycle Model Intercomparison Project

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1495	9	71	16	71	16	How closely do these "emergent constraints" that are derived from models actually reflect real constraints in the physical worl? How sensitive are they to model physics/parameterizations? Are there scale-dependent aspects to them? [Graham Feingold, United States of America]	Taken into account. The text on emergent constraints has been extended to address these issues.
9-1496	9	71	16	71	37	The right panel of Figure 9.44 is described in detail, but the left panel not at all? [Farahnaz Khosrawi, Sweden]	Taken into account. Left panel explained.
9-1497	9	71	16	71	37	I like the chapter's focus on process-oriented and regime-oriented metrics such as Bony and Dufresne and Williams and Webb, as these are metrics have a relatively low risk of getting a good score for the wrong reasons i.e. due to some fortuitous compensation of errors (as explained nicely at top of p60 in this chapter) – a lot of thought has gone into these metrics using physical understanding and so any constraint on them is promising. But we have to present the success of “emergent constraints” like Hall and Qu and Boe et al with some caution because they are currently based on a set of climate models that could all have a common structural error. It will be interesting to see where CMIP5 lies on Fig 9.44 but again even if the match with CMIP3 is good, it is no guarantee that a new relationship might emerge in some subsequent generation of climate models when some structural error has been improved. So please can a cautionary sentence be added to this paragraph explaining that there is the possibility that these emergent constraints arise due to some common structural error across the models. This is relevant to my comments about “discrepancy” and the importance of accounting for structural errors in the models, in section 9.2.2.7. [David Sexton, UK]	Taken into account. Yes, the emergent constraint technique does have limitations, especially if the models share a common bias, or if the mechanism underpinning the relationship between current and future climate parameters is not credible. In the case of snow albedo feedback, there are clear physical reasons for the relationship between the seasonal cycle and climate change -- namely that the surface albedo in snow-covered areas varies widely across the models (from 10% to 50%). The models can't all be right with regard to this number. If the seasonal cycle snow albedo feedback biases were reduced by eliminating these model biases, this would be an improvement in overall ensemble quality because the spread reduction would occur for reasons that make good physical sense. Of course, we agree with the reviewer that this process, once complete, would not guarantee the models are completely free of errors in snow albedo feedback, especially if the models share some common structural bias. It would only ensure errors made obvious by analyses such as that shown in the left panel of Figure 9.44 have been eliminated. We have included text further emphasizing the limitations of the emergent constraint technique, as the reviewer requested.
9-1498	9	71	16	72	9	The following papers propoased and applied another method to find metrics well relating to variations of climate projections and impact assessments: Shiogama, H., S. Emori, N. Hanasaki, M. Abe, Y. Masutomi, K. Takahashi, and T. Nozawa (2011), Observational constraints indicate risk of drying in the Amazon basin, Nat Commun, 2, 253. Abe, M., H. Shiogama, T. Nozawa, and S. Emori (2011), Estimation of future surface temperature changes constrained using the future-present correlated modes in inter-model variability of CMIP3 multimodel simulations, J. Geophys. Res.-Atmos., 116. Yoshimori, M. and A. Abe-Ouchi (2011): Sources of spread in multi-model projections of the Greenland ice-sheet surface mass balance. J. Climate, doi: 10.1175/2011JCLI4011.1 [Hideo Shiogama, Japan]	Taken into account. References added in Section 9.8.
9-1499	9	71	19	71	19	Could it be that you ment to cite this paper: Mahlstein I. and R. Knutti (2011). Ocean heat transport as a cause for model uncertainty in projected Arctic warming. J. Climate, 24, 1451-1460 instead of the 2010 Climate Dynamics? [Irina Mahlstein, Switzerland]	Taken into account. Reference removed.
9-1500	9	71	31	71	31	What is NINO3? [Farahnaz Khosrawi, Sweden]	Taken into account - removed.
9-1501	9	72	1	72	1	Is the 2C referred to a global mean or an Arctic mean? [Martin Juckes, UK]	Taken into account - clarified. Refers to annual mean global surface temperature

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1502	9	72	13			Annan and Hargreaves 2011 did not discuss the "effective number of independent models", a concept which seems to be rather vague, incoherent and inconsistent across the rest of the literature you cite here. I suggest it would be better to omit this paragraph altogether at this time, since this concept does not appear to be usefully quantified anywhere in the literature. Note that Masson and Knutti provided no quantitative analysis of the number of independent models. At a bare minimum, surely you must mention that this concept is not clearly defined or understood. [James Annan, Japan]	Taken into account. The paragraph has not been omitted, but the text on the effective number of independent models has been revised.
9-1503	9	72	14	72	15	Rather than saying "they are considerd to be able to provide a scientifically sound preview of the climate to come", you should be more specific and say, in line with this chapter's assessment: "jointly, climate models describe possible evolutions of the climate system, while paying tribute to the uncertainties we face". This statement would make clear that the purpose of forecasting climate change is not so much about getting the single correct answer, but about accurately gauging the uncertainties that lie ahead. [Gregor Betz, Germany]	Taken into account. This paragraph has been removed.
9-1504	9	72	14			"is much smaller..." It would pay to be explicit here: much smaller than the number of participating models. [Robert Pincus, USA]	Taken into account. Text revised.
9-1505	9	72	14			A citation to doi:10.1029/2004GL022241 would be appropriate here [Robert Pincus, USA]	Taken into account. Reference added.
9-1506	9	72	19	72	20	I don't think it is helpful to end with this paragraph, i.e. with an open-ended statement that models suffer from a multitude of systematic errors and that we don't know what this means. The chapter needs to make a clearer statement as to how useful models are for projection (and attribution), and for what scales, and for what variables. For example their representation of temperatures, and many of the associated processes leads to confidence in temperature projections (including characteristics of extremes) particularly at large scales, and higher confidence than than, say for rainfall. [Robert Colman, Australia]	Taken into account. We agree that it was not appropriate to end the chapter with an open-ended pargraph like this. The text has been revised accordingly.
9-1507	9	72	19	72	20	Such a sentence, although honest, casts doubts about any reliable predictability of the models. [François GERVAIS, France]	Taken into account. We agree that it was not appropriate to end the chapter with an open-ended pargraph like this. The text has been revised accordingly.
9-1508	9	72	19	72	25	"All models suffer from a multitude of systematic errors" – you need to take a step back and look at what has been achieved. This is an incredibly complex system, with chaotic motions on many different spatial and temporal scales. Much of the system is poorly observed. And yet the models all track the evolution of global mean temperature over the last century. Take a look at the temporal evolution of SST in the MIROC4h model – the detailed representation of ocean currents is amazing. Sure, there are systematic errors if you look at the details, and it is important the scientists maintain a strong focus on these errors – but if you are trying to review the state of the science, this paragraph is seriously misleading. [Martin Juckes, UK]	Taken into account. We agree that it was not appropriate to end the chapter with an open-ended pargraph like this. The text has been revised accordingly.
9-1509	9	72	20			unkown to unknown [Larry Thomason, United States of America]	Taken into account. We agree that it was not appropriate to end the chapter with an open-ended pargraph like this. The text has been revised accordingly.
9-1510	9	72	21	72	23	It is stated that confidence is based on how well models represent important feedbacks, but no conclusion is drawn as to what this confidence is, or what conclusions can be validly drawn from it. Yet, importantly, we do now have very high levels of confidence in water vapour/lapse rate feedback, for example, and in addition there is relatively good agreement between models, and model errors in the control climate (upper level humidity) have little impact on their feedback strength (John and Soden). We have bolstering observational evidence is well. This then gives us very high levels of confidence in the overall magnitude of the global temperature change projected over the coming century for example, albeit with a range of uncertainty mainly due to cloud responses and feedbacks. I think this chapter needs to state this somewhere in the conclusions. [Robert Colman, Australia]	Taken into account. We agree that it was not appropriate to end the chapter with an open-ended pargraph like this. The text has been revised accordingly.
9-1511	9	72	28	73	29	FAQ 9.1: I suggest addition of a brief italicised "high level" answer at the beginning of this FAQ, which covers both whether climate models are getting better, and how this is known. This would be in line with the standard WG1 FAQ style). [David Wratt, New Zealand]	taken into account -- FAQ revised

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1512	9	72	30	72	30	The real question is are they any good at all? Unless they can doi better than conventional weather forecasts the answer is no.. [VINCENT GRAY, NEW ZEALAND]	rejected -- the question has to do with climate simulation, not weather forecasting, which are different issues.
9-1513	9	72	30	73	7	".......... So, yes" The answer to the question comes rather late in this text, after a long sequence of lame excuses for perceived shortcomings. Yes, we know climate models are getting better! Note that the question asked is about improvements in the models – in the context of this report it is probably important to explain that substantial improvements in the models abilities to represent the myriad processes that influence climate my only yield minor reductions in policy relevant uncertainties, or possibly only minor improvements in the confidence with which uncertainties can be stated. The question posed, is not, however, about reductions in policy relevant uncertainties, so it should be answered with a clear affirmative. [Martin Juckes, UK]	taken into account -- FAQ revised
9-1514	9	72	32	72	32	Climate models are much more than "pieces of software." They are encapsulations of what we understand about how the climate system works. [David Randall, USA]	noted
9-1515	9	72	33	72	33	delete marvellously [Robert Colman, Australia]	accepted
9-1516	9	72	33			I would remove "marvellously" as although we may feel this way about it, others may not. [Adam Scaife, United Kingdom of Great Britain & Northern Ireland]	accepted
9-1517	9	72	35			"Vastly better" I guess it refers to the phenomena attempted to be described by the models, a great improvement in the code (or in an a priori evaluation, as mentioned by McWilliams JC 2007 Proc Natl Acad Sci 104:8709–8713). It seems to me that "vastly better" in a a posteriori sense is exagerated. [Ramon de Elia, Canada]	noted - but compared to models of 1990, ESMs are indeed 'vastly better'
9-1518	9	72	36	72	36	Add: this new generation of models include more processes, and also considers more subsystems. [Claudio Cassardo, Italy]	taken into account -- FAQ revised
9-1519	9	72	36			"Much higher spatial resolution" it seems exagerated too. [Ramon de Elia, Canada]	taken into account -- FAQ revised
9-1520	9	72	37	72	37	"(i.e., they resolve much finer-scale detail)" It's not just details, it's emergent processes that coarse resolution models do not capture at all. [James Christian, Canada]	noted -- thanks
9-1521	9	72	39	72	39	delete 'generally'. Better without it and the text does not claim it is universal. [Robert Colman, Australia]	taken into account -- FAQ revised
9-1522	9	72	40	72	42	Perhaps rephrase this sentence emphasizing that model "improvement" in terms of module extensions does not in principle lead to a reduction of uncertainty, even though the become "better". I'm not sure if a temporary degration gives the proper impression what model development means in terms of improvement and uncertainty. [Bram (Abraham) Bregman, Netherlands]	taken into account -- FAQ revised
9-1523	9	72	40	72	42	What can happen when a change is made in a single model should not be mixed up with a question about the progress in a collection of dozens of models. Please clarify that "overall degradation" here refers to a single model, not the multi-model mean (or, if you do believe it applies to the multi-model mean, provide some evidence). [Martin Juckes, UK]	taken into account -- FAQ revised
9-1524	9	72	40			I do appreciate this line and I agree that the issue is important. But the phrase "perhaps only temporarely" puts into question the very essence of climate modelling --if it is not "only temporarely" we are wasting our time. Is it worth including this into a FAQ if no appropriate discussion of the issue is carried out elsewhere? [Ramon de Elia, Canada]	rejected -- unclear what comment means, and FAQ does build on material presented in chapter
9-1525	9	72	41	72	41	"new sources of errors": what are these new sources of error? Adding an approximate representation of some process which was previously unrepresented simply results in the errors associated with that process having a different form – but it is misleading to call them "new". [Martin Juckes, UK]	taken into account -- FAQ revised
9-1526	9	72	44	72	45	The standard WG1 FAQ style is for FAQs to be readable "stand-alone" and for them not to refer to chapters or sections. I suggest rewording of this sentence to something like: Chapters assessing model performance have occurred in all of the IPCC Working Group I reports, beginning with the First Assessment and including the Fifth Assessment. [David Wratt, New Zealand]	taken into account -- FAQ revised

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1527	9	72	47			I suggest deletion of the word "either" since no "or" phrase is provided. [David Wratt, New Zealand]	accepted
9-1528	9	72	52	72	52	remove "(" before "Gleckler" [Hai Lin, Canada]	editorial
9-1529	9	72	52			The standard style for WG1 FAQs does not include references. I suggest removal of the references to Glecker et al and to Murphy et al, - and if necessary some expansion of the text to explain key concepts from these references. [David Wratt, New Zealand]	accepted
9-1530	9	72	55	72	56	The standard style for WG1 FAQs does not include references. Is it possible to reword this sentence removing the reference to Reichler and Kim ? [David Wratt, New Zealand]	accepted
9-1531	9	72				FAQ 9.1: FAQ is currently incomplete regarding reasons for improvements in modeling. The focus is currently on metrics. We suggest the discussion could be considerably broadened by considering improved process understanding, observational constraints, process couplings, multi-model ensembles etc.. [Thomas Stocker/ WGI TSU, Switzerland]	taken into account -- FAQ revised
9-1532	9	72				FAQ 9.1, Fig 1: Note that RMSE will be not generally understood by an FAQ reader, so term will need to be explained in clear language. [Thomas Stocker/ WGI TSU, Switzerland]	accepted -- caption rewritten
9-1533	9	73	2	73	2	Include "results" or "intercomparisons" after "Their" [Farahnaz Khosrawi, Sweden]	accepted -- text revised
9-1534	9	73	4	73	4	sentence incomplete [ANNALISA CHERCHI, Italy]	accepted -- text revised
9-1535	9	73	5			Suggest adding a caveat regarding aspects of the simulation of observed climate that are not converging quickly as a function of model development efforts. [J. David Neelin, United States]	taken into account -- FAQ revised
9-1536	9	73	7	73	7	"So yes": Please rephrase and write e.g. "Indeed" or "In fact". [Farahnaz Khosrawi, Sweden]	editorial
9-1537	9	73	7	73	7	"So, yes, climate models are getting better, and we can demonstrate this with quantitative performance metrics" Possibly this can be expressed more formally? I think in here, in FAQ 9.1, the 2 imbedded questions are not well separated. Question 1: are climate models getting better? In principle, this question can be answered - assuming that it is crystal clear how to assess the models. In practice, it is answered on a balance of evidence, including (still, I believe - or not?) expert judgment. Question 2: How Would We Know? This question addresses how is the model assessment done. Hence I ask: is objective of Chapter 9, also to assess the level of maturity and success of the calculation of the performance metrics? Last, general comment: Chapter 9 concern 4 classes of models, what is it referred to here? [Elisa Manzini, Germany]	noted -- FAQ must stand alone, and be readable be a general audience. so it is distinct from the Chapter 9 main text. There both methods and results are discussed.
9-1538	9	73	10	73	10	"will be well simulated" instead of "be well simulated" [Farahnaz Khosrawi, Sweden]	rejected -grammatically incorrect
9-1539	9	73	10	73	10	Include "be", so that it reads it "may not be sufficient". [Farahnaz Khosrawi, Sweden]	editorial
9-1540	9	73	10	73	10	insert 'be' between 'not' and sufficient'. [Zhaomin Wang, UK]	editorial
9-1541	9	73	22	73	27	FAQ 9.1 Figure 1: Can the figure caption be reworded to make it more understandable to a general reader, who might not understand technical terms like RMSE ? [David Wratt, New Zealand]	accepted -- caption rewritten
9-1542	9	74	1	109	20	There is a large number of references which are directly related to the WMO Scientific Assessment of Ozone Depletion (2011) and SPARC CCMVal, but in most cases the cited papers refer only to results derived from Chemistry-Climate Models where the focus is laid on the stratosphere only. I think that the link between these activities and AR5 is almost weak so far. More details must be provided about the role of CCMs regarding conclusions in AR5; maybe it is given in another chapter?! [Martin Dameris, Germany]	Taken into account. The FOD paragraph on stratospheric ozone was a placeholder since ozone from CMIP5 models was not available at the time of the FOD. This is updated in the SOD with results from CMIP5 models with interactive stratospheric chemistry. Moreover, the SOD will only assess the models that are being used in later chapters; these are from CMIP5, not CCMVal.
9-1543	9	74	1	109	20	Cheng, A., and K.-M. Xu, 2011: Improved low-cloud simulation from a multiscale modeling framework with a third-order turbulence closure in its cloud-resolving model component. J. Geophys. Res., 116, D14101, doi:10.1029/2010JD015362.	noted

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the Earth's top-of-atmosphere radiation budget. J. Climate, 22, 748–766. doi: 10.1175/2008JCLI2637.1 [Kuan-Man Xu, USA]	
9-1544	9	74	1			References to add from suggestions in text: [J. David Neelin, United States]	noted
9-1545	9	74	1			Neelin, J. D., A. Bracco, H. Luo, J. C. McWilliams, and J. E. Meyerson, 2010: Considerations for parameter optimization and sensitivity in climate models. Proc. Nat. Acd. Sci., 107(47), doi:10.1073/pnas.1015473107 [J. David Neelin, United States]	noted
9-1546	9	74	1			Neelin, J. D., O. Peters, and K. Hales, 2009: The transition to strong convection. J. Atmos. Sci., 66, 2367-2384.DOI: 10.1175/2009JAS2962. [J. David Neelin, United States]	noted
9-1547	9	74	1			Sahany, S., J. D. Neelin, K. Hales, and R. Neale, 2012: Temperature-moisture dependence of the deep convective transition as a constraint on entrainment in climate models. J. Atmos. Sci., in Press. [J. David Neelin, United States]	noted
9-1548	9	74	1			Holloway, C. E. and J. D. Neelin 2009: Moisture vertical structure, column water vapor, and tropical deep convection. J. Atmos. Sci., 66, 1665-1683. DOI: 10.1175/2008JAS2806.1 [J. David Neelin, United States]	noted
9-1549	9	74	1			Neelin, J. D., O. Peters, J. W.-B. Lin, K. Hales and C. E. Holloway, 2008: Rethinking convective quasi-equilibrium: observational constraints for stochastic convective schemes in climate models Phil. Trans. Roy. Soc. Lond. A, 366, 2581-2604. doi:10.1098/rsta.2008.0056. [J. David Neelin, United States]	noted
9-1550	9	74	1			Neelin, J. D., 2007: Moist dynamics of tropical convection zones in monsoons, teleconnections and global warming. In The Global Circulation of the Atmosphere, T. Schneider and A. Sobel, Eds, Princeton University Press, Princeton, 385pp. [J. David Neelin, United States]	noted
9-1551	9	74	1			Hall, A., X. Qu, and J. D. Neelin, 2008: Improving predictions of summer climate change in the United States Geophys. Res. Lett., 35, L01702, doi:10.1029/2007GL032012. [J. David Neelin, United States]	noted
9-1552	9	74	1			Neelin, J. D., M. Munnich, H. Su, J. E. Meyerson, and C. E. Holloway, 2006: Tropical drying trends in global warming models and observations. Proc. Nat. Acd. Sci., 103, 6110-6115. [J. David Neelin, United States]	noted
9-1553	9	74	1			Wallace, J. M., E. M. Rasmusson, T. P. Mitchell, V. E. Kousky, E. S. Sarachik, and H. von Storch, 1998: On the structure and evolution of ENSO-related climate variability in the tropical Pacific: lessons from TOGA. J. Geophys. Res., 103, 14241-14259 [J. David Neelin, United States]	noted
9-1554	9	74	1			Jackson, C. S., M. K. Sen, G. Huerta, Y. Deng, and K. P. Bowman, Error reduction and convergence in climate prediction, J. Climate, 21, 6698-6709, 2008, 8 citations, doi:10.1175/2008JCLI2112.1 [J. David Neelin, United States]	noted
9-1555	9	74	1			Covey et al. 2012 submitted -- should request from Curt Covey Summarizing uncertainty quantification work by the group at Lawrence Livermore national labs [J. David Neelin, United States]	noted
9-1556	9	74	1			Ming, Y and co-authors 2012: I don't have a precise reference for this but suggest contacting [J. David Neelin, United States]	noted
9-1557	9	75	50	75	51	Please use full reference:Beare, R.J., MacVean, M.K., Holtslag, A.A.M., Cuxart, J., Esau, I., Golaz, J-C., Jimenez, M.A., Khairoutdinov, M., Kosovic, B., Lewellen, D., Lund, T.S., Lundquist, J.K., McCabe, A., Moene, A.F., Noh, Y., Raasch, S. and Sullivan, P.P, 2006: An intercomparison of Large-Eddy Simulations of the stable boundary layer. Boundary-Layer Meteorol., 118, 247-272. [Albert A.M. Holtslag, Netherlands]	Editorial - corrected in online library
9-1558	9	77	21	77	25	This is the same reference twice. [Farahnaz Khosrawi, Sweden]	Editorial - Duplicate deleted from library
9-1559	9	77	54	77	55	The citation should be: Brown, J.N., A.V. Fedorov and E. Guilyardi, (2011) How well do coupled models replicate ocean energetics relevant to ENSO? Climate Dynamics, 36, 2147–2158, doi:10.1007/s00382-010-0926-8. (Jaclyn N. Brown is different person to Josephine R. Brown). [Josephine Brown, Australia]	Editorial - Reference updated in library

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1560	9	80	21	80	22	Please use full reference:Cuxart, J., A.A.M. Holtslag, R.J. Beare, E. Bazile, A. Beljaars, A. Cheng, L. Conangla, M. Ek, F. Freedman, R. Hamdi, A. Kerstein, H. Kitagawa, G. Lenderink, D. Lewellen, J. Mailhot, T. Mauritsen, V. Perov, G. Schayes, G-J. Steeneveld, G. Svensson, P. Taylor, W. Weng, S. Wunsch, and K-M. Xu, 2006: Single-column model intercomparison for a stably stratified atmospheric boundary layer. Bound.-Layer Meteor., 118, 273-303. [Albert A.M. Holtslag, Netherlands]	Editorial - Reference updated in library
9-1561	9	86	12	86	15	These two references are obviously the same! [Andreas Sterl, Netherlands]	Editorial - Cause for duplication unclear
9-1562	9	93	28	93	29	Pages are missing. [Farahnaz Khosrawi, Sweden]	Editorial - Reference updated in library
9-1563	9	93	37	93	44	Same reference is repeated twice [HASIBUR RAHAMAN, India]	Editorial - Reference updated in library, cause for duplication unclear
9-1564	9	108	3	108	3	TABLE9.1: Given the increased heterogeneity in model tops and overall vertical resolution, it would be particularly useful, to know the number of levels in two part of the atmosphere: (1) surface 100 hPa and (2) 100 hPa to model top. Can this be listed in the table? This information will help / better enable to distinguish between changes in model performance / projection associated with different vertical resolution in the troposphere and those associated with different vertical resolution in the stratosphere. Note: at present there is redundant info in the table, as well as errors (eg, MPI-ESM-xx wrong top; HadGEM2-ES twice). I also have another suggestion: obviously, there are a number of models used in slightly different configuration (eg, the series of gfdl models). To improve the readability of the table, and to reduce the chances of errors, try to use a model version as baseline, and report for the other versions only the difference from the baseline. [Elisa Manzini, Germany]	Taken into account. Indication of low top /high top will be provided for final version. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://http://q.cmip5.ceda.ac.uk/. Additional suggestion could not be taken into account due to time constraints.
9-1565	9	110				Table 9.1: The entries for many models need completion. I will send the CLI details on the ACCESS1.0 model shortly. (Details are too extensive to give here.) [Anthony Hirst, Australia]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1566	9	110				Table 9.1: Up to know it is everywhere xx% written. I hope this will be updated. [Farahnaz Khosrawi, Sweden]	Taken into account. Information removed and discussion made in text.
9-1567	9	110				Table 9.1: it would be better to have separate tables for the AOGCMS and the ESMs, and also for AMIP models. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. No clear separation can be made so it was decided to keep all models in the same table
9-1568	9	110				Table 9.1: It would be better to put high-top and low-top models in separate tables. They are not simply different resolutions, they represent different functionality. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Rejected. Depending on people's particular interests, many different organisations of the table are possible and the current choice was made by Chap 9 LAs. We nevertheless will add the low top /high top information
9-1569	9	110				Table 9.1: It would be useful to separate atmospheric chemistry and the carbon cycle in this table. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Rejected. Depending on people's particular interests, many different organisations of the table are possible and the current choice was made by Chap 9 LAs.
9-1570	9	112	3	112	3	EC-EARTH model - the institution is called Europe , but I think it is ECMWF [Claudio Cassardo, Italy]	Rejected. EC-Earth is a European consortium and ECMWF is not part of this consortium as it has no mandate to perform climate studies.
9-1571	9	112		122		Table 9.1: In gfdl-esm2g and gfdl-cm2p1 top model level is shown as 0 (zero) is it correct ? Also some specific model name is also missing like NIMO,MOM4 etc. Table 9.3 should be Table 9.2 [HASIBUR RAHAMAN, India]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1572	9	117		117		Table 9.1: HadGEM2-CC and HadGEM2-ES main references are Collins et al., 2011 (GMD 4, 1051-1075) and The HadGEM2 Development Team, 2011 (GMD 4, 723-757). [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
							http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1573	9	117		117		Table 9.1: HadGEM2-CC does have aerosols, in the same way as HadGEM2-ES. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1574	9	117				References to Collins et al. for the HadGEM2 model should be updated to Collins et al. Geosci. Mod. Devel. 2011 [William Collins, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1575	9	117				Table 9.1, entry on p.117 for HadGEM2-CC has some errors: column 2, model refs should be Martin et al., 2011, GMD and Collins et al., 2011, GMD; column 3 Bellouin reference should be updated to 2011 (as per column 5) [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1576	9	117				Table 9.1, entry on p.117 for HadGEM2-ES has some errors: column 2, model refs should be Martin et al., 2011, GMD and Collins et al., 2011, GMD; column 3 is wrong - HadGEM2-ES DOES HAVE AEROSOLS! Use the same Bellouin (2011) reference as for HadGEM2-CC. Column 4 - I don't recognise the Davies et al reference here - can you confirm? It does not appear in the reference list for the chapter [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1577	9	118				MPI-ESM-LR, MPI-ESM-MR and MPI-ESM-P are of vintage 2012 (not 2009) [Marco Giorgetta, Germany]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1578	9	118				MPI-ESM-LR, MPI-ESM-MR and MPI-ESM-P have an atmospheric grid top of 0.01 hPa (not 0.1 hPa) [Marco Giorgetta, Germany]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1579	9	118				MPI-ESM-P uses the same ocean model as MPI-ESM-LR, hence the entries in the "Ocean" column should be the same [Marco Giorgetta, Germany]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1580	9	118				MPI-ESM-MR: ocean top level is missing [Marco Giorgetta, Germany]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1581	9	118				The ocean models of MPI-ESM-LR, MPI-ESM-MR and MPI-ESM-P use the same vertical grid, but the information is different for each of the three models [Marco Giorgetta, Germany]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1582	9	121		121		Table 9.1: HadGEM2-ES appears twice in this table. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Editorial. Corrected
9-1583	9	122				Comments on entry for CSIRO-Mk3-6-0 in Table 9.2: Vintage: It is probably more accurate to write 2009, since various changes were made to the code between 2007 and 2009. Main reference(s): We have a new main reference, (Rotstayn et al., ACPD, 2012). Hopefully, we will be able to report that it is published in ACP by the time of the Second Order Draft. It is much more comprehensive than Rotstayn et al. (2010) as a reference that documents the changes that were made between Mk3.5 and Mk3.6, and also describes what was done with the forcings etc. for CMIP5. There are probably too many references listed for CSIRO-Mk3.6.0 under "main references", especially once you get down to Sato et al. (a data set that is used, but not really part of the model). I'd probably just list Rotstayn et al. (2012) in that column, subject to its acceptance by ACP. Flux corrections: nil. Aerosol references: These are OK, but I'd replace Rotstayn et al. (2010) with Rotstayn et al. (2011), which describes the dust scheme. Atmosphere: Replace Rotstayn et al. (2010) with the 2012 paper, and add the references for Gordon et al. (2002, 2010) to the reference list. Top level ~4.5 hPa. Land ice: not implemented. Ocean: Is a modified version of GFDL MOM 2.2. We don't understand what you want for "Top BC": My colleague suggests "rigid lid, salinity flux", but that appears not to be what you are looking for. References: Gordon, H. B., Rotstayn, L. D., McGregor, J. L., Dix, M. R., Kowalczyk, E. A., O'Farrell, S. P., Waterman, L. J., Hirst, A. C., Wilson, S. G., Collier, M. A., Watterson, I. G., and Elliott, T. I., 2002: The CSIRO Mk3 Climate System Model, Technical Paper No. 60, CSIRO Atmospheric Research, Aspendale, Vic., Australia, 134 pp. Available online at http://www.cmar.csiro.au/e-print/open/gordon 2002a.pdf. Gordon, H. B., O'Farrell, S. P., Collier, M. A., Dix, M. R., Rotstayn, L. D., Kowalczyk, E. A., Hirst, A. C., and Watterson, I. G., 2010: The CSIRO Mk3.5 Climate Model, Technical Report No. 21, The Centre for Australian Weather and Climate Research, Aspendale, Vic., Australia, 62 pp. Available online at http://www.cawcr.gov.au/publications/technicalreports.php. Rotstayn, L. D., Collier, M. A., Mitchell, R. M., Qin, Y., Campbell, S. K., and Dravitzki, S. M. (2011): Simulated enhancement of ENSO-related rainfall variability due to Australian dust, Atmos. Chem. Phys., 11, 6575–6592, doi:10.5194/acp-11-6575-2011. Rotstayn, L. D., S. J. Jeffrey, M. A. Collier, S. M. Dravitzki, A. C. Hirst, J. I. Syktus, and K. K. Wong (2012). Aerosol-induced changes in summer rainfall and circulation in the Australasian region: a study using single-forcing climate simulations. Atmos. Chem. Phys. Discuss. (in press). [Leon Rotstayn, Australia]	Taken into account. This table is generated automatically from the on-line CMIP5 questionnaire as filled up by the modelling groups and available from http://q.cmip5.ceda.ac.uk/. We will verify contents with the modelling groups.
9-1584	9	126	1	126	7	This figure is quite confusing, in particular in part 3. One can deduce that Land has interactions only with Biosphere, ocean and atmosphere. Actually, Land has links with Carbon cycle and Atmos. Chemistry as well [Eric Martin, France]	Accepted. Figure eliminated.
9-1585	9	126	5	126	7	Figure 9.1. is not clearly comprehensible, it needs more explanation (in text and/or below the figure) [Ladislav Metelka, Czech Republic]	Accepted. Figure eliminated.
9-1586	9	126				Figure 9.1 is colorful, but I do not see its benefits compared to the text alone. Conceptually, it is unclear why "atmospheric chemistry" must have its own hexagon as it is completely embedded in the atmosphere. Further I think it is wrong not to connect the "carbon cycle" directly to the land. Also, if the carbon cycle is shown, which is not a geographically confined model component like for example atmosphere or ocean, then I wonder why the energy cycle or the hydrological cycle are missing, which both are at the core of climate dynamics. For the carbon cycle it is also problematic to show a single set of arrows, as the related process models are embedded in separate compnents and have their own sets of theory, observation, numerical implementation and parameter adjustments. I think this sketch is not sufficient. [Marco Giorgetta, Germany]	Accepted. Figure eliminated.
9-1587	9	126				Fig. 9.1: This figure is very confusing. Is biosphere not connected to atmosphere? Is carbon cycle not connected to land? I would suggest to put biosphere across the boundary between ocean and land, put atmos chemistry within atmosphere, and put carbon cycle overlapped with atmosphere, land and ocean. The color scheme is also confusing. For example, purple is used for both "parameter adjustment" and "carbon cycle". [Jialin Lin, United States of America]	Accepted. Figure eliminated.
9-1588	9	126				Figure 9.1: more explanation is needed. It is not obvious what the difference is between ESM and AOGCM. [Irina Mahlstein, Switzerland]	Accepted. Figure eliminated.
9-1589	9	126				This figure is not very clear. [Gunilla Svensson, Sweden]	Accepted. Figure eliminated.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1590	9	126				This figure is pretty opaque to me [Larry Thomason, United States of America]	Accepted. Figure eliminated.
9-1591	9	126				Fig. 9.1 (part 3) is a bit mis-leading, because there are also connections between Biosphere and Carbon cycle, between Land and Carbon cycle, and netween Biosphere and Atmos chemistry. [Zhaomin Wang, UK]	Accepted. Figure eliminated.
9-1592	9	126				Fig 9.1: The key messages that you want to convey with this figure are not clear to us. In particular the figure concept seems hard to understand. Several key links are missing, and so is the Cryosphere component of the Earth System. The role of hexagons is not clear to us. [Thomas Stocker/ WGI TSU, Switzerland]	Accepted. Figure eliminated.
9-1593	9	127	1	127	3	I suggest to put in bold the colored text, for making it more readible in the pdf file. [Claudio Cassardo, Italy]	Editorial. The figure has been made more clear
9-1594	9	127				Is that not a figure for chapter 1? [Olivier Boucher, France]	noted. Ch09 is where CMIP5 experiment is described in detail and so figure is best placed here
9-1595	9	127				Wouldn't it be better to put this information into a table? [Farahnaz Khosrawi, Sweden]	rejected. Figure provides compact and visual representation of CMIP5 experiments
9-1596	9	127				Does this even require a figure? [Larry Thomason, United States of America]	noted. Ch09 is where CMIP5 experiment is described in detail and so figure is best placed here
9-1597	9	128	1	129	8	Figures 9.3 and 9.4: Surely there is a more up to date observational dataset for surface air temperature? It is not very convincing to readers to show model means for 1985-2005 compared against observations published in 1999 - what time period are these meant to represent? [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	Taken into account. Newer observatonal data have been used.
9-1598	9	128	5	128	6	Skip the abbreviation MME [Farahnaz Khosrawi, Sweden]	Editorial. The figure and caption have been completely revised.
9-1599	9	128				Fug 9.3: The reference periods are inconsistent netween text, figure and legend. Also, how can the observations for at least 5 years in the 21sr century come from a 1999 publication? [Peter Guttorp, USA]	Taken into account. The figure and caption have been completely revised.
9-1600	9	128				Figure 9.3 (bottom panel): it would be much easier to see the size of average errors if a more suitable colour bar was used. At the moment, the whole globe is a shade of pink and it is very difficult to diagnose the size of errors in a particular region with any accuracy. [Chris Roberts, Uk]	Taken into account. The figure and caption have been completely revised.
9-1601	9	128				It would be awfully interesting to see a few figures like this one with cmip3 vs cmip5…. [Larry Thomason, United States of America]	Noted. We have considered this, but it is not justified because there is very little difference in the CMIP3 and CMIP5 multi-model averages.
9-1602	9	128				Figs 9.3 and 9.4: we suggest to also use additional datasets available (and discussed in Chapter 2), or to at least refer back to Chapter 2 and to justify your selection of using Jones, 1999, only. [Thomas Stocker/ WGI TSU, Switzerland]	Taken into account. The figure and caption have been completely revised.
9-1603	9	128				Figs 9.3 and 9.5: We note the inconsistent use of 'model climatology' vs. 'model mean' in the titles. [Thomas Stocker/ WGI TSU, Switzerland]	Taken into account. The figure and caption have been completely revised.
9-1604	9	129	5	128	5	Skip the abbreviation MME [Farahnaz Khosrawi, Sweden]	Editorial. The figure and caption have been completely revised.
9-1605	9	129				A usual convension is to use blue for negative values and red for positive. Please change or use completely different colors. [Gunilla Svensson, Sweden]	Editorial. The figure and caption have been completely revised.
9-1606	9	130	1	130	1	In fig. 9.5, in the central panel the scale is not centered in the color interval, thus there are not blue portions in the figure. The first (wrong) visual impact from this figure is that the model is drier everywhere. I suggest to center the scale between the colors, in order that white color indicates the smallest absolute bias. Concerning the bottom panel, even in this case the choice of the scale appears misleading, as it evidences almost only errors in the tropical regions: I will suggest to use non-equal intervals in order that also errors in the rest of the world can be seen, eventually highligthing the use of a specific scale in the legenda. [Claudio Cassardo, Italy]	Editorial. The figure and caption have been completely revised.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1607	9	130	1	130	1	If possible, convert units in mm/day or in a more comprehensible unit. [Arona DIEDHIOU, France]	Editorial. The figure and caption have been completely revised.
9-1608	9	130	5	130	5	Skip the abbreviation MME [Farahnaz Khosrawi, Sweden]	Editorial. The figure and caption have been completely revised.
9-1609	9	130				fig. 9.5 – colour scale is poor here. Linear scale on difference plot only shows up errors in regions of high precip – it looks like there are virtually no precip errors over land or outside tropical pacific! Also need better units? Suggest mm/day as more meaningful than kg/m2/s [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Editorial. The figure and caption have been completely revised.
9-1610	9	130				Figure 9.5: would it help to show results in percentages? [Irina Mahlstein, Switzerland]	Taken into account. The figure and caption have been completely revised.
9-1611	9	131				Figure 9.6: Units are missing in the left panel (which is also the case for many other figures) [Irina Mahlstein, Switzerland]	Editorial. The figure and caption have been completely revised.
9-1612	9	131				Fig 9.6 right, and throughout. when giving plots as function of latitude, make the independent variable sine(latitude), not latitude. Otherwise there is great visual distortion of the effects at high latitudes. [Stephen E Schwartz, USA]	Rejected; there is no absolute standard for plotting line plots as function of latitude; will keep current format for reasons of simplicity and consistency with other chapters.
9-1613	9	131				It would seem that the differences between model and obsd in right hand col should also be presented; these differences seem rather large, up to 15 W m-2. Surely these differences deserve discussion; what are the implications; simply stating "major challenge" page 9-23, line 47, hardly seems to do justice to the matter. Page 24, lines 27-29 seem to move in that direction. [Stephen E Schwartz, USA]	Taken into account - The figure has been updated with the CMIP5 models. We have decided to stay with the absolute value presentation rather than showing differences, as it is important to provide some insight into the relative magnitude of the error, which the maps do not. The second half of the comment refers to the text, not the figure and has been taken into account in updating the text to the new CMIP5 results.
9-1614	9	132	6	132	6	I do not think there are solid evidence for "Ozone depletion increased after 1960" [Zhaomin Wang, UK]	Rejected - there is solid evidence that ozone depletion increased after 1960 as equivalent stratospheric chlorine (ESC) values steadily increased throughout the stratosphere, see WMO (2011) - reference added.
9-1615	9	132	7			please explain "equivalent stratospheric chlorine" [Martin Dameris, Germany]	Taken into account.
9-1616	9	133				Fig, 9.8: The ticks on x-axis may be removed. [Jialin Lin, United States of America]	Taken into account. The figure and caption have been completely revised.
9-1617	9	133				The green bars in this plot are misleading in combination with the dash-marks. Either represent the average by a colored dash mark, or make a box-and-whisker plot. [Andreas Sterl, Netherlands]	Taken into account. The figure and caption have been completely revised.
9-1618	9	134		134		is this a placeholder for a similar figure using CMIP5 models? If not, I suggest omitting it on the grounds that it is dated. [Philip Mote, USA]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1619	9	134				Fig. 9.9. The legend needs to explain the acronyms in the vertical legend (or give a place to look them up). Are the two first columns on the same color scale as the rest of the figure? It is a bit puzzling to have them all be negative. [Peter Guttorp, USA]	Editorial. The figure and caption have been completely revised.
9-1620	9	134				Am I suppose to see a number hiding in here somewhere? This is pretty darn indigestible (to me anyway) [Larry Thomason, United States of America]	Taken into account. The figure and caption have been completely revised.
9-1621	9	135	1	135	2	I expect this figure (9-10) or one like it will become an "icon" of this report as was the case for Fig 9.5 in AR4. However by plotting anomaly, rather than temperature itself, vs year, the figure minimizes the variation among the models. I would think that it is essential to include a plot of Global mean sfc temp itself from the models, rather than (or in addition to) anomaly.	Noted. We will add the time-mean over the reference period for each model and the observations. Note that in constructing GMST from observed station data, temporal anomalies are used; hence for GMST

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						I call attn to such a plot Tredger, E. Thesis, 2009, http://cats.lse.ac.uk/homepages/edward/TREDGER_Thesis.pdf page 71. On the evaluation of uncertainty in climate models. PhD thesis, London School of Economics, London. Also Stevens, Bjorn and Stephen E. Schwartz, 2011: Observing and Modeling Earth's Energy Flows. Surveys of Geophysics, revised January 2012. http://www.mpimet.mpg.de/fileadmin/staff/stevensbjorn/Documents/StevensSchwartz2012.pdf Figure 11. In this context I call attention to the language at page 20 line 34-36: "since the effective climate sensitivity depends on the state of the climate system, it is necessary for climate models to reproduce the observed state as accurately as possible to minimize the effects of state-related errors on projections of future climate (Senior and Mitchell, 2000)." As shown by Tredger the spread in GMST among the AR4 models was 3K, a spread that is well greater than the increase in any of the modeled GMST over the twentieth century, and well beyond the threshold of concern of increase in GMST, and certainly into the range of differences that might be expected to affect the effective sensitivity of the climate models or the actual climate. It would thus seem that, depending on the spread of the AR5 models that is revealed in a Tredger type plot, the effect of such spread on model sensitivity would need to be discussed. [Stephen E Schwartz, USA]	the anomaly is the primary variable. Note further that some models are tuned against observed GMST, some others are not (see tuning box). Hence, the full meaning of different absolute GMST in different models is complex. Thanks for pointing out text on p. 20, line 34-36. We will add a figure to the chapter showing absolute GMST over reference period (and/or) control run against model climate sensitivity. Text on p. 20 line 34-36 will be adapted.
9-1622	9	135	1	135	2	I am pleased to see specific models identified in the figure (contrast AR4, Fig 9.5, where they were not). May I recommend a supplementary material for the chapter in which the numerical data of the plots are presented so that they may be readily analyzed by the scientific community. [Stephen E Schwartz, USA]	Rejected. It is beyond the scope of this assessment to provide easy access to processed model output that is already publicly available through the CMIP5 archive.
9-1623	9	135		135		why not continue the plot to the latest year for which data are available (2010 for FOD, 2011 for SOD)? Also, the use of 1961-90 climatology artifically reduces the intermodel variance during that period and visually suggests that uncertainty is smaller then than during other periods. I suggest using a 20th century average for the calculation of anomalies, to remove this artificial visual effect. [Philip Mote, USA]	Noted. Will attempt to extend time series.
9-1624	9	135				Fig. 9.10: I strongly suggest extending this figure through 2011, using, as necessary, model RCP simulations (e.g., RCP4.5) for the periiod 2006-2011. This would enhance the credibility of the treatment here. [Anthony Hirst, Australia]	Noted. Will attempt to extend time series.
9-1625	9	135				Figure 9.10: Please improve the quality of this figure. As it is now the colors are quite difficult to distinguish. [Farahnaz Khosrawi, Sweden]	Noted. Will do.
9-1626	9	135				Lines for the specific models are too thin; labels for model key is too small [Larry Thomason, United States of America]	Noted. Will improve.
9-1627	9	136	5	136	5	Include "given for all climate models and data considered". [Farahnaz Khosrawi, Sweden]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1628	9	136	5	136	5	I don't see the connection to Figure 3a. [Farahnaz Khosrawi, Sweden]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1629	9	136	6	136	6	This comment is more confusing than helpful. [Farahnaz Khosrawi, Sweden]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1630	9	136	9	136	10	Which symbols are the climate models and which are the satellite datasets? [Farahnaz Khosrawi, Sweden]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1631	9	136				Figure 9.11: Explain in caption what W and Tct are. [Irina Mahlstein, Switzerland]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1632	9	136				Figure 9.11: Surely the caption for this is incorrect: is it not just a plot of the trend in W as a function of the TLT trend? (I think this is incorrect in the original paper too). The caption also refers to Table 1 and Figure 3a,	Taken into account. The figure and caption have been completely revised using CMIP5 data.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						which must be incorrect. [Gill Martin, United Kingdom of Great Britain & Northern Ireland]	
9-1633	9	137	4	137	4	after 'averaged' should use 'over' instead of 'from'. [Zhaomin Wang, UK]	Considered. Figure caption revised.
9-1634	9	138	4		10	The top left plot of Figure 9.13 does not, as stated in the caption, show surface air temperature change between the LGM and present day. The ocean measurements represent SST. See Figure 2 of Hargreaves et al CP, 2011 for convincing evidence of why this is important. The air and ocean SST LGM-present values are vastly different at the high latitudes due to the large increase in sea ice at the LGM! [Julia Hargreaves, Japan]	accepted. Note that the major changes were made to this figure
9-1635	9	138	8	138	8	What is MH? [Farahnaz Khosrawi, Sweden]	Editorial. Mid-Holocene
9-1636	9	138	9	138	10	"On the right figures the red line highlights the root-mean square of the inter-model differences": I cannot see clearly this red line. [Claudio Cassardo, Italy]	Taken into account. Not that these mapshave been supressed of the figure
9-1637	9	138				Plot quality on upper left panel is poor [Larry Thomason, United States of America]	noted
9-1638	9	142				labelling on key is too small [Larry Thomason, United States of America]	Taken into account. The figure and caption have been completely revised using CMIP5 data.
9-1639	9	142				Here, it should be clarified if human-contributed sulfate aerosols were included in driving the models (see comments 38). [Zhaomin Wang, UK]	Noted. This information should be available from the METAFOR documentation in time for it to be used for the final draft.
9-1640	9	143	4	143	10	The variable on the y axis is the temperature or the theta (=potential temperature)? The two variables are similar but not equal… Also, the sentence " Show quantitatively … palaeoclimate" appears useless. [Claudio Cassardo, Italy]	Taken into account. Figure and legen have been updated
9-1641	9	143	5			“parenthesis open” missing before “triangles” [Andreas Sterl, Netherlands]	editorial. Corrected
9-1642	9	143	6	143	6	Put the site numbers in brackets or skip them. As it is now it is quite confusing. [Farahnaz Khosrawi, Sweden]	editorial
9-1643	9	143	9	143	10	“Show quantitatively .... “ What does this mean? Leftover from discussion? [Andreas Sterl, Netherlands]	Taken into account. Text clarified
9-1644	9	143		143		Figure 9.18 - Is there a reason the symbols in the legend have a different shape from the symbols in the figure? Also the black symbols for paleo observations are difficult to distinguish from the model colors; suggest thickening, or using a different shape symbol. Finally, connecting the two geographic points with straight lines carries an implication that the properties of sea water in geographic space would be connected along that line. If that is not true, or not well known, then using the lines is misleading and the symbols should be shown simply as symbols without the connecting line segment. [Philip Mote, USA]	Taken into account. Figure redrawn and updated
9-1645	9	143				Figure 9.18: It is quite confusing what it is shown here. It would be worth to use also triangles in the legend and maybe add a second legend on the left side (or extend the existing one on the right side) with the second station. Another possibility would be to split the figure into two, one for each station. [Farahnaz Khosrawi, Sweden]	Taken into account. Figure redrawn and updated
9-1646	9	143				Figure 9.18: Open and filled symbols are different in plot (triangle) and legend (circle) [HASIBUR RAHAMAN, India]	Taken into account. Figure redrawn and updated
9-1647	9	144				Figure 9.19: A legend showing which color represents which color will hopefully be given in the updated version of the figure. [Farahnaz Khosrawi, Sweden]	Taken into account. Figure redrawn and updated
9-1648	9	144				Figure 9.19: In figure legend is not there for the different models [HASIBUR RAHAMAN, India]	Taken into account. Figure redrawn and updated
9-1649	9	145				Figure 9.20: Include a,b in the figure or write top and bottom in the figure caption. Which figure shows CMIP3 and which CMIP5? [Farahnaz Khosrawi, Sweden]	Taken into account. Figure redrawn and updated
9-1650	9	146	7	146	7	This reference is not in the reference list. [Farahnaz Khosrawi, Sweden]	Taken into account. Figure redrawn and updated

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1651	9	146				Figure 9.21: Also here a legend for the different models will hopefully be included in the updated version of this figure. [Farahnaz Khosrawi, Sweden]	Taken into account. Figure redrawn and updated
9-1652	9	146				Figure 9.21: No legend is there is figure [HASIBUR RAHAMAN, India]	Taken into account. Figure redrawn and updated
9-1653	9	147	1	147	2	In the figure, the two bands of CMIP3 and CMIP5, mean +/- std, are partially overlapped: that of CMIP5 covers that of CMIP3. The band of CMIP3 is this not well visible. I suggest to use a transparent color for the CMIP5 band, or for both, in order to see both bands. [Claudio Cassardo, Italy]	Accepted
9-1654	9	147	1	150	9	In Fig. 9.22, there is a generic mention of CMIP3/CMIP5 models ensemble, without an indication of how many models are used. In Fig. 9.23, 14 CMIP3/5 models are used. In Fig. 9.24, 12 CMIP3/5 models (called "the 12" CMIP* models as if there are only 12 models) are used. In Fig. 9.25, just 9 models. Could it be possible to use the same number of models everywhere? The use of a different number of models (which ones?) for each figure is questionable. [Claudio Cassardo, Italy]	Taken into account. Caption upodated. An attempt will be made to follow this suggestion. However, this is the matter of current avalability of particular types of experiments with individual models.
9-1655	9	147	5	147	10	For consistency with the rest of the report, the reference period for the mean seasonal cycle of sea ice extent as simulated by CMIP5 models should be 1985-2006. [Thierry Fichefet, Belgium]	Rejected. This figure is to show the difference between CMIP3 (as in AR4) and CMIP5. That is why the AR4 baseline is used
9-1656	9	147				Figure 9.22: Write "Sea ice extent" instead of "SI" in the figure titles [Farahnaz Khosrawi, Sweden]	Taken into account. The titles are removed.
9-1657	9	148				Figure 9.23: Why are different color bars used (different scale for the numbers)? Why do the negative numbers occur? What does it mean? [Farahnaz Khosrawi, Sweden]	Noted. The 4 figures to the left (A) show the NUMBER of CMIP5 models (which cannot be negative). The 4 figures to the right (B) show the DIFFERENCES btw the numbers of CMIP5 and CMIP3 models (which can be negative).
9-1658	9	148				Figure 9.23: state clearly whether CMIP5-CMIP3 or other way round. [Irina Mahlstein, Switzerland]	Accepted
9-1659	9	149	5	149	5	Write "green" and "blue" instead of "greenish" and "blueish" [Farahnaz Khosrawi, Sweden]	Accepted
9-1660	9	149				Figure 9.24: Could we get this for Antarctic sea ice trends as well? [Marcus Sarofim, USA]	Accepted
9-1661	9	149				Why NSIDC and not Had ISST? To be consistent with figure 9.23. [Andreas Sterl, Netherlands]	Rejected. The figure is adopted from published studies.
9-1662	9	149				It is not possible to see the extent of the grey shaded region. [Gunilla Svensson, Sweden]	Accepted
9-1663	9	149				There should be two panels, the other one for SH March Sea Ice Extent. [Zhaomin Wang, UK]	Accepted
9-1664	9	150	4	150	5	"for February": it could be convenient to remember here the CMIP5 simulation period, allowing thus a comparison with the one of (Robinson & Frei, 2000). [Claudio Cassardo, Italy]	Noted. An update of Robinson and Frei (2000) is used here.
9-1665	9	150				Figure 9.26: The left figure and bottom figure looks somewhat strange. I would suggest that you put the x-axis at -20 and -40 respectively, thus at the bottom of the y-axis and include a dotted zero line to mark the zero in the figure. [Farahnaz Khosrawi, Sweden]	Rejected. It is usual to put the x-axis at y=0, and we think it would be confusing to do otherwise in the figure.
9-1666	9	150				Please choose another color for the observed permafrost regions - it is almost possible to see that line. [Gunilla Svensson, Sweden]	Taken into account. Fig.9.25 is modified
9-1667	9	151	5	151	5	It would be very helpful to put an estimate of the observed error bars on the figure. Sabine et al. And etc. [Ronald Stouffer, USA]	Accepted - figure revised to include error bar on the observations.
9-1668	9	151	7	151	7	I don't see a dotted line in this figure. [Farahnaz Khosrawi, Sweden]	Accepted - figure caption revised.
9-1669	9	151				fig 9.26 – did you diagnose this for land based on land carbon stores, or fluxes? If by fluxes then be careful how different models have defined and reported land-use emissions. May be better to look at land carbon storage as a diagnostic [CHRIS JONES, United Kingdom of Great Britain & Northern Ireland]	Noted. The fluxes were integrated to give cumulative uptake. We will check that this agrees with the change in carbon stores.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1670	9	151				Figure 9.27: Same here. Move the x-axis to the left and include a zero line. [Farahnaz Khosrawi, Sweden]	Rejected. It is usual to put the x-axis at y=0, and we think it would be confusing to do otherwise in the figure.
9-1671	9	151				Observation based estimates; would be helpful to put uncertainty on that to provide context to differences with models. [Stephen E Schwartz, USA]	Accepted - figure revised to include error bar on the observations.
9-1672	9	152	7	152	7	Observation-based estimates provided by the Global Carbon Project are indicated with black asterisk, not as the dotted line [Ladislav Metelka, Czech Republic]	Accepted - figure caption revised.
9-1673	9	152				Caption refers to observations in dotted line; I dont see it. [Stephen E Schwartz, USA]	Accepted - figure caption revised.
9-1674	9	153	1	154	8	Unit of AOT is not reported in the figures nor in their legenda. [Claudio Cassardo, Italy]	Accepted -- AOT has been replaced by AOD (preferred term)
9-1675	9	153		154		You use the term AOT aerosol optical thickness; the quantity is AOD, aerosol optical depth, which is column integral of extinction coefficient; AOT refers to integral along a specified path; see AMS glossary of Met. Not a big deal but might as well get it right. [Stephen E Schwartz, USA]	Taken into account -- combined with comment 9-1675.
9-1676	9	153				This is a good figure but more work is needed on the color scale (obviously there is no relative error between -200 and -100%). Can you have more colors in the 0 to 60% relative error range? [Olivier Boucher, France]	Taken into account -- Color scale is designed to reproduce color scale in Kinne et al AEROCOM paper as closely as possible. We have added contours to distinguish positive errors.
9-1677	9	153				According to the caption the quantity that is shown is relative error. It would be much more valuable, indeed, I think, essential to present absolute error, in units of AOD. This readily translates into error in aerosol radiative effect (toa flux with aerosol minus without). An absolute error of 0.1 in AOD of a scattering aerosol translates to error in flux of 3 W m-2 [McComiskey A., Schwartz S. E., Schmid B., Guan H., Lewis E. R., Ricchiazzi P., and Ogren J. A. J. Geophys. Res. 113, D09202, doi:10.1029/2007JD009170 (2008)]. Lets say that the AOD = 0.1; then rel error between 0.25 and 0.5 is abs error of .025 to .05 or forcing error of 0.75 - 1.5 W m-2. Even if ocean all looks to be within ±0.25 rel error, that still encompasses much possible error in W m-2, so the scale needs to be much more finely divided. One really needs to see the calc'd and measd aod, and absolute error, not just relative differences. [Stephen E Schwartz, USA]	Taken into account -- Figure is designed to reproduce figure Kinne et al AEROCOM figure.
9-1678	9	154				The brightening trend is not discussed in chapter 7, it is in chapter 2, page 53. I have little confidence in the GACP aerosol trends over ocean and this is also stated in chapter 2 (page 53, lines 32-33, remember how long it took to get the T trend right in the satellite record). This is not a very informative plot in my opinion except to say that there is no obvious trend neither in the obs not in the models. You would need to explain why some models have stratospheric AOD included and others not. [Olivier Boucher, France]	Taken into account -- Figure has been modified to address these issues, as follows: (1. The time range has been expanded to include time period from 1850s to present, and AODs retrieved for present and inferred for 1850 are plotted for comparison; (2. The GACP data has been replace with the MODIS-based data-assimilation-quality product of Zhang and Reed.
9-1679	9	154				Figure 9.29: A legend is missing. Which model is represented by which curve/color? [Farahnaz Khosrawi, Sweden]	Accepted -- Model legend has been added.
9-1680	9	154				Figure 9.29: Legend is missing also the individual model plots are not visible clearly. [HASIBUR RAHAMAN, India]	Taken into account -- Combined with comment 9-1679.
9-1681	9	154				It seems that the models bear little resemblance to observations or to each other, say in annual cycle. Errors are substantial in terms of forcing; see comment on page 9-153. Models should be identified; some discussion needed to the effect that one model (specify) incorp aerosol effects; others did not. Citation to observations should be given. [Stephen E Schwartz, USA]	Accepted -- citation to observations is given, and models are identified in the legend.
9-1682	9	154				The thin dashed/dotted lines look quite constant. No volcanoes? If so, mention! The thick black line is not explained. The thick red line neither. [Andreas Sterl, Netherlands]	Accepted -- Demarcation lines are now explained. Absence of volcanic aerosols from models lacking prognostic stratospheric aerosols is detailed.
9-1683	9	154				minor lines are too thin [Larry Thomason, United States of America]	Taken into account -- combined with comment 9-1682

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1684	9	154				The vertical dashed line close to 92 needs an explanation, and for other lines. [Zhaomin Wang, UK]	Taken into account -- combined with comment 9-1682
9-1685	9	155	4	155	5	suggest to use 'CMIP3 ensemble means (red lines)' [Zhaomin Wang, UK]	taken into account. Note that the figure has been removed
9-1686	9	155	4	155	6	Is that an average over all CMIP3 models? How many models have been considerd? [Farahnaz Khosrawi, Sweden]	Taken into account. Note that this figure has been removed
9-1687	9	155				Figure 9.30: for the tropics, either the annual average or the average over MAM + SON seasons is more relevant, because the diurnal cycles are strongest during the inter-monsoon period rather than during the monsoon seasons of DJF and JJA. Perhaps the author of the referenced paper, who is also in the contributing author list, could be contacted to revise the figure? [Tieh-Yong Koh, Singapore]	Taken into account. Figure updated with input from the CA lead authors of the corresponding referenced paper
9-1688	9	156	4	155	5	Which models are shown? [Farahnaz Khosrawi, Sweden]	Taken into account. Model are now referenced
9-1689	9	156				Figure 9.31: Please add a legend to this Figure. [Farahnaz Khosrawi, Sweden]	taken into account. Legen revised
9-1690	9	156				Figure 9.31: as for Figure 9.30 above [Tieh-Yong Koh, Singapore]	Taken into account. Figure updated with input from the CA lead authors of the corresponding referenced papers
9-1691	9	156				figures are too small, fonts are too fuzzy and generally too small, overall image quality is low [Larry Thomason, United States of America]	Noted.
9-1692	9	157	1			The figure is cropped so that "OLR" looks like "OLF". If possible, update with CMIP5. [Philip Mote, USA]	taken into account. Figure updated and typo corrected
9-1693	9	157	5	157	5	Empirical' [Zhaomin Wang, UK]	editorial. Figure and legend updated
9-1694	9	157	13	157	13	Figure 9.32 Sperber,Gualdi year is missing [HASIBUR RAHAMAN, India]	editorial. corrected
9-1695	9	157				Figure 9.23: This figure is not really helpful and could be omitted. [Farahnaz Khosrawi, Sweden]	rejected. This was a place holder. MJO is recieving lots of attention for variability and change
9-1696	9	157				Figure 9.32 Legend shouls be AVHRR OLR not OLF [HASIBUR RAHAMAN, India]	editorial. Figure and legend have been updated
9-1697	9	157				the rare figure that could be smaller [Larry Thomason, United States of America]	noted
9-1698	9	158				The caption to this figure appears wrong. It says first that shading shows monsoon precipitation in mm/day, but then implies that a 'threat score' is shown. This is not defined. The caption says that 1.0 indicates perfect agreement, but the plotted quantity appears to be anomalies centred on 0. [Nathan Gillett, Canada]	Taken into account.
9-1699	9	159	5	159	12	Something is not clear to me in this figure. The caption mentions "a) several simulations" and "b) … for a subset of the same models". But in a) there are four models, while in b) there are 9 models. How 9 can be a subset of 4? [Claudio Cassardo, Italy]	Taken into account. Figure and legend have been updated
9-1700	9	159				Can observations be added to this? [Nathan Gillett, Canada]	Taken into account. The figure has been updated
9-1701	9	159				Are the two panels interchanged? [Andreas Sterl, Netherlands]	Taken into account. Figure and legend have been updated
9-1702	9	159				Some of th eline colors are much too light (cyan?), particularly the ECHO-G line in the upper frame. [Larry Thomason, United States of America]	Taken into account. Figure updated
9-1703	9	160	4	160	9	It is not clear at which temporal period are referred the averages in the figure. [Claudio Cassardo, Italy]	Taken into account - Figure caption has been expanded considerably.
9-1704	9	161	6	161	8	"ECMWF reanalysis in (b) refers to the European Centre for Medium Range Weather Forecasts (ECMWF) 15-year reanalysis (ERA15) as in (a)": in the 2nd draft, I suggest to standardize as to have two different labels for the same dataset is misleading. Another most important point is that I do not understand why, in the figure, the	Taken into account. Figure, data and legend have been updated. SST re-analysis are different reconstructions and as such can have different

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
						largest difference (which is the cause of the different scale) is in the ERA15 reanalyses, that should be equal in the two graphics (models can differ, but data should be equal!). [Claudio Cassardo, Italy]	spectral characteristics.
9-1705	9	161	10	161	11	Year is missing in the reference [HASIBUR RAHAMAN, India]	Taken into account. Figure, data and legend have been updated.
9-1706	9	162	4	162	5	Include "blue" and "red" in parantheses after CMIP3 and CMIP5. [Farahnaz Khosrawi, Sweden]	Taken into account. Figure, data and legend have been updated.
9-1707	9	163	4	163	15	It seems essential to include a sea ice error metric in this figure. [Thierry Fichefet, Belgium]	Rejected - the figure is intended to cover only temperature and precipitation extremes.
9-1708	9	163				In this figure it is clear that the two reanalysis are very different in terms of representation of the diurnal temperature range (tdtr) since the models perform so different for this quantity. Compared with ERA40 (why not ERA interim?) three models have the darkest red and one model the darkest blue which tells you that the relative error is large - this is in big contrast to the statements in the text that the diurnal cycle is well represented. [Gunilla Svensson, Sweden]	Noted - as we have narrowed down the choice of indices shown in the updated figure, tdtr is no longer shown. However, the apparent inconsistency between text and figure has been checked. It may be related to differences in underlying data and definitions used.
9-1709	9	163				Is this the only way to display this information? It is indecipherable to me [Larry Thomason, United States of America]	Noted - this is intended to follow the style of Figure 9.9 (new 9.7), which is one of well-known ways to display relative model errors. The updated Figure 9.38 (new 9.37) more exactly follows the style of Figure 9.9 (new 9.7), which is hopefully easier to understand.
9-1710	9	164	4	164	6	Please specify the years also for the model simulations. [Claudio Cassardo, Italy]	taken into account. Figure caption revised
9-1711	9	164				Figure 9.39 :Legends and plots are not visible properly [HASIBUR RAHAMAN, India]	taken into account. Figure revised.
9-1712	9	164				Figures are way too small, lines too thing, image quality is poor to the point of making the figure useless [Larry Thomason, United States of America]	taken into account. Figure revised.
9-1713	9	165	1	165	1	What the smaller box into figure a stands for? [Claudio Cassardo, Italy]	Taken into account. The caption was amended to provide an explanation.
9-1714	9	165				the light grey in the upper frame is much too light [Larry Thomason, United States of America]	Taken into account. Figure updated
9-1715	9	167	5	167	5	What do you mean with modern? Please be more precise. [Farahnaz Khosrawi, Sweden]	Taken into account. Figure updated and moved to section 9.4.1
9-1716	9	167	5	167	5	The regions for which the figure has been performed, thus North Atlantic Europe and Tropics should be mentioned somewhere in the caption. [Farahnaz Khosrawi, Sweden]	taken into account. Note that his figure has been removed
9-1717	9	167				Figure 9.42: I am not sure if this figure is really helpful. More explanations are needed to describe what is shown, what the purpose of this figure is and why these regions have been chosen. [Farahnaz Khosrawi, Sweden]	taken into account. Note that the figure has been removed
9-1718	9	168				Love this figure, the most important figure in the chapter [Larry Thomason, United States of America]	Noted - thank you.
9-1719	9	170	4	170	4	Shaded? I guess you mean colored? [Farahnaz Khosrawi, Sweden]	Considered. This figure has been removed since the revised text refers to chapter 12 for detailed discussion on this issue.
9-1720	9	170				Figure 9.45: This figure is difficult to understand. How does one has to read this figure? Does that mean that the range for the sea ice extent for a temperature increase of 0.5 degrees is between 0 and 5 x 10^6 km2? Thus covering everything from all to nothing? [Farahnaz Khosrawi, Sweden]	Considered. This figure has been removed since the revised text refers to chapter 12 for detailed discussion on this issue.
9-1721	9	170				Figure 9.45: there is a new and final version of this Figure as paper is accepted. Please contact me. [Irina Mahlstein, Switzerland]	Considered. This figure has been removed since the revised text refers to chapter 12 for detailed discussion on this issue.

Comment No	Chapter	From Page	From Line	To Page	To Line	Comment	Response
9-1722	9	170				lines are too thin [Larry Thomason, United States of America]	Considered. Figure removed, see above.
9-1723	9	171	5	171	5	What is the abbreviation RSME standing for? [Farahnaz Khosrawi, Sweden]	taken into account -- figure and caption to be replaced
9-1724	9	171				Figure FAQ 9.1: The text on the y-axis is unreadable. [Farahnaz Khosrawi, Sweden]	taken into account -- figure and caption to be replaced