Compilation of Comments on the Public Review Draft of CCSP Synthesis and Assessment Product 1.1: "Temperature trends in the lower atmosphere – steps for understanding and reconciling differences"

I. Introduction

The 45-day public comment period for CCSP Synthesis and Assessment Product 1.1 concluded on January 4, 2006. All public comments received during this period were individually evaluated in accordance with the *Guidelines for Producing CCSP Synthesis and Assessment Products.* This compilation provides a record of the comments received and the Author Team responses.

The 3rd draft of the CCSP Synthesis and Assessment Product 1.1 reflects consideration of all the public comments. Subsequent to the comment period, an open public meeting was held in Chicago, Illinois on February 8-9, 2006, to address the resolution of the comments. Following the Chicago meeting, the revised 3rd draft of CCSP Synthesis and Assessment Product 1.1 was completed in accordance with the rules of the Federal Advisory Committee Act. On March 15 2006 the 3rd draft was posted on the CCSP website. In conformance with *Guidelines for Producing CCSP Synthesis and Assessment Products*, the final version of CCSP Synthesis and Assessment Product 1.1 will be released subsequent to consideration and approval by the CCSP Interagency Committee and the National Science and Technology Council.

II. Names of Commenters

Comments were received from one team and from eleven individuals:

Names:	Dr. William Chameides, Dr. James Wang, and Dr. Lisa Moore
Organization:	Environmental Defense, New York NY
Area of expertise:	Atmospheric science (Chameides), atmospheric science (Wang), ecology (Moore)

Name:	David Douglass
Organization:	Dept of Physics and Astronomy, University of Rochester
Area of Expertise:	Not Given

Name:	Haroon Kheshgi
Organization:	ExxonMobil Research & Engineering Company, Annandale, NJ
Area of Expertise:	earth system models, paleoclimate, attribution, integrated assessment, mitigation

Name:	Michael MacCracken
Organization:	Climate Institute
Area of Expertise:	Climate Change

Name: Alastair B. McDonald
Organization: The Open University, Wimborne, Dorset BH21 1BP, U.K.
Area of Expertise: Amateur Earth System Scientist

Name: Jim Meyer
Organization: Not Given
Area of Expertise: Not Given

Name: Roger Pielke, Sr.
Organization: Colorado State University
Area of Expertise: Weather and climate; Original Convening Lead Author of CCSP Chapter 6

Name: Professor Alan Robock
Organization: Department of Environmental Sciences, Rutgers University
Areas of expertise: Climate data analysis, climate modeling

Name: S. Fred Singer
Organization: University of Virginia/SEPP
Areas of Expertise: Atmospheric Temperature Trends

Name: R. E. Swanson, MsME,
Organization: Independent
Areas of Expertise: Research Engineer

Name: Kevin E. Trenberth
Organization: National Center for Atmospheric Research, Climate Analysis Section
Area of Expertise: Climate Analysis

Name: Derek Winstanley
Organization: Illinois State Water Survey, Illinois Department of Natural Resources
Area of Expertise: Climatology

III. Report Section Sorting Structure

The comment sorting routine followed the Report section structure:

Preface
Executive Summary
Chapter 1
Chapter 2
Chapter 3
Chapter 4

Chapter 5
Chapter 6
Appendix A (Statistical Appendix A)
Appendix B (Members of the Assessment / Synthesis Product Team)
Glossary

Responses to comments that were not addressed to a specific report location were labeled General and are included at the end of this compilation following the Appendix B and Glossary comments/responses.

IV. Response Sorting/Labeling System

For the purpose of responding to the comments, responses were labeled with the commenter's name and the Report section addressed. As an example of the labeling system:

Doe ES-1 would be John Doe's first comment on the Executive Summary
Doe ES-2 would be John Doe's second comment on the Executive Summary
Doe CH1-1 would be John Doe's first comment on Chapter 1

Preface Comments and Responses:

Chameides, et.al., Pre-1, Page 1, Line 1 (We suggest a succinct abstract that states up front and explicitly that one of the reasons this assessment was conducted was that the previous discrepancy between surface and tropospheric observations was used to challenge the correctness of models and the whole idea of greenhouse gas-induced global warming. The findings in this assessment bolster the consensus view that recent warming is due in large part to anthropogenic greenhouse gases.) Chameides, Wang & Moore, Environmental Defense

Response: A two-paragraph abstract that addresses the reviewers' concerns has been has been inserted in front of the Preface and also included in the Executive Summary.

Chameides, et.al., Pre-2, Page 1, Line 12 (Add a sentence or two explaining why this topic is relevant to decision-making: recent research has resolved the reported differences between observations and models that were used to argue against the existence or cause of climate change.) Chameides, Wang & Moore, Environmental Defense

Response: See previous response.

Chameides, et.al., Pre-3, Page 1, Line 14 (The preface gives a very good historical overview of the problem. It can be strengthened in two ways. First, it should state at the beginning why temperature trends in the lower atmosphere matter—why has this topic received so much attention? Why should a policymaker care? Second, since the current draft of the report does not mention the key findings until the 15th page, the preface leaves the reader with the sense that the problem is still unresolved, when in fact this is not the case.) Chameides, Wang & Moore, Environmental Defense

Response: A sentence was added to the first paragraph stating that several earlier discrepancies have been resolved.

Chameides, et.al., Pre-4, Page 8, Line 171 (The report "promises to be of significant value to decision-makers" but it is not clear how policymakers would use the document. The relevant conclusions—that recent analyses and corrections have resolved previous inconsistencies between observations and models, and that the results further strengthen the evidence for anthropogenic interference with the climate system—are not given the prominence they deserve.) Chameides, Wang & Moore, Environmental Defense

Response: A sentence has been added at the beginning of the paragraph to clarify this.

Chameides, et.al., Pre-5, Page 8, Lines 172-177 (The sentence "Readers of this Report will find that new observations, data sets, analyses, and climate model simulations enabled the Author Team to resolve many of the perplexities…" does not explain what the resolutions were. The paragraph goes on to say that the Report "has had an important impact" on the IPCC FAR, but again does not say why. This section will be much more accessible to non-specialist audiences if it explicitly explains the main results of the Report: that previous discrepancies between observations and models have been resolved and that the results add further support to the overwhelming scientific consensus that human activity is affecting the climate system.) Chameides, Wang & Moore, Environmental Defense

Response: A sentence has been added to this paragraph indicating that additional evidence in support of anthropogenic influences on climate change.

Kheshgi Pre-1, Page 10, Line 210: In, for example, ES Figure 3 the label "Mid to Upper Troposphere" is inconsistent with the preface line 210 (table 1) definition of T2 which is "Mid to Lower Stratosphere": this should be corrected for consistency throughout the document.

Response: Corrected

MacCracken Pre-1, Page 5, Line 104: In that the system is chaotic, it is unrealistic to expect the models to "replicate" the observed temperature changes, except within some statistical bounds. Such a caveat needs to be added.

Response: Sentence added to reflect this.

MacCracken Pre-2, Page 5, Lines 107-110: The words "are not" are too definitive, even given the list of suggested reasons, as the system is chaotic and so one should not actually be expecting replication of the record, except within some bounds. Also, the word "serious" should be deleted, as it is not indicated how serious the departure might be.

Response: Deleted the word "serious", but since the words "are not" are followed by adequate this helps to qualify the statement. Ensemble runs are used to help assess the chaotic nature of the climate systems. This is not specifically mentioned in this paragraph.

MacCracken Pre-3, Page 5, Lines 116-117: Rephrasing is needed, for not all levels of the atmosphere respond in the same way either, and we want to understand that as well.

Response: Done

MacCracken Pre-4, Page 5, Lines 119-120: This report also clears up a number of misrepresentations of past supposed conclusions, so makes progress in understanding what is going on; it does not just outline steps for doing this. A much more forthright statement is needed here about what this assessment effort has actually done—it has to a large degree actually done the reconciliation and has made clear that the remaining gaps in our understanding do not introduce a serious challenge to the model representations of the climate system and of climate change.

Response: A specific sentence has been added indicating this (last sentence of "Focus of this Synthesis/Assessment Report")

MacCracken Pre-5, Page 5, Line 123: The phrase "reducing the uncertainties" is really not very helpful to decision makers—the glossary includes no definition of "uncertainties" and there is really no indication of what level of uncertainty matters nor a metric of how far it needs to be reduced, etc. It would be more useful to say that the results presented here have improved our understanding of surface-atmosphere coupling to a large degree, such that there no longer remains any supposedly serious disagreement between observations and the results of climate change models, wiping away the main excuse that has existed in the minds of many doubters with regard to accepting the IPCC projections of climate change.

Response: The Preface makes it clear that conclusions of the report are in the Executive Summary, but a sentence has been added to indicate that the report provides more evidence for anthropogenic influences on climate.

MacCracken Pre-6, Page 6, Line 127: The word "analysis" should be "analyses" given all the work that has gone into this effort.

Response: Done

MacCracken Pre-7, Page 6, Lines 140-141: It is not yet clear that the full significance of the new findings has been incorporated into the report except in quite superficial ways. These new findings really do help to resolve a number of the key issues that had existed, and this needs to be made more clear in the text of the report rather than by just adding a sentence here or there.

Response: The reader is referred to the Executive Summary for the full set of conclusions, but additional text has been added to indicate the significance of this Report.

MacCracken Pre-8, Page 12, Lines 240-242: The sentence here is presumably referring to Appendix B, which actually features a listing of something called the "Assessment/Synthesis Product Team" which is not, mainly, the authors of the report (see general comment). The Appendix should be featuring the scientists who wrote the report and are responsible for it—not the staff support for the effort.

Response: The Author Team is now listed on a separate page immediately following the Table of Contents. Appendix B has been removed from the Report.

Executive Summary Comments and Responses:

Chameides, et.al., ES-1, Page 3, Line 49 (We suggest that the sentence begin with "*Observations—*" to make it immediately clear that this result relates to data sets, unlike the next point, which has to do with simulations.) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 2, Page 3, Lines 50-52 (Recent studies (referred to in the draft as "some data sets") have shown that many of the earlier "majority of data sets" are flawed, but this fact is not mentioned here. Thus, these two sentences imply that there is still a lot of uncertainty about relative rates of warming. This section is an excellent opportunity to emphasize the progress that has been made in this area of research and to highlight the importance of these findings.) Chameides, Wang & Moore, Environmental Defense

Response: The reviewers have misunderstood what data have been used in the Report. 'data sets' here refers only to the latest versions of all data sets. 'earlier' (versions of) data sets are not considered in this Report. This point is now noted specifically in the first and third bullets. Further, there really is still a lot of uncertainty about observed rates of warming – this is a clear conclusion of the Report, and the wording here reflects this. No change.

Chameides, et.al., ES- 3, Page 3, Line 54 (We suggest that the sentence begin with "*Model simulations—*" to differentiate this section from the one above it (i.e., make it clear this point is about models rather than about observations).) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 4, Page 3, Lines 59-62 (This is a critical point that should be given more prominence at the very beginning of the document.) Chameides, Wang & Moore, Environmental Defense

Response: Given the focus of the Report on changes in vertical temperature profiles, the emphasis given here is correct. This Section deals with "New Results and Findings" and, strictly, this bullet (and the one below it) are not new results – so undue emphasis would create a more obvious conflict with the Section heading.

Chameides, et.al., ES- 5, Page 3, Lines 64-65 (This is a critical point that should be given more prominence at the very beginning of the document.) Chameides, Wang & Moore, Environmental Defense

Response: Given the focus of the Report on changes in vertical temperature profiles, the emphasis given here is correct. This Section deals with "New Results and Findings" and, strictly, this bullet is not a new result – so undue emphasis would create a more obvious conflict with the Section heading.

Chameides, et.al., ES- 6, Page 3, Lines 68-69 (Recent research (here referred to as "newer observed data sets") has shown that many of the earlier "majority of observed data sets" are flawed, but this fact is not mentioned here. Thus, these two sentences imply that there is still conflict between observations and models. This section is an excellent opportunity to emphasize that this problem has been resolved.) Chameides, Wang & Moore, Environmental Defense

Response: The reviewers have misunderstood what data have been used in the Report. Earlier data sets are not considered in this Report, so the reviewers' assumption that the "majority of observed data sets" are earlier data sets is incorrect. The Report deals only with the most recent versions of all data sets. The view of the expert author team is that the "problem" of observed data set differences has not been resolved. This is what the Report states both here and in the individual Chapters.

Chameides, et.al., ES- 7, Page 5, Lines 100-107 (The last sentence, "The second explanation is judged more likely", can be stronger. Santer *et al.* (2005) demonstrated that only corrected data produce long-term trends that are consistent with our understanding of physics.) Chameides, Wang & Moore, Environmental Defense

Response: Some reviewers thought this conclusion was too strong, while others thought it was too weak. This implies that the current text has achieved something close to the right balance. In order to explain how this conclusion and its specific wording was arrived at, additional text (extracted from Chapter 5) has been added (in Section 4 of the Executive Summary). Other minor wording changes have been made in response to comments of other reviewers. As an example, here is the first mention of this issue in the revised version of the Executive Summary:

"Although the majority of observed data sets show more warming at the surface than in the troposphere, some observed data sets show the opposite behavior. Almost all model simulations show more warming in the troposphere than at the surface. This difference between models and observations may arise from errors that are common to all models, from errors in the observational data sets, or from a combination of these factors. The second explanation is favored, but the issue is still open."

Chameides, et.al., ES- 8, Page 6, Line 129 (Remove the extra space between "according to" and "location".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 9, Page 10, Line 189 (Remove hyphen from "Independently-performed".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 10, Page 17, Lines 331-341 (Throughout the Figure 2 caption, change "degC" and "degF" to "°C" and "°F", respectively.) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 11, Page 19, Lines 386-388 (Among the "number of observed data sets" that do not show amplification are those that have been shown to be flawed. Because it does not include this important fact, this paragraph implies that there is still a great deal of uncertainty.) Chameides, Wang & Moore, Environmental Defense

Response: The reviewers have misinterpreted this statement (see responses to ES-2 and ES-6). "the most recent" has been added to the text to clarify this.

Chameides, et.al., ES- 12, Page 20, Lines 396-397 (This sentence would make more sense, especially for non-specialist readers, if it included specific examples of "independent physical evidence supporting substantial tropospheric warming".) Chameides, Wang & Moore, Environmental Defense

Response: New text has been added here (from Chapter 5) to explain why the "second explanation is more likely" (which was on lines 394, 395 of the original version). The increase in the height of the tropopause has been added as an example. The new text is as follows:

"This inconsistency between model results and observations could arise due to errors common to all models; due to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or due to a combination of these factors. The new evidence in this Report – model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming (such as the increasing height of the tropopause) – favors the second explanation. Reconciliation of observational uncertainty is a pre-requisite for resolving to what extent model error exists."

Chameides, et.al., ES- 13, Page 20, Line 408 (Remove the hyphens from "previously-ignored" and "spatially-heterogeneous".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 14, Page 21, Line 412 (Remove the hyphen from "spatially-heterogeneous".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 15, Page 21, Line 414 (Remove the hyphen from "spatially-heterogeneous".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 16, Page 22, Line 439 (Change "model/observed similarities and differences" to "similarities and differences between observations and model results".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 17, Page 22, Lines 441-442 (Change "model/observed data inconsistencies" to "inconsistencies between observations and model results".) Chameides, Wang & Moore, Environmental Defense

Response: Done

Chameides, et.al., ES- 18, Pages 23-24 (The different colored rectangles are indistinguishable on black and white printing. The caption identifies them, but it would still be nice to have a visual difference for those who may read a black and white printout.) Chameides, Wang & Moore, Environmental Defense

Response: Cross-hatching has been put on the red rectangles.

Chameides, et.al., ES- 19, Pages 23-24 (These figures give equal weight to the data sets, including some that have been shown to be flawed.) Chameides, Wang & Moore, Environmental Defense

Response: The reviewers have misunderstood what data have been used in the Report. The only data used are the latest versions of all data sets. While some have (very recently) been shown to have potential problems (such as the likely bias in tropical radiosonde data noted by Sherwood et al. and Randel and Wu, cited (e.g.) in Chapter 5), these are still unresolved issues and it is not possible to assign relative credibility levels to the different data sets. Note that it is not yet known the extent to which the problems in individual station data that is noted in the above two references has also affect the homogenized radiosonde data sets used in the Report.

Chameides, et.al., ES- 20, Page 25, Lines 490-491 (Please give examples of "variables other than temperature" that should be compiled.) Chameides, Wang & Moore, Environmental Defense

Response: List of variables added – as now given in Chapter 6.

Douglas ES-1, P2, L 43-44, **Quote from report:** "These changes are in accord with our understanding of the effects of radiative forcing agents and with model predictions." **Comment:** Model predictions are not documented or referenced.

Response: The relevant information is given in Chapters 1 and 5.

Douglas ES-2, P3, L 56-57, **Quote from report:** "Given the range of observed results and the range of model results, there is no inconsistency between models and observations at the global scale." **Comment:** There is no definition of range. It is up to the subjective choice of the author.

Response: "range" is used in its normal English language sense – as the difference between the lowest and highest.

Douglas ES-3, P3, L68-70, **Quote from report:** "The majority of observed data sets show more warming at the surface than in the troposphere, while some newer observed data sets show the opposite behavior. Almost all model simulations show more warming in the troposphere than at the surface." **Comment:** Correct, but other statements in the report contradict this statement.

Response: There are no statements in the Report that contradict this summary, and the reviewer does not identify any such statements.

Douglas ES-4, P5 L93-94, **Quote from report:** "Given this range of results, there is no conflict between observed changes and the results from climate models." **Comment:** The data do not support the "no conflict" characterization.

Response: In the opinion of the expert author team, this is a correct conclusion. The text has been revised slightly to clarify the issue. The new text, which explains what is meant by "no conflict", is:

"Given the range of model results and the overlap between them and the available observations, there is no conflict between observed changes and the results from climate models."

Douglas ES-5, P5, L 100-101, **Quote from report:** "On decadal and longer time scales, however, while almost all model simulations show greater warming aloft, most observations [tropics] show greater warming at the surface." **Comment:** Correct, but other statements contradict this.

Response: There are no statements in the Report that contradict this summary, and the reviewer does not identify any such statements.

Douglas ES-6, P5, L103-107, **Quote from report:** "These results have at least two possible explanations, which are not mutually exclusive. Either amplification effects on short and long time scales are controlled by different physical mechanisms, and models fail to capture such behavior; and/or remaining errors in some of the observed tropospheric data sets adversely affect their long-term temperature trends. The second explanation is judged more likely." **Comment:** No basis for selecting 2nd explanation.

Response: The basis for this conclusion is given in Chapter 5. The wide range of trends in the available observed data sets requires that most of these data sets must give incorrect trends – at most, only one trend value can be correct, and it is entirely possible that none are correct. Furthermore, errors in the observed radiosonde data in the tropics have already been identified in the cited papers by Sherwood et al. and Randel and Wu. Text from Chapter 5 has been added to clarify these points. The new text is:

"This inconsistency between model results and observations could arise due to errors common to all models; due to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or due to a combination of these factors. The new evidence in this Report – model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming (such as the increasing height of the tropopause) – favors the second explanation. Reconciliation of observational uncertainty is a pre-requisite for resolving to what extent model error exists."

Douglas ES-7, P6, L127-128, **Quote from report:** "Temperature trends at the surface can be expected to be different from temperature trends higher in the atmosphere because:" **Comment:** Climate shift of 1970's not listed.

Response: The apparent climate shift around 1976 is described in Chapter 3 (and also mentioned in the Statistical Appendix). This shift is not so evident at the surface, and is only clear in the tropospheric data from radiosondes. It does not affect any of the conclusions regarding changes over the period 1958 to present (and obviously has no relevance to changes over the satellite era). This issue is not considered important enough by the author team to include in the Executive Summary. Nevertheless, it should be noted that the short-term warming that occurred at this time is consistent with the physics of amplification, whereby the tropospheric temperature change should be (and is) greater than the surface change.

Douglas ES-8, P13, L262-3, **Quote from report:** "Since 1979, due to the considerable disagreements between tropospheric data sets, it is not clear whether the troposphere has warmed more than or less than the surface." **Comment:** Not true. Do Thorne and Free agree?

Response: This is true. It is the considered opinion of the expert author team. Thorne is a member of this team, and, of course, he agrees. Free, who has been consulted at numerous times by the author team and who has participated in some of the meetings of the team, also agrees.

Douglas ES-9, P21, L429-431, **Quote from report:** "Figures 3 and 4 summarize the new model results used in this Report, together with the corresponding observations. Figure 3 gives global-mean results, while Figure 4 gives results for the tropics (20 S to 20 N)." **Comment:** Fig 3 and 4 refers to figs 5.3 and 5.4 and tables 5.3 and 5.4. in Chap 5. These plots and tables are new and have not been peer reviewed. See later comments.

Response: This is incorrect. These results have been published in ScienceNote; also note that the text describing Fig. 4 has been modified in response to other reviewers' comments. The new text is:

"For global averages (Fig. 3), models and observations generally show overlapping rectangles. A potentially serious inconsistency, however, has been identified in the tropics. Figure 4G shows that the lower troposphere warms more rapidly than the surface in almost all model simulations, while, in the majority of observed data sets, the surface has warmed more rapidly than the lower troposphere. In fact, the nature of this discrepancy is not fully captured in Fig. 4G as the models that show best agreement with the observations are those that have the lowest (and probably unrealistic) amounts of warming (see Chapter 5, Fig. 5.6C). On the other hand, as noted above, the rectangles do not express the full range of uncertainty, as they do not account for uncertainties in the individual model or observed data trends."

Kheshgi ES-1, Page 2, Line 42: Suggest adding after "Report," the phrase "all data sets show that"; this provides the objective basis supporting this statement

Response: Text changes have been made to cover this point.

Kheshgi ES-2, Page 2, Line 43: It is not clear what is meant (statistically?) by the term "substantially" or why this applies to the lower layers and not the stratosphere. Suggest deleting the term "substantially".

Response: This comes from earlier Chapters. "substantially" has been deleted.

Kheshgi ES-3, Page 2, Line 46: Suggest adding after "1950s," the phrase "all radiosonde data sets show that"; this provides the objective basis supporting this statement.

Response: Done

Kheshgi ES-4, Page 3, Line 49: Suggest adding after ")," the phrase "all radiosonde and satellite data sets show that"; this provides the objective basis supporting this statement.

Response: Judged not necessary in an Executive Summary. The statement implies this in the absence of wording to the contrary. No change.

Kheshgi ES-5, Page 3, Line 54: Suggest adding a footnote after "new" describing what is new about these simulations as opposed to older simulations such as those in Hansen et al (Science 1998, vol. 281, p 930-931) which show amplification in lapse rate trend. In addition, much clearer information is needed to explain what is "new" in models that has affected the amplification evident in models.

Response: Text modified slightly to clarify "new". Further details are given in Chapter 5.

Kheshgi ES-6, Page 3, Lines 45-57: Suggest replacing the final sentence with "The wide range of observed and modeled global trend differences overlap and are, therefore, in this sense consistent." It is important to point out how large these ranges are, since such overlap not a very severe test of climate models

Response: This bullet point (dealing with global-mean temperatures) has been reworded to read:

"The most recent climate model simulations give a range of results for changes in global-average temperature. Some models show more warming in the troposphere than at the surface, while a slightly smaller number of simulations show the opposite behavior. There is no fundamental inconsistency between these model results and observations at the global scale."

Kheshgi ES-7, Page 3, Lines 59-65: These two paragraphs cover the topic of attribution of climate change which does not seem to be in the scope of this assessment product. Suggest deleting these two paragraphs. If a discussion of attribution is retained, then suggest that it reflect the text in the summary of Chapter 5 in lines 6567-6577 which state: "This chapter has evaluated a wide range of scientific literature dealing with the possible causes of recent temperature changes, both at the Earth's surface and in the free atmosphere. It shows that many factors – both natural and human-related – have probably contributed to these changes. Quantifying the relative importance of these different climate forcings is a difficult task. Analyses of observations alone cannot provide us with definitive answers. This is because there are important uncertainties in the observations and in the climate forcings that have affected them. Although computer models of the climate system are useful in studying cause-effect relationships, they, too, have limitations. Advancing our understanding of the causes of recent lapse-rate changes will best be achieved by comprehensive comparisons of observations, models, and theory – it is unlikely to arise from analysis of a single model or observational dataset." If a discussion of attribution is retained, then also suggest that discussion be added to clarify the scope of this SAP's assessment of attribution Vs SAP1.3's discussion of attribution, which is clearly in its scope (and title of the draft prospectus).

Response: It is not possible to add this much detailed text in the Executive Summary – which is meant only to summarize information and results given in individual Chapters. The text in these two bullets is a direct paraphrasing of bullet points given in Chapter 5, where full details and justification is given. I have added "over the past 50 years" to the second bullet (original lines 64, 65) to clarify the time interval and to be consistent with the Chapter 5 wording.

Kheshgi ES-8, Pages 4-5, Lines 89-94: These sentences consider global, not tropical temperatures. Suggest moving first two sentences to global section following line 292 although this is mostly redundant and might be mostly deleted. Suggest deleting 3rd sentence since it is redundant with that in the previous section

Response: The Section on tropical results ends with (original text) line 70. The text noted here is meant to be more general. This has been clarified by inserting "global and tropical" on line 72. The 3rd sentence is considered to be an important reminder of an important point that warrants repetition.

Kheshgi ES-9, Page 5, Lines 96-107: While this section does a good job describing the differences in model results, it does a poor job explaining why. We should know why models are giving such a wide range of amplification, why the range in the tropics is different than for global averages. This report gives the impression that we do not know what the models are doing? If this is so, then this indicates a gap in understanding that should be addressed. If this in not so, then this understanding needs to be outlined here.

Response: Models do not show a wide range of amplification in the tropics (see original text line 99). Additional text has been added (from material in Chapter 5) that should cover the reviewer's concerns.

Kheshgi ES-10, Page 5, Line 101: A final sentence should be added here saying why roughly half the models show global/decadal amplification, while almost all show tropical/decadal amplification.

Response: Text has been added to explain differences between the tropics and other latitudes. The new text is:

"Over the period since 1979, for global-average temperatures, the range of recent model simulations is almost evenly divided among those that show a greater global-average warming trend at the surface and others that show a greater warming trend aloft. The range of model results for global average temperature reflects the influence of the mid- to high-latitudes where amplification results vary considerably between models. Given the range of model results and the overlap between them and the available observations, there is no conflict between observed changes and the results from climate models."

Kheshgi ES-11, Page 8, Line 161: Table 1: I could not find justification in the underlying text for volcanoes for the "short-term" label on stratospheric warming, and not for the surface or troposphere. Suggest removing "short-term". Simply looking at Figure 1 suggests a similar time-scale for the stratosphere and troposphere, however, a much greater effect relative to other variation in the stratosphere than in the troposphere.

Response: "short-term" has been deleted and text added to the caption to explain that there are response time differences between the stratosphere and the troposphere (see. e.g., Wigley et al., JGR (2005)).

Kheshgi ES-12, Page 9, Lines 171-172: Reasons should be given here for why outside of the tropics there may be attenuation as opposed to amplification.

Response: This is covered in Chapter 1. See also response to ES-10.

Kheshgi ES-13, Page 18, Line 352: Suggest including a discussion how this question is interpreted; what is meant by "reconciled with our understanding". The text in this section jumps to statements about attribution that have unclear relation to this question.

Response: A discussion here of the word 'reconcile' would not be appropriate – this should be covered (if at all) in the Preface. The issue of giving attribution results is an issue for Chapter 5 to justify (and the response to this is given in Chapter 5's responses to this reviewer's comments). The Executive Summary rests on (and must summarize) material in the other Chapters.

Kheshgi ES-14, Pages 18-20, Lines 355-397: This text is essentially all redundant with earlier text in this summary. Suggest deleting this redundant text and including a paragraph describing the additional information that is available in the data sets that goes beyond just their trends, and how this may be reconciled with our understanding.

Response: Again, the Executive Summary rests on (and must summarize) material in the other Chapters. The structure of the Executive Summary is to give some general overview and then to go through the specific key points given in the individual Chapters. Because of this there is some unavoidable duplication of material.

Kheshgi ES-15, Page 19, Line 387: Suggest replacing "A number of observed" with "Most" to be consistent with earlier text.

Response: "A number of the most recent" has been changed to "Most of the most recent".

Kheshgi ES-16, Page 22, Line 442: Suggest replacing "fully overlapping rectangles" with "rectangles with extensive overlap" to improve clarity. --

Response: The word "fully" has been deleted.

Kheshgi ES-17, Page 23, Line 447: In Figure 3 the label "Mid to Upper Troposphere" is inconsistent with the preface line 210 (table 1) definition of T2 which is "Mid to Lower Stratosphere": this should be corrected for consistency throughout the document.

Response: Presumably the reviewer means that this should be "Mid Troposphere to Lower Stratosphere". The Figure has been corrected.

MacCracken ES-1, Page 2, Lines 46-47: Are not these results also consistent with model predictions, as was the finding in the first bullet (lines 42-44). Similar treatment needs to be given to this finding.

Response: Text added as in first bullet.

MacCracken ES-2, Page 3, Lines 68-70: This phrasing is really inappropriate. First, it seems to imply that science is more a matter of voting than of real understanding. It seems to me quite unscientific to be giving equal weight to datasets of differing credibility (and differing histories of having to be corrected and updated) based on not only how they are constructed but in how they are consistent or inconsistent with other evidence. Second, the phrasing places the datasets that are based on the most rigorously presented methodologies in the second phrase. Saying that they are the "newest" really does not do justice to what has been done and found. This bullet needs to be replaced with something like "Observational data sets that account for all of the adjustments and biases that have been found to be important to generating an accurate representation of atmospheric behavior show more warming in the troposphere than at the surface, in agreement with model simulations. This new finding supercedes an earlier finding that was based on datasets that had not adequately considered shortcomings in the observational methodology."

Response: The use of "majority" is correct English language usage, and simply a statement of fact. The word "newer" has been deleted. The result here does not supersede any earlier findings, since previous reviews did not isolate changes in the tropics.

MacCracken ES-3, Page 4, Lines 79-80: There is really too little context provided here. For example, it is not explained why this matter is "crucial." This could be addressed by referring to an appendix that presented the issue in a bit more detail and went through the history, as suggested in one of my general comments, indicating that early (mis)interpretations of the MSU results were being widely touted by skeptics to suggest that there were shortcomings in the model simulations and therefore to question the very strong overall understanding of the climatic changes expected to result from increasing the GHG concentrations. The preceding paragraph and subsequent text might be fine for a scientific audience had this issue been just some remote, hidden matter always discussed in highly technical terms at far off meetings, but this issue has been front and center in discussions of climate change by a variety of politicians and other interests, and much more context needs to be provided.

Response: Text modified to make the issue of amplification, and the reason why it is "crucial", clearer. The word "crucial" is no longer used. The new text is:

"The issue of changes at the surface relative to those in the troposphere is important because larger surface warming (at least in the tropics) would be inconsistent with our physical understanding of the climate system, and with the results from climate models. The concept here is referred to as "vertical amplification" (or, for brevity, simply "amplification"): greater changes in the troposphere would mean that changes there are "amplified" relative those at the surface."

MacCracken ES-4, Page 4, Lines 82 and 85: In both lines, change "amplification" to "vertical amplification" so that there is no confusion over the issue of "horizontal amplification" that is talked about as due to albedo feedback, etc. in high latitudes.

Response: This distinction is now clarified.

MacCracken ES-5, Page 4, Line 87: Again, the reference to "most data sets" is given too much prominence—voting is not what matters (the use of the word "show" seems to implicitly imply that these data sets are still considered credible). Over the period since 1979, there has really been only one group generating the satellite data set, and though there have been many versions as they have corrected successive problems, this phrasing indicates that there are many data sets that show this (the radiosonde ones may as well, but that should be mentioned separately). A better phrasing might be "Since 1979, incompletely corrected versions of the satellite and radiosonde records have both shown slightly greater warming at the surface." That is, make clear that this conflict no longer exists rather than giving it any further credence.

Response: The use of "most" is correct. Whether or not some data sets are better than others is not an issue that is resolved in the Report. Unfortunately, the conflict does still exist.

MacCracken ES-6, Page 4, Line 91 to page 5, line 94: This supposed "evenly divided" state with regard to global changes encompasses many more processes than what are being described here—such as albedo feedback, ability to represent surface inversions, etc. Making the statement in this way thus creates more doubt about all of this than is justified here—indeed, it really does not matter so much for the world which is changing more given the significant disconnect between surface and atmospheric temperature variations that is described in the text. Thus, this statement really is adding to confusion in a misleading manner.

Response: New text has been added to explain the "evenly divided" result. The full details suggested cannot be given here, partly because they are the concern of Chapter 5, but also because such detail would not be appropriate in this Summary chapter.

MacCracken ES-7, Page 5, Line 98: Change "amplification" to "vertical amplification"

Response: No change. The distinction has been clarified earlier.

MacCracken ES-8, Page 5, Line 104: "different" than what?

Response: "different" in this context does not require a qualifier.

MacCracken ES-9, Page 6, Line 131: Change to read ""smoothed out by the motions of the atmosphere so the patterns"

Response: Done

MacCracken ES-10, Page 8, Line 161: The entry for "Volcanic Eruptions" in column 2 needs to be modified to indicate that the response differs by location and type of eruption and by season, as winter warming can occur over some continental areas. This chart seems to make everything seem too simple.

Response: The point here is to give a simple overview. The Table caption says "effects on global-, annual-mean temperatures" – note the added emphasis.

MacCracken ES-11, Page 9, Lines 169-170: Does this reasoning also imply to volcanic eruptions and the cooling that results—in that it seems to, this also might be mentioned.

Response: The text already states that amplification is "largely independent of the type of forcing".

MacCracken ES-12, Page 13, Lines 262-263: This statement seems to imply that no progress in understanding has come since 1979 and that all analyses done over this time are equally valid. This is simply not the case—that there has been disagreement is a result of the failure to early on understand how to correct the observational datasets for their various biases and problems, and this needs to be indicated as the reason. This sentence should be rewritten to something like "Recent advances in improving and correcting the observational record for biases and problems has resolved most of the disagreement that has existed over the past 20 years regarding the relative warming of the troposphere and the surface." Waiting to make this point until later is not adequate—an affirmative result from this assessment effort needs to be stated. [Note—I don't think this disagreement actually goes back to 1979, even though the data may—a bit of rephrasing is needed.]

Response: The reviewer has apparently misunderstood "since 1979". This refers to changes in temperature from 1979, not to changes in the records themselves. Unfortunately, disagreements between data sets have not been resolved.

MacCracken ES-13, Page 13, Lines 265-268: It seems irresponsible to not be giving error bounds on the various estimates—just saying "about" is really not enough to gain an understanding about whether the results are or are not in agreement or significantly different.

Response: This information comes straight from Chapter 3, where confidence limits are given. Please recall that this is an Executive Summary.

MacCracken ES-14, Page 13, Lines 270-271: Change "that trends" to "that estimates of trends" and "troposphere" to "tropospheric". More generally, this phrasing is rather misleading—of course, errors likely remain in everything (so "very likely" is technically correct), but are these errors very likely to be as significant as past errors? Given that the various interlocking sets of measures are all now much more in tune than before when there were significant conflicts, it would seem quite likely that the remaining errors are not as important as before, and this too should be stated—we do now have greater overall confidence, both because we have looked much more intensively and because of the way that various findings better fit together.

Response: Wording changes made as suggested. "Very likely" uses the terminology lexicon given in the Preface. There are still significant conflicts between the various tropospheric temperature data sets (see original text lines 262, 263; lines 273, 274; and also Fig. 4 in the Statistical Appendix.)

MacCracken ES-15, Page 13, Line 272: Change "cases" to "causes"

Response: Done

MacCracken ES-16, Page 14, Lines 287-289: "large" relative to what? It is really not clear that these differences are particularly important, given that linear trend analysis for the stratosphere is problematic due to volcanic effects, interferences, etc. Remember, this executive summary is for the wider audience, and the statements need to be made very carefully so as to give the right overall impression, and not a misimpression because scientists simply want to resolve each and every difference.

Response: "large" deleted.

MacCracken ES-17, Page 15, Figure 1: In the key, change "Volcanic Eruption" to "Major Volcanic Eruption" to make clear that only the very largest are being indicated. More generally, is it not the case (and even likely) that some of the wiggles might be a result of the influence of smaller volcanic eruptions than those listed (e.g., in the 1970s)?

Response: Change made.

MacCracken ES-18, Page 16, Lines 311-317: As indicated in my general comment, I would urge dropping the phrase "a time coincident with a previously identified climate regime shift." First, this shift is based largely on the radiosonde data, and these data sets are being updated and changed, so this whole notion needs a new look. At least as to how it relates to the analysis of the tropospheric temperature record, it was rather arbitrarily determined and is based upon a particular choice of end points for the analysis. It is also not clear how the incomplete spatial coverage of the radiosonde record may contribute to this apparent shift, if that is what it is. Second, it is not clear if it is a natural fluctuation, driven by a natural forcing (e.g., volcanic), or perhaps human-induced (e.g., due to a change in sulfur emissions around that time)—or something else. It is also not clear it should really be referred to as a "climate regime shift" as it occurred mainly in one area of the world and involved mainly the atmospheric circulation—and not surface temperature, as is later pointed out in this report. There is also no discussion of how this might relate to other such shifts, in other variables, at other times, in other places, etc. It is simply an extraneous bit of speculation—mere coincidence does not prove anything. So, in my view, it should be discarded.

Response: The reviewer's skepticism regarding the apparent shift around 1976 is understandable, but the published literature supports the current text. This text comes from material in Chapter 3. Some of the current text has been deleted in accord with the reviewer's suggestion.

MacCracken ES-19, Page 16, Line 316: I simply do not see a "rapid rise" in the mid-1970s. I see some fluctuations occurring, and how they line up might give an impression of a rapid rise, but also may be purely coincidental, etc. This report needs to be much more questioning about all of this, and not simply accept such claims and analyses (and in any case, this is really not the subject of this report, so why cover this supposed shift at all?

Response: The statements here reflect earlier Chapters, and are consistent with the published literature (which examines different 'models' for the temperature change, such as a step change versus a gradual (but noisy) trend). The text has been shortened slightly in response to this comment. The wording is now "a major part" not "the major part" (original text line 315), and the use of "appears" (line 316) seems to be sufficiently circumspect.

MacCracken ES-20, Page 18, Line 346: Change "estimating" to "estimating and deciphering" as understanding what is going on is why we really use models.

Response: Done

MacCracken ES-21, Page 19, Lines 366-367: This phrasing here is very misleading. Natural factors cannot even come close to "fully explain[ing]" the changes over the past 50 years—indeed, natural factors would likely have been causing a cooling, yet this statement seems to indicate that natural factors are not so far off doing so. This is a phrasing out of statistics and its hidden message will be lost on the average reader. This sentence, and others like it in this report need to be changed to something like: "While natural factors are likely to have contributed to some of the climate fluctuations of the past 50 years, the overall warming can only be explained as mainly the consequence of human influences."

Response: The Executive Summary can only use (or paraphrase) the wording that is used in Chapter 5. In defense of this statement, it should be noted that "natural factors" refers to both internally generated changes and externally forced (solar and volcanic) changes. It is possible that the reviewer has forgotten about internally generated changes, which are virtually impossible to quantify (except as a range of possibilities). The text has been expanded to explain what is meant by "natural factors". For example, in the Key Findings section it states"

"Natural factors (external forcing agents like volcanic eruptions and solar variability and/or internally generated variability) have influenced surface and atmospheric temperatures, but cannot fully explain their changes over the past 50 years."

MacCracken ES-22, Executive Summary, Page 19, Lines 377-379: This statement seems to be giving equal credence to data sets that do not merit equal credence. Science is not a vote—it relies on close examination and consideration of the strengths and weaknesses of various lines of evidence, and in this case it needs to be made clear that the data sets that account for the biases and inadequacies that have been identified now give a consistent result between the surface and the troposphere. This statement, near the end of the Executive Summary, should not be based on the situation before the extensive analysis of this report, but on the situation after all this work.

Response: The Report makes no statements about the relative credibility of the various data sets, so the claim here that there are some data sets that have been used that "do not merit equal credence" is not supported in the Report. Further, "majority" is used here simply in its usual English language sense – not to imply that there has been some sort of vote.

MacCracken ES-23, Page 20, Lines 408-410: How is this known to the level of definitiveness of "does not"? I don't know of studies that have really looked closely at how the spatial heterogeneity of forcings has been looked at in terms of how the atmospheric circulation might be affected and how such changes might be seen by the time varying observation network, etc. Given the large areas where there is low correlation between surface and tropospheric fluctuations on a short-term basis, how can this statement be made with such certainty?

Response: This statement comes directly from Chapter 5.

MacCracken ES-24, Page 21, Line 423: Change "analyses" to "analyzes" and delete "globally"—maybe substitute 'from around the world"

Response: Done

MacCracken ES-25, Page 22, Line 445: Again, it is not good scientific practice to be summarizing the observational results as sort of a vote, especially when not considering the relative credibility of the datasets that are included. This report has made a major advance in understanding, and this needs to be indicated.

Response: The Report makes no statements about the relative credibility of the various data sets, so the claim here that there are some data sets that have been used that "do not merit equal credence" is not supported in the Report. Further, "majority" is used here simply in its usual English language sense – not to imply that there has been some sort of vote.

MacCracken ES-26, Page 27, Line 524: Reference should be made to the appendix on statistical techniques at this point, particularly making the point to consider the limitations and pitfalls of "linear trend analysis" (something that, for example, Pat Michaels' projections of temperature trend based on fluctuations over the past 30-35 years fails to account for). At the very least, connect this statement to footnote 4, where reference is made to the appendix and the point is made that linear trend analysis is not always the best approach (especially when a signal is rising out of a fluctuating baseline).

Response: Reference added; see Appendix A for more information on linear trends.

Robock ES-1, p. 8. Table 1: "Increased loading of sulfate (SO4) aerosol" should be changed to "Increased loading of sulfate (SO4) aerosol in the troposphere." As written, it is unclear, because sulfate in the stratosphere had different effects.

Response: Done.

Singer ES-1, P2 line 44: Misleading. The cooling trend until about 1976 is not explained by anthropogenic forcing (see, e.g., IPCC-TAR, SPM)

Response: This is incorrect. Our understanding includes internally generated variability; which is more than sufficient to explain this cooling even in the presence of an externally forced warming trend. The text does not say "anthropogenic".

Singer ES-2, P2 line 46-47: One should explain that 1950 is during a cool period; therefore a temp increase since 1950 is obvious. [*Singer*]

Response: The text says "from the late 1950s", not from 1950. The claim that the time around 1950 was a cool period cannot be supported. The text here refers to low and mid tropospheric temperatures, and we have no reliable data for these atmospheric layers prior to 1958.

Singer ES-3, P3 line 50: A crucial result that speaks against both anthropogenic warming and the results of GH models [*Singer*]

Response: The reviewer has taken this phrase out of context. To assess this result one must examine the totality of relevant evidence. This is what the Report does, and, by so doing, reaches quite a different conclusion.

Singer ES-4, P3 line 56-57: The claim that "there is no inconsistency between models and observations at the global scale," is an artful evasion of the fact that there IS inconsistency when the time scale is confined to 1979 to 2005 (after the major climate shift of 1976-78) – and especially in the Tropics, the region most relevant to detecting any human influence. [See my comment on Chap 5, p4, line 82-83] [*Singer*]

Response: The text is not meant to be evasive. The text has been revised and now reads:

"The most recent climate model simulations give a range of results for changes in global-average temperature. Some models show more warming in the troposphere than at the surface, while a slightly smaller number of simulations show the opposite behavior. There is no fundamental inconsistency between these model results and observations at the global scale."

The inconsistency in the tropics is clearly stated, and the implications explained and justified. The revised text reads:

"In the tropics, the agreement between models and observations depends on the time scale considered. For month-to-month and year-to-year variations, models and observations both show amplification (i.e., the month-to-month and year-to-year variations are larger aloft than at the surface). This is a consequence of relatively simple physics, the effects of the release of latent heat as air rises and condenses in clouds. The magnitude of this amplification is very similar in models and observations. On decadal and longer time scales, however, while almost all model simulations show greater warming aloft (reflecting the same physical processes that operate on the monthly and annual time scales), most observations show greater warming at the surface.

These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report favors the second explanation. Reconciliation of observational uncertainty is a pre-requisite for resolving to what extent model error exists."

Singer ES-5, P3 line 59-62: This conclusion is contradicted by the data (see Chapter 5)

Response: There is a large literature supporting this conclusion, cited in Chapter 5. The reviewer cites no evidence to support his claim.

Singer ES-6, P3 line 64-65: This conclusion is contradicted by the data (see Chapter 5)

Response: There is a large literature supporting this conclusion, cited in Chapter 5. The reviewer cites no evidence to support his claim.

Singer ES-7, P3 line 68-70: This crucial result speaks against any significant human climate effect. [*Singer*]

Response: The reviewer has taken this phrase out of context. To assess this result one must examine the totality of relevant evidence. This is what the Report does, and, by so doing, reaches quite a different conclusion.

Singer ES-8, P4 line 72-77: Statement is obscure and ignores Tropics [*Singer*]

Response: The statement here is unarguably correct. There have been changes in our understanding (many, many new publications), and the NRC and IPCC statements require modification because of this. The statement refers to both the global-means and the tropics (with this now clarified by a text addition).

Singer ES-9, P4 line 89-90: Dissimulation; it slides over the result of line 87 [*Singer*]

Response: It cannot be denied that this is a complex issue, as stated here. This is not dissimulation, but a fact.

Singer ES-10, P5 line 93-94: The assertion of "no conflict" is unwarranted by observations [*Singer*]

Response: The text has been expanded to explain why there is no conflict. The revised text is:

"Over the period since 1979, for global-average temperatures, the range of recent model simulations is almost evenly divided among those that show a greater global-average warming trend at the surface and others that show a greater warming trend aloft. The range of model results for global average temperature reflects the influence of the mid- to high-latitudes where amplification results vary considerably between models. Given the range of model results and the overlap between them and the available observations, there is no conflict between observed changes and the results from climate models."

Note that this refers to global-mean data.

Singer ES-11, P5 line 101: Crucial for showing anthropogenic warming to be minor. [*Singer*]

Response: The text at issue here is "On decadal and longer time scales, however, while almost all model simulations show greater warming aloft, most observations show greater warming at the surface", referring to the tropics. As noted above, to assess this result one must examine the totality of relevant evidence. This is what the Report does, and, by so doing, reaches quite a different conclusion, viz.

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report favors the second explanation. Reconciliation of observational uncertainty is a pre-requisite for resolving to what extent model error exists."

Singer ES-12, P5 line 103-107: The simplest explanation is one not mentioned. Namely: amplification on monthly and inter-annual time scales confirms merely that a moist convective atmosphere is in accord with theory; however, the absence of such amplification on a decadal time scale shows that the models overestimate GH warming.

The alternative explanation given here, which blames any disagreement between data and model results on errors and uncertainties, is unsatisfactory. It appears to be more ideological than scientific. [*Singer*]

Response: As there is no physical reason to expect amplification to depend on time scale, the conclusion that the problem rests with at least some of the observational data is quite logical. Note that the conclusions here are both cautious and fully justified. The following text extract explains this:

"This inconsistency between model results and observations could arise due to errors common to all models; due to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or due to a combination of these factors. The new evidence in this Report – model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming (such as the increasing height of the tropopause) – favors the second explanation."

Singer ES-13, P13 line 262-263: Analysts (Free et al, Thorne et al) who publish radiosonde data do not accept this statement about complete uncertainty. [*Singer*]

Response: The text does not say "complete uncertainty", but notes the unarguably correct point that there is "considerable disagreement between tropospheric data sets". Both Thorne (who is one of the author team) and Free (who participated in some aspects of the Report development) agree with this.

Singer ES-14, P13 line 267-268: There is no such thing as "tropospheric temp" unless one first defines a weighting function (with altitude) [*Singer*]

Response: "tropospheric temperature" is used on line 265 of the original text. Weighting functions are given in Chapter 2 (Fig. 2.2), and the terminology is given in the Preface.

Singer ES-15, P13 line 275: What is the evidence for "spurious cooling" in the tropical troposphere? [*Singer*]

Response: Publications by Sherwood et al. and Randel and Wu, cited in (e.g.) Chapter 5, provide the evidence.

Singer ES-16, P13 line 278-281: This paragraph may be out of date. The cause of the difference between the RSS and UAH-v5.2 values has not yet been established [*Singer*]

Response: This paragraph is not out of date. The developers of these data sets are part of the author team. It is true that the differences between the latest RSS and UAH data sets have not been fully resolved, but the contributing factors are known and are described in the Report.

Singer ES-17, P18 line 357-364: We don't see any evidence for the claimed anthropogenic influence in the climate record. The "fingerprint" results claimed in IPCC-SAR have been discredited. [*Singer*]

Response: The reviewer does not identify "we"? There is a vast literature on fingerprint studies, much of which is reviewed in Chapter 5. None of this literature has been discredited.

Singer ES-18, P19 line 371-379: While global-mean results may not show discrepancies, the more relevant tropical data show significant differences between surface and tropospheric trends, which indicate that anthropogenic effects are minor. [*Singer*]

Response: That the surface and the troposphere show different trends is not disputed. The issue is whether these differences are in accord with physical understanding as encapsulated in model simulations. There are model/observed data differences here; but the conclusion of the expert group of authors of the Report is that:

"This inconsistency between model results and observations could arise due to errors common to all models; due to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or due to a combination of these factors. The new evidence in this Report – model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming (such as the increasing height of the tropopause) – favors the second explanation."

This is a carefully worded and fully justified conclusion.

Singer ES-19, P20 line 391-397: The simplest explanation is one not mentioned. Namely: amplification on monthly and inter-annual time scales merely confirms that a moist convective atmosphere is in accord with theory; however, the absence of such amplification on a <u>decadal</u> time scale shows that the models overestimate GH warming.

The alternative explanation given here, which blames any disagreement between data and model results on errors and uncertainties, is unsatisfactory. It appears to be more ideological than scientific. [*Singer*]

Response: The point here is that there is no physical reason to expect amplification to depend on time scale. As there is no physical reason to expect amplification to depend on time scale, the conclusion that the problem rests with at least some of the observational data is quite logical. Note that the conclusions here are both cautious and fully justified. The following text extract explains this:

"This inconsistency between model results and observations could arise due to errors common to all models; due to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or due to a combination of these factors. The new evidence in this Report – model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming (such as the increasing height of the tropopause) – favors the second explanation."

Singer ES-20, P20 line 408-410 Spatially heterogeneous forcings in climate modals may not influence "amplification" (i.e., ratio of troposphere to surface trends); but anyway, it is not replicated in observed geographic temp changes. [*Singer*]

Response: The reviewer appears to be confusing signal and noise here.

Singer ES-21, P24 Line 464: The most relevant figure for judging human influence is Fig. 4G. But this figure is drawn in a misleading way – as can be seen by comparing with the original Fig. 5.4G (chapter 5, page 54, line 1027). [*Singer*]

Response: It is correct that this Figure could be misleading, and the text has been reworded to avoid this possibility. However, the results in Fig. 5.4G are identical to those in Fig. 4G – in fact, the present Figure is more accurate in defining the model range because the data are not binned. Binning actually makes the overlap appear greater in Fig. 5.4G than here. We note that is wrong to focus on a single set of results, as the reviewer is doing here, and to ignore the discussion of these results given in the text.

The revised text states:

"For global averages (Fig. 3), models and observations generally show overlapping rectangles. A potentially serious inconsistency, however, has been identified in the tropics. Figure 4G shows that the lower troposphere warms more rapidly than the surface in almost all model simulations, while, in the majority of observed data sets, the surface has warmed more rapidly than the lower troposphere. In fact, the nature of this discrepancy is not fully captured in Fig. 4G as the models that show best agreement with the observations are those that have the lowest (and probably unrealistic) amounts of warming (see Chapter 5, Fig. 5.6C). On the other hand, as noted above, the rectangles do not express the full range of uncertainty, as they do not account for uncertainties in the individual model or observed data trends.

The potential discrepancy identified here is a different way of expressing the amplification discrepancy described in Section 4, item (5) above. It may arise from errors that are common to all models, from errors in the observational data sets, or from a combination of these factors. The second explanation is favored, but the issue is still open."

Trenberth GEN-1& ES-1, There is, in my view, too much emphasis on linear trends and nowhere a clear statement that linear trends are not a good fit to the data (the appendix in fact claims otherwise but gives examples chosen to make this so). This is especially so in the stratosphere with the volcanic perturbations, in the tropics with ENSO, and it is also true especially for longer intervals such as 1958 to 2004 where the trends in troposphere and stratosphere are very different after 1976 from those before then. As a result, sampling issues and sensitivity to small differences at start and end of series is real. It makes a big difference whether the trends begin in 1976 or 1979. This becomes a major issue for comparisons with model results that do not have such a shift or ENSOs in the right sequence and magnitude. Error bars are missing in many places, including 2 figures in exec summary.

Response: It is true that a linear trend has disadvantages when the behavior of a time series is not expected to be linear. Nevertheless, there is no single metric that can replace the trend value, and the reviewer has offered no constructive suggestion in this regard. The texts of the Exec. Summary and the Statistical Appendix contain many statements recognizing this obvious deficiency and explaining why, nevertheless, the trend is still a useful descriptor of a gross characteristic of a time series. Here are some examples:

Statistical Appendix

"Over the present study period (1958 onwards), the expected changes due to anthropogenic effects are expected to be approximately linear. In some cases, natural factors have caused substantial deviations from linearity (see, e.g., the lower stratospheric changes in Fig. 1B), but the linear trend still provides a simple way of characterizing the overall change and of quantifying its magnitude.

Alternatively, there may be some physical process that causes a rapid switch or change from one mode of behavior to another. In such a case the overall behavior might best be described as a linear trend to the changepoint, a step change at this point, followed by a second linear trend portion. Tropospheric temperatures from radiosondes show this type of behavior, with an apparent step increase in temperature occurring around 1976 (see Chapter 3, Fig. 3.2a).

Step changes can lead to apparently contradictory results. For example, a data set that shows an initial cooling trend, followed by a large upward step, followed by a renewed cooling trend could have an overall warming trend. To state simply that the data showed overall warming would misrepresent the true underlying behavior. A linear trend may therefore be deceptive if the trend number is given in isolation, removed from the original data. Nevertheless, used appropriately, linear trends provide the simplest and most convenient way to describe the overall change over time in a data set, and are widely used."

Executive Summary

"Many of the results in this Report (and here in the Executive Summary) are quantified in terms of linear trends, i.e., by the value of the slope of a straight line that is fitted to the data. A simple straight line is not always the best way to describe temperature data, so a linear trend value may be deceptive if the trend number is given in isolation, removed from the original data. Nevertheless, used appropriately, linear trends provide the simplest and most convenient way to describe the overall change over time in a data set, and are widely used. For a more detailed discussion, see the Appendix."

It should be clear from these extracts that we are well aware of the issues on trends raised by the reviewer, and that we have discussed them openly and in a balanced way.

Another point the reviewer should realize is that the Executive Summary is just that, a summary of material presented elsewhere in the Report. As such, if linear trends are used as a descriptor elsewhere in the Report (which is indeed the case) then these results must be presented, in this form, in the Exec. Summary. The decision to use linear trends as a primary descriptor was not taken lightly, and was made jointly by the whole expert author team.

To suggest, furthermore, that "the appendix in fact claims (that linear trends are useful) but gives examples chosen to make this so", is incorrect. The Appendix gives a range of representative examples, including the time series for stratospheric temperature changes that the reviewer lists as a contrary example. Time series that show the apparent step change in tropospheric temperature are illustrated in the Exec. Summary.

The reviewer also claims that error bars (or confidence intervals) should be given in various Figures. Again, this was a decision not taken lightly, and made jointly by the whole expert author team. A number of factors were considered. Here are some points noted in the Statistical Appendix:

"While it may be common practice to use error bars to illustrate C.I.s for trends of individual time series, when the primary concern (as it is in many parts of this Report) is the comparison of trends, individual C.I.s can be misleading. A clear example of this is given in Fig. 4 (based on information in Figs. 2 and 3). Individual C.I.s for the three MSU T2 series overlap, but the C.I.s for the difference series show that there are highly significant differences between the three data sets. Because of this, in some cases in this Report, where it might seem that error bars should be given, we consider the disadvantage of their possible misinterpretation to outweigh their potential usefulness. Individual C.I.s for all trends are, however, given in Tables 3.2, 3.3, 3.4 and 3.5 of Chapter 3; and we also express individual trend uncertainties through the use of significance levels. As noted in Section (9) below, there are other reasons why error bars can be misleading."

Note that the C.I. information is given in the Report in all cases. In some cases, as explained in the quoted text, we considered it best not to give such information the prominence it would receive if illustrated graphically. In other cases we considered that graphical representation would make the Figures messy and more difficult to digest for our intended lay audience. (NOTE: See also the response to Trenberth GEN-1)

Trenberth GEN-2 & ES-2, The summary is also deficient on issues of land vs. ocean. This is related to max vs. min changes and how those would be seen in the troposphere vs. surface; i.e., expect max. to be seen from deeper mixing but not min. Surface changes are much larger over land than ocean and muted in troposphere (see chapter 1), but in troposphere changes are more zonally symmetric and larger over oceans than at surface. This relates to the issue of where and how the surface can increase more than troposphere. Chapter 1 makes the point that there are really not good reasons why these should be strongly linked, yet much of the report misses this point. In chapter 4, where huge differences occur over Africa in T2LT, it does not come to grips with this issue (note also that the diurnal cycle of surface temperature is order 30ºC over the Sahara).

Response: These are criticisms of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. (NOTE: See also the response to Trenberth GEN-2)

Trenberth GEN-3 & ES-3, There is little discussion of issues on urban heat island effects etc. It is briefly mentioned in chapter 4 but inadequate. It is a complex issue and the effects are real, so it while one can say that the global mean is OK because it is not contaminated by unrepresentative very local UHI effects, those changes are real. This is not dealt with in the report. There is now quite a bit of literature related to the "weekend effect" whereby statistics differ by weekday and presumably relate to aerosols and interactions with clouds.

Response: These are criticisms of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. (NOTE: See also the response to Trenberth GEN-3)

Trenberth GEN-4 & ES-4, This is supposed to be an assessment. It falls short especially in chapters 2 and 3, where it should refer ahead to chapter 4. In chapter 4 there is some useful assessment but it falls back on "all datasets are equal" in spite of strong evidence otherwise. This is a major limitation of the report.

Response: These are criticisms of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. (NOTE: See also the response to Trenberth GEN-4)

Trenberth GEN-5 & ES-5, The report pretends that the radiosondes are global, and insufficient accounting is made of the fact that they are not close to that. Zonal means are also biased by land distribution. Errors of 0.2ºC can occur in global means from the distribution of sondes (Hurrell et al 2000) although effects on trends seems to be modest (0.03ºC decade-1) this is not guaranteed.

Response: These are criticisms of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. (NOTE: See also the response to Trenberth GEN-5)

Trenberth GEN-6 & ES-6, Very little account is taken of the works that show major shortcomings in the radiosondes (Sherwood et al 2005, Randel and Wu 2005) in chapters 2 and 3. They are discussed in chapter 4 and conclusions drawn that sondes are biased cold but then this is ignored elsewhere. There is no sound basis for believing the profiles in Fig 3.7, for instance.

Response: These are criticisms of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. (NOTE: See also the response to Trenberth GEN-6)

Trenberth GEN-7 & ES-7, The UAH record has once again been revised but the new T2LT values are at odds with surface temperature trends. Chapter 4 falls short in not presenting maps of this difference. Accordingly, this dataset ought to also be discounted. Given the UAH algorithm that is designed to minimize trends, this dataset ought to be given lower weight, but no commentary appears on this issue.

Response: This is a criticism of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. Note that the author team did not think, on the basis of published or "in press" research, that is was possible to assign relative credibility levels to individual data sets. (NOTE: See also the response to Trenberth GEN-7)

Trenberth GEN-8 & ES-8, The reanalyses are not considered seriously for no good reason other than opinions that are baseless. For NCEP, these fears are well grounded and some references are given but for ERA-40, major efforts went into bias correction and a major advantage of ERA-40 is that all observations were assimilated at the exact time they were made, overcoming diurnal cycle issues, a major advantage relative to all the other datasets. The bias corrections to the sondes in ERA-40 likely makes them better than the sonde records themselves. Nevertheless the reanalyses are seriously flawed and have to be used with care (see Trenberth and Smith 2005; given below under chapter 1).

Response: These are criticisms of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. (NOTE: See also the response to Trenberth GEN-8)

Trenberth GEN-9 & ES-9, In places the document is unduly dumbed down to the point where the text is not factual. Why is it necessary to have an appendix that is dominated by basic statistical text book material?

Response: This is a criticism of individual Chapters. The Executive Summary can only summarize what is in the individual Chapters, and a joint author decision was taken to include in the Exec. Summary only those items identified as key points in the individual Chapters. The fact that the Report is meant to be read by an audience with widely ranging backgrounds required that some material be presented in simple terms – the pejorative "dumbed down" is not appropriate. If there are factual errors as a result of attempts to explain concepts in simple terms, then a more constructive criticism would have been to point out the specific cases. The reasons for including a comprehensive Statistical Appendix have been outlined in the specific responses to comments on this Appendix. . (NOTE: See also the response to Trenberth GEN-9)

Trenberth GEN-10 & ES-10, What is the vintage of this report? It mostly does not include papers submitted or in press but there are exceptions? It would help to make clear the time frame and cut off for considering literature.

Response: This is not relevant to the Executive Summary. A response to this question is now given in the Preface. (NOTE: See also the response to Trenberth GEN-10)

Trenberth GEN-11 & ES-11, The report is very long, not generally readable as a result, and contains a lot (far too much) basic tutorial material.

Response: This is not directly relevant to the Executive Summary. The fact that the Report is meant to be read by an audience with widely ranging backgrounds required that some material be presented in simple terms. (NOTE: See also the response to Trenberth GEN-11)

Winstanley ES-1, Page 2, Lines 25-26; and **Winstanley CH5-1:** In the Executive Summary, the focus of the report is broadened from that stated in the Preface (to understand the causes of differences between independently produced data sets) to also include understanding of the causes of the temperature changes themselves, which are addressed in Chapter 5. Whereas much attention is given in the report to addressing the strengths and weaknesses of different observed temperature trends, little attention is paid to documenting the strengths and weaknesses of the models whose outputs are compared with observations. The models also are used to understand causes of the differences among the observed trends and to understand the causes of the trends. Since there is considerable reliance on models in comparing observations with theoretical expectations and in evaluating the causes of observed changes, similar critique of the strengths and weaknesses of models should be included in the report as is given to the critique of the strengths and weaknesses of observations.

Response: The statement here properly reflects the charge defined by the original questions, and the format of the Report as a whole and the Executive Summary in particular follows this charge. Model issues are discussed at length in Chapter 5. (NOTE: See also the response to Winstanley CH5-1)

Winstanley ES-2, An Executive Summary often is used as a stand-alone document and should provide all the necessary information for those who use it as such. It would be improved by providing at the start the context for the report, i.e., the purpose and scope of the report, the controversy and uncertainties in scientific understanding that gave rise to the report, and the origin of the questions. As the Executive Summary should provide information contained only in the report itself, the report would benefit from adding an Introduction that includes such information. Currently, there is no Introduction to the report. Information that typically is included in an Introduction is incorporated in the Preface, so the contents of the Preface also should be reviewed once an Introduction is incorporated.

Response: As the reviewer states, this background material is given in the Preface – and so it would not be appropriate to duplicate it in the Executive Summary.

Reviewer (Winstanley) comment (cont): The Executive Summary and a new Introduction also should explain the importance to the climate system and decision makers of vertical temperature profiles in the atmosphere; for example, actual temperatures and variations in temperature at the Earth's surface and in the atmosphere, and the rate of change of temperature with height (lapse rate) influence the stability of the atmosphere, convection, and precipitation. It is important to understand spatial and temporal variations in lapse rates to understand the climate system and climate change. In understanding climate change it is important to be able to determine the causes of observed changes in the climate system and to establish data accuracy and consistency between model simulations and observations. If there are inconsistencies among different observational data sets, among model simulations, and among observational data sets and model simulations, these reduce our confidence in understanding the climate system and in future climate scenarios projected by these models. The issue of uncertainty and/or confidence should be addressed explicitly.

Response: These issues are addressed throughout the main body of the Report, and summarized here in the Executive Summary.

Reviewer (Winstanley) comment (cont): The Executive Summary also would benefit from a clear summary of the new understanding that this report brings to addressing the contentious differences between independently produced data sets of atmospheric temperature trends from the surface through the lower stratosphere reported in earlier reports, and the causes of the changes and differences. Much of the needed information is included in the current draft (and the Preface), but in a format and location that would make it difficult for decision makers and non-scientists to discern clearly what has been resolved since the NRC and IPCC reports, and what remains unresolved.

Response: As the reviewer states, "Much of the needed information is included in the current draft" (my emphasis). It is encouraging that the reviewer was able to see this.

Winstanley ES-3, Page 3, lines 54-62: The main findings reported here are that 1) there is no inconsistency between models and observations at the global scale, 2) there is clear evidence of human influences on the climate system, and 3) the observed patterns of change cannot be explained by natural processes alone. Points 2 and 3 add nothing new and provide nothing of relevance to this report and should be deleted as "important new results" in the Executive Summary and **Chapter 5**.

Response: The new results supporting these statements are given in Chapter 5.

Reviewer (Winstanley) comment (cont): Point 1 is a gross overgeneralization and a more carefully crafted statement of our understanding of the strengths and limitations of existing observational data sets and models would be more appropriate.

Response: The text has been revised to read:

"The most recent climate model simulations give a range of results for changes in global-average temperature. Some models show more warming in the troposphere than at the surface, while a slightly smaller number of simulations show the opposite behavior. There is no fundamental inconsistency between these model results and observations at the global scale."

Reviewer (Winstanley) comment (cont): In **Chapter 6**, for example (page 2, lines 49-51), it states that "There remain differences between independently estimated temperature trends for the surface, troposphere and lower stratosphere, and differences between the observed changes and model simulations, that are, as yet, not fully understood."

Response: This information is also given in the Executive Summary. The observational differences (and similarities) are illustrated in Fig. 1 and described in the accompanying text. Model/observed differences are shown in Fig. 3 and 4. Chapter 6 gives recommendations for improving our understanding of the reasons for differences between observational data sets that purport to measure the same thing, and these recommendations are repeated in the Executive Summary.

Reviewer (Winstanley) comment (cont): A key finding in **Chapter 4** (p. 3, lines 70-75) is that uncertainties in tropospheric data are the main reason why it is difficult to determine whether the troposphere has warmed more or less than the surface. The difference in trend between the lower troposphere and mid-upper troposphere is not well characterized by the existing data (p. 38, lines 808-809).

Response: This is stated in the Executive Summary as:

"Tropospheric temperatures: All data sets show that the global- and tropical-average troposphere has warmed from 1958 to the present, with the warming in the troposphere being slightly more than at the surface. For changes from 1979, due to the considerable disagreements between tropospheric data sets, it is not clear whether the troposphere has warmed more than or less than the surface."

with the clear implication that …

"Errors in observed temperature trend differences between the surface and the troposphere are more likely to come from errors in tropospheric data than from errors in surface data."

Reviewer (Winstanley) comment (cont): Chapter 4 also recognizes that structural uncertainties are difficult to assess in an absolute sense (p.40, lines 848-849) and there may be systematic biases that remain after appropriate homogenization methods have been applied (p. 5, lines 117-120).

Response: See (e.g.) Exec. Summary lines original text lines 270, 271, and lines 278, 279.

Reviewer (Winstanley) comment (cont): Chapter 3 recognizes considerable disagreement among tropospheric (p. 2, lines 58-61) and stratospheric (p. 51, lines 979-980) datasets.

Response: The observational differences (and similarities) are illustrated in Fig. 1 and described in the accompanying text. See also Exec. Summary original text lines 270, 271, and lines 278, 279.

Reviewer (Winstanley) comment (cont): If it is a goal of the report to address the causes of the temperature changes in the atmosphere, the report should do this and summarize the findings in the Executive Summary. Stating that there is "…clear evidence of human influences on the climate system …"(lines 60-62) and that "The observed patterns of change cannot be explained by natural processes alone, nor by the effects of short-lived species" (lines 64-65) does not specifically address the causes of observed vertical temperature changes or the roles of external forcings, internal forcings, and internal variability.

Response: The Exec. Summary is meant to summarize information given in the individual Chapters. The above statements come directly from Chapter 5. The text has been expanded to explain what is meant by "natural factors", viz.

"Natural factors (external forcing agents like volcanic eruptions and solar variability and/or internally generated variability) have influenced surface and atmospheric temperatures, but cannot fully explain their changes over the past 50 years."

Reviewer (Winstanley) comment (cont): The extent to which it is known that internal variations of the climate system are represented reliably in current climate models and may be a contributing cause of observed climate changes regionally and globally, including vertical temperature changes, should be addressed in the report.

Response: The text has been expanded to explain what is meant by "natural factors", which includes internally generated variability. The representation of internal variability in a model cannot emulate what has occurred in the real world – essentially each model realization is a separate universe with its own chaotic weather variability and associated lower frequency climatic variability. This is explained further in the Statistical Appendix, and also in Chapter 5. It should be noted that the statistical character of internally generated variability in climate models is, in most models, similar to that in the real world.

Winstanley ES-4 and **Winstanley CH5-2:** Due to the fundamental climatological importance of lapse rates, the Executive Summary should contain a summary of what we know about lapse rates regionally and globally and how well regional and global climate models simulate actual temperatures and lapse rates.

Response: This is precisely what the Exec. Summary does, based on material in the individual Chapters. There is little that is said in the Report at the regional level because signal-to-noise ratio problems preclude separation of signal from noise.

Reviewer (Winstanley) comment (cont): The draft Executive Summary says nothing about the fundamental subject of lapse rates. **Chapter 2, page 30, lines 541-543** state that explaining atmospheric and surface trends demands relative accuracies of a few hundredths of a degree per decade in global time series of both surface and upper-air observations and **Chapter 3, Section 7.2**, contains limited information on lapse rates. **Chapter 3, lines 986-988** acknowledges that "Most of the observational work to date has not examined lapse rates themselves, but instead has used an approximation in the form of a vertical temperature difference."

Response: These comments refer to individual Chapters, not specifically to the Exec. Summary. The Exec. Summary can only include material given in the individual Chapters.

Reviewer (Winstanley) comment (cont): In **Chapter 3**, with a summary in the Executive Summary, there needs to be discussion of the implications for climate studies of not reporting actual temperatures and lapse rates, and not comparing observed lapse rates with modeled lapse rates.

Response: All available observational studies of lapse rates per se are summarized in the individual Chapters. Model/observed comparisons dealing directly with lapse rates are covered in Chapter 5. There is very little published literature on either of these subjects.

Reviewer (Winstanley) comment (cont): Also, there should be discussion of the implications for the questions posed of using a surrogate lapse-rate approximation in climate studies. As a focus of the report is to compare observed and modeled vertical temperature variations, **Chapter 5** should include a statement about the accuracy of models in simulating decadal lapse rates, as well as changes in lapse rates.

Response: Model/observed comparisons dealing directly with lapse rates are covered in Chapter 5. There is very little published literature on this subject.

Reviewer (Winstanley) comment (cont): The global climate system is a composite of regional climates and more discussion of regional lapse rates and changes in lapse rates would give readers more confidence that global analyses represent the composite of regional conditions accurately. That comprehensive regional-scale analyses of lapse rates have not been conducted is recognized in **Chapter 5, lines 862-866**.

Response: It is correct that there is little that is said in the Report at the regional level, reflecting the paucity of literature dealing with lapse rate changes at the regional level. This is at least partly because signal-to-noise ratio problems preclude separation of signal from noise. Further, defining changes in lapse rates per se is much more difficult than defining changes at a particular level – so there is a paucity of suitably accurate data. This is why we have resorted to lapse rate proxies, as in Figures 3 and 4 of the Exec. Summary and the corresponding material in Chapter 5.

Reviewer (Winstanley) comment (cont): The Executive Summary should incorporate recognition of the importance of comprehensive regional analyses of lapse rates and state that they have not been conducted, if this is an accurate statement.

Response: As this point is not made in Chapter 6, it cannot be made here.

Reviewer (Winstanley) comment (cont): The report also should discuss the implications for the climate system (e.g., stability and precipitation) of reported spatial and temporal variations in vertical temperature differences and lapse rates.

Response: These are interesting issues, but, as explained in the Preface and as should be clear from the specific questions that define the scope of the Report, these issues do not fall within the charge set for this Report.

Winstanley ES-5 and **Winstanley CH5-3:** All major climate reports (e.g., IPCC, NRC, CCSP) adopt the approach of examining only temperature differences, either from one time period to another or between the surface and some height above the Earth's surface. This approach, adopted in reporting both observed temperature changes and modeled temperature changes, excludes explicit reporting of actual temperatures. A differential approach is appropriate in addressing many aspects of climate change, but also has limitations, which need to be addressed.

Response: The reviewer does not state what the specific limitations are, so no response is possible.

Reviewer (Winstanley) comment (cont): Particularly when discussing lapse rates or vertical temperature differences, actual temperatures and changes in actual temperatures are of great importance in evaluating the stability of the atmosphere and precipitation. By focusing only on temperature differences and avoiding actual temperatures conceals some important issues relating to model limitations, which are important in comparing differences between observed temperature changes and modeled temperature changes, and in evaluating the causes of temperature changes.

Response: The present Report is not concerned specifically with evaluating climate models, which is the focus of another CCSP Report in preparation. It is noted that analyses of other variables would be useful, but there are no such analyses currently that are of direct relevance to the charge for this Report. The relevant text in the revised Executive Summary (reflecting Chapter 6) is:

"Efforts should be made to develop new or reprocess existing data to create climate quality data sets10 for a range of variables other than temperature (e.g. atmospheric water vapor content, ocean heat content, the height of the tropopause, winds and clouds, radiative fluxes, and cryospheric changes). These data sets should subsequently be compared with each other and with temperature data to determine whether they are consistent with our physical understanding. It is important to create several independent estimates for each variable in order to assess the magnitude of construction uncertainties."

Reviewer (Winstanley) comment (cont): Kunkel *et al.* ("Can CGCMs simulate the Twentieth Century "Warming Hole" in the central United States?", in press, *Journal of Climate*, and attached with these comments) show major differences between the observed evolution of mean annual 20th Century temperature in Central North America (CNA) and mean annual temperature simulated by global climate models. There are significant differences between the observed and modeled temperature changes, and large differences between observed and modeled temperatures. The models simulate CNA mean annual temperature to an accuracy of only +/- 3°C. This raises the question as to the credibility of models in simulating regional changes in temperature of a few tenths of a degree when the accuracy of the models in simulating mean annual temperature of the region spans a range of 6°C.

Response: This is precisely why the present Report focuses on larger scales, averages over the tropics or the whole globe.

Reviewer (Winstanley) comment (cont): This is consistent with the finding in the Third Assessment Report of the Intergovernmental Panel on Climate Change that "Nearly all regional temperature biases are within the range of +/- 4°C " (Giorgi and Hewitson, 2001, p.592 and figure 10.2(a)).

Response: It is not clear what point is being made here by the reviewer. There are a number of studies that show that, at the spatial scales considered in the present Report, models give externally forced changes that are largely independent of errors in the baseline climate.

Reviewer (Winstanley) comment (cont): The draft Chapter 5 concludes that "When run with natural and human-caused forcings, model global-mean temperature trends for individual atmospheric layers are consistent with observations" (page 4, lines 79-80). The knowledge that there are large discrepancies between observed temperatures and modeled temperatures at the regional scale should be incorporated in Chapter 5 and the Executive Summary and the significance of these biases for global syntheses discussed.

Response: No -- this is precisely why the present Report focuses on larger scales, averages over the tropics or the whole globe. There are regional differences between model simulations and observations, but these tend to cancel out over larger areas. Signal-to-noise ratio problems are more serious at the regional level making the interpretation of regional results very difficult. Some of these issues (including the issue of poorly defined regional forcings) are discussed in Chapter 5.

Reviewer (Winstanley) comment (cont): Also, it must be asked what is the significance of these model limitations when evaluating lapse rates and changes in lapse rates? A bias in simulating surface temperature of +/- 3 °C must have major implications for understanding the stability of the atmosphere and precipitation regionally.

Response: These aspects are beyond the scope of the present Report. Further, even if stability and precipitation issues were within the scope of the Report, the reviewer gives no support for the claim that "A bias in simulating surface temperature of +/- 3 oC must have major implications for understanding the stability of the atmosphere and precipitation regionally" (my emphasis).

Reviewer (Winstanley) comment (cont): When climate models simulate mean annual temperature across a range of 6°C or more, how well do they simulate lapse rates and changes in lapse rates? Is it only surface temperature values that are inaccurate, or do the inaccuracies extend into the atmosphere above? What are the implications of such inaccuracies when evaluating the causes of observed temperature changes of a fraction of a degree? How accurately do global climate models simulate actual temperatures in other regions of the world and globally?

Response: The key issue is simulation of change. There are many studies that show that models can simulate changes even when there are biases in the base state.

Reviewer (Winstanley) comment (cont): What does it mean to conclude that "there is no inconsistency between models and observations at the global scale" when studying vertical variations in temperature and temperature changes?

Response: The text has been modified to state:

"The most recent climate model simulations give a range of results for changes in global-average temperature. Some models show more warming in the troposphere than at the surface, while a slightly smaller number of simulations show the opposite behavior. There is no fundamental inconsistency between these model results and observations at the global scale."

Here, fundamental inconsistency means a sufficient difference to cause us to suspect serious problems with either our physical understanding of the climate system or with current climate models. The above statement is the considered judgment of the expert author team.

Winstanley ES-6: Some findings in the Chapters are important and should be reported in the Executive Summary. For example:

Chapter 1, page 24, lines 477-479 recognize that "major relevant forcings are important to simulate 20th Century temperature…" and **Chapter 5, page 22, lines 466-468**, reports that it is difficult to answer "whether those forcings most important for understanding the differential warming problem are reliably represented [in current climate models]." **Chapter 6, page 14, lines 331-333,** state that "many of the forcings are not yet well quantified." **Chapter 1** (p. 15, lines 328-329) also recognizes that only in the past few years have climate models included time varying estimates of a subset of the forcings that affect the climate system. **Chapter 5** (p. 14, line 301) recognizes that most models undergo some form of "tuning". The fact that many climate forcings, and internal climate variations, are not well quantified and that most models are "tuned" leads one to question the veracity of the alleged lack of inconsistency between models and observations at the global scale (**Executive Summary**, p. 3, line 57).

Response: All findings that are judged to be key findings in the individual Chapters are given in the Executive Summary. The reviewer appears not to understand what is meant by "tuning" in the context of AOGCM development. In fact, "tuning" is not the correct word (a fault of the Chapter 5 authors), and the process neither considers nor does it affect simulated changes in climate.

Reviewer (Winstanley) comment (cont): Chapter 2, page 5, lines 123-125 state that all observations contain some errors and biases and **Chapter 2, page 2, lines 50-51**, states that measurements from all systems require adjustments and this report relies on adjusted datasets. There is a lack of traceable standards (**line 65**) and reference stations (**line 71**) and most observing systems have not retained complete metadata (**line 73**). Reanalysis trends are not always reliable (**Chapter 2, page 17, lines 348-350**).

Response: These statements are correct – but they are not relevant to the Exec. Summary, which clearly cannot repeat every point made in the individual Chapters. All findings that are judged to be key findings in the individual Chapters are given in the Executive Summary.

Reviewer (Winstanley) comment (cont): Chapter 2, page 2, lines 53-57, state that land-surface temperature records yield trends that are reasonably similar on large (e.g., continental) scales, that the ocean surface record suffers from more serious sampling problems and changes in observing practices, and that upper-air datasets likely give reliable indications of directions of change but some questions remain regarding the precision of measurements.

Response: These statements are correct – but they are not relevant to the Exec. Summary, which clearly cannot repeat every point made in the individual Chapters. All findings that are judged to be key findings in the individual Chapters are given in the Executive Summary.

Reviewer (Winstanley) comment (cont): Chapter 2, page 30, lines 532-536, state that most observing systems are generally able to quantify well the magnitude of change associated with shorter time scales, for longer time scales the observing systems face significant challenges.

Response: These statements are correct – but they are not relevant to the Exec. Summary, which clearly cannot repeat every point made in the individual Chapters. All findings that are judged to be key findings in the individual Chapters are given in the Executive Summary.

Winstanley, ES-7 and **Winstanley, CH5-4:** The discussion on models includes consideration of internal and external forcings as drivers of climate variations and change.

Response: There is no such thing as "internal forcings", so this is probably a typo by the reviewer. The standard distinction is between external forcing and internal variability.

Reviewer (Winstanley) comment (cont): There is no explicit recognition that natural internal variations of the climate system can bring about climate variations and change, and that internal variability needs to be considered as a factor when attributing causes of observed or modeled change.

Response: Internal variability is considered in numerous places in the Report. Virtually all D&A (detection and attribution) work assumes as a null hypothesis that changes are due solely to internal variability, and seeks to demonstrate the existence of external forcing effects by rejecting the null hypothesis. This is explained in Chapter 5, and also in the Statistical Appendix.

Reviewer (Winstanley) comment (cont): Kunkel *et al.* ("Can CGCMs simulate the Twentieth Century "Warming Hole" in the central United States?" in press, *Journal of Climate*, and attached to these comments) demonstrate that "…the warming hole is not a robust response of contemporary CGCMs to the estimated external forcings. A more likely explanation based on these models is that the observed warming hole involves external forcings combined with internal dynamic variability that is much larger than typically simulated."

Response: All variations are the combined effects of external forcing and internal variability. Even if climate models seriously underestimated internal variability for some limited spatial region, this would not affect any of the conclusions drawn in this Report.

Reviewer (Winstanley) comment (cont): The models produce substantially less variability of critical north Atlantic sea surface temperature than observed.

Response: This is largely correct, but only on time scales of decades or longer – and it is not true for all models.

Reviewer (Winstanley) comment (cont): From this, I conclude that the deficiencies of models to represent the internal dynamics of the climate system adequately can lead to erroneous attribution of climate variations and change to internal and external forcing factors.

Response: Again, this is not an issue that is of concern to the Exec. Summary, as it is not discussed in any individual Chapter. The relevance of a model's underestimate of decadal variability in the North Atlantic (or internal variability in general) to large scale simulations of changes in vertical temperature profile changes is not stated – indeed, the relevance is exceedingly unlikely. Further, D&A work accounts for uncertainty in the magnitude of internally generated variability.

Reviewer (Winstanley) comment (cont): Chapter 1, page 11, lines 230-231 recognizes that "unforced variability could be substantial" and states that "Chapter 5 provides more details on models and their limitations (see particularly Box 5.1 and 5.2)". However, **Chapter 5** does not incorporate recognition of the importance of internal variations in its discussions of the causes of reported changes in vertical temperature profiles. It should do so.

Response: D&A studies do account for internally generated variability, as noted above. However, there have been very few such studies of lapse rate changes per se.

Reviewer (Winstanley) comment (cont): Chapter 2, page 31, lines 556-560, recognizes the importance of internal modes of climate variability on regional scales and states that identifying the patterns and separating the influences of such modes from the warming signal is required.

Response: True, this would be an advantage since it would be a way to reduce noise and increase signal-to-noise ratios. However, standard optimized detection techniques do this already, albeit in a more sophisticated way. The reviewer seems to unaware of this.

Reviewer (Winstanley) comment (cont): The extent to which the report is able to identify the internal modes of climate behavior and separate these from internal and external forcings should be addressed in **Chapter 5** and summarized in the **Executive Summary**.

Response: Standard optimized detection techniques do this, as explained in Chapter 5 (and references cited therein).

Reviewer (Winstanley) comment (cont): Kunkel *et al.* ((“Can CGCMs simulate the Twentieth Century “Warming Hole” in the central United States?”, in press, *Journal of Climate,* and attached to these comments) demonstrate that model simulations, even simulations from the same model, are highly sensitive to initial conditions.

Response: This is well known, and a primary reason why we run multiple realizations with AOGCMs. This is explained in Chapter 5 and in the Statistical Appendix. The key point, however, is that it is the internally generated noise that is sensitive to initial conditions, not the externally forced signal.

Reviewer (Winstanley) comment (cont): Chapter 5 should incorporate this reference on page 14 and include as a Key Finding on model limitations (section to be added) the fact that noticeably different regional simulations of changes in atmospheric temperature profiles probably can result from model simulations that employ the same atmospheric model and the same climate forcings.

Response: These issues concern only the noise, not the signal. The work by Kunkel is not relevant to the Report, partly because it is regional, and partly because it does not address the topics that the present Report is concerned with.

Reviewer (Winstanley) comment (cont): Chapter 5, part of a much needed discussion on model limitations (parallel to the extensive discussions on the limitations of observational data throughout the draft report) should be discussion of the implications of a lack of explicit treatment of internal variability as a cause of climate variability and change and the lack of explicit treatment of model initialization.

Response: This is incorrect. We do consider internal variability. This is a stochastic component of AOGCM output that serves to obfuscate the underlying externally forced signal(s). Each realization from an AOGCM (with different initialization) has a different realization of internal variability (like a set of parallel universes, none of which is our "observed" universe). We average multiple runs to reduce this noise. Unfortunately, we cannot to this in the real world – we only have one of these. So we must use appropriate statistical methods to account for the noise. Internally generated variability in the observations is considered directly through these methods, and through the calculation of confidence intervals.

Reviewer (Winstanley) comment (cont): Also, different treatment of internal variations of the climate system and initial conditions should be included in the list on Page 7 of Chapter 5 of the reasons why climate simulations differ.

Response: This is covered in Chapter 5.

Reviewer (Winstanley) comment (cont): A key finding of **Chapter 5** should be that it is important to account for model uncertainty and limitations in comparisons between modeled and observed temperature changes. In the present draft, it is recognized only that observational uncertainty should be accounted for (page 6, lines 128-130).

Response: In fact, this point has been made in the modified version of Chapter 5, and the change reflected in the Exec Summary (see text under "OTHER FINDINGS") in Section 4.

<u>Chapter 1 Comments and Responses:</u>

MacCracken CH1-1, Chapter 1, Page 2, Line 49: Solar heat also warms the atmosphere—not just the surface. Also "properties" should be changed to "properties and processes"

Response: **Text revised and sense is incorporated.**

MacCracken CH1-2, Page 2, Line 53: Change "results" to "generally results" as there are inversions.

Response: **Text revised and sense is incorporated.**

MacCracken CH1-3, Page 2, Line 54: Change "of the troposphere" to "of the convectively mixed troposphere" to give an indication of what is defining the troposphere.

Response: **It is not always convectively mixed up to the tropopause.**

MacCracken CH1-4, Page 2, Lines 60-61: Variation also occurs due to the type of land, land cover, land use, etc.

Response: **Text revised and sense incorporated.**

MacCracken CH1-5, Page 2, Line 63: Insert to read "quickly smoothed out by the motions of the atmosphere, contributing …" to give an indication of how the smoothing occurs.

Response: **Done.**

MacCracken CH1-6, Page 3, Line 70: Change to "in winter over continents and sea ice/snow cover" as inversions also are important across the Arctic Ocean and in Antarctica. [On line 71 change "temperatures" to "temperature".]

Response: **Incorporated.**

MacCracken CH1-7, Page 3, Line 76: Change to read "due to temporal and spatial changes" as the changes are not only in space.

Response: **Done.**

MacCracken CH1-8, Chapter 1, Page 3, Line 130: Change "where" to "in which" for general readability.

Response: **Done.**

MacCracken CH1-9, Page 12, Line 246 (Table 1.1): While this table is based on the IPCC bar chart, a serious failing of the IPCC chart was its very limited indication of what was meant by the various levels of confidence. For example, no indication was really provided that a number of the forcings with "very low" confidence do not significantly contribute to the limits of our confidence in the results of the climate model projections. For example, the variations in solar radiation have been observed to be quite small (so one would think this forcing is reasonably well understood), and so this reference is to long-term changes where data are lacking, but for which the net effect is very likely small compared to the other forcings. Similarly, associating aircraft contrails with very low confidence is about what is almost certainly a quite small value. I would urge the authors to rework this column, indicating the likelihood that the uncertainty in understanding of the particular forcing would noticeably impact their overall analysis and findings.

Response: **Accepted. Global-mean forcing estimates from IPCC TAR have been added. This allows an easier discrimination of forcings that have larger values and a high degree of confidence from those that are estimated to have smaller values and/or a lower level of confidence. Table caption and text have also been revised to**

convey this information. Do not agree that long-term solar forcing is well understood just because the observed variations over the last 2.5 decades have been small. We do not wish to bring in climate response considerations here as this is a topic for chapter 5 and ES.

MacCracken CH1-10, Chapter 1, Page 13, Line 279: This homogenization by the atmosphere is itself a regional response—for example, while the sulfate aerosols create a regional forcing, the atmosphere would be responding to this in ways that affect the atmospheric circulation over a somewhat larger region, and even have some global influence. Indeed, one might call this homogenization a "climate regime shift" in response to the forcing, given how that term has earlier been used. The text included here should be modified to make it clear that this smoothing indeed generates a response, and over a somewhat larger region than the forcing.

Response: Accepted. Sentence is revised, in particular "homogenization" is replaced by "atmospheric processes and motions".

MacCracken CH1-11, Page 14, Lines 292-302: As the focus narrows to regions and finer scales (e.g., megalopolises), it is going to also be important to account for thermal emissions from energy use. Modeling studies were done in the 1970s by, for example, Washington and Chervin, looking at the impacts of thermal emissions resulting from the combustion of fossil fuels, and a recent relook at this question that I took makes it clear that these emissions could be adding a few tens of watts per square meter over reasonably sized regions. Thus, this paragraph needs to be changed to also mention the potential for influences from thermal emissions (and perhaps tying these to the discussion of potential biases affecting urban area surface observations).

Response: Agree, could be of relevance locally. But, robust estimates do not exist at present, especially for larger spatial scale contexts. It is also not clear that metrics have been devised to assess this forcing and thus we do not include it. However, they could merit consideration for climate change over a small urban region. Urban area surface observations are not within the scope of this chapter.

MacCracken CH1-12, Page 20, Line 412: Change "observations" to "specification" to really be clearer about what is being done.

Response: Done.

MacCracken CH1-13, Page 20, Lines 419-421: The warming during the first half of the 20^{th} century is only in part due to natural factors—there was a clear human influence on the Southern Hemisphere, and in the Northern Hemisphere, while there was a sort of counterbalancing of the GHG forcing and the aerosol forcing on the large scale, these were not spatially coincident forcings, so the "smoothing" by the atmosphere would be expected to cause some sort of response (e.g., a change in the atmospheric circulation—and is this a "regime shift"?). So, at the least, change the text to read "century mainly ascribed to natural forcings (primarily an absence of major volcanic eruptions and a natural increase in solar radiation), with unforced variations and adjustments to human-induced influences also playing some role, and the warming …"

Response: **Test revised to indicate that that the warming in the first half of the 20th century is "mostly" due to natural forcings, and that in the second half has been "mostly" due to human-induced increases of GHGs.**

MacCracken CH1-14, Page 20, Line 425: Change "aerosols" to "human-contributed aerosols" to make clear this was not a natural influence.
Response: **Text at this point is making a general point about the entire 20th century, so sentence is retained to include tropospheric aerosols in a general sense, although the references cited obviously weigh in more heavily on the anthropogenic component.**

MacCracken CH1-15, Page 21, Lines 430-435: Again, has it been reestablished with the revised data sets and with more general analysis techniques than simply choosing a breakpoint at a convenient point that the regime shift is real (especially given that it does not affect surface temperatures nearly as much)? Do we know (and with what level of confidence?) that this is not an artifact of the spatial coverage of the radiosonde network? Does its spatial extent really merit this being so prominently featured? Do we really know that the change in slope of the NH temperature trend was associated with this event—which might have caused the other? So much attention to this shift seems to me to give too much acceptance to the rather arbitrary time lines used by Pat Michaels to suggest that there really was a significant shift rather than a close occurrence of opposing fluctuations.

Response: **Accepted. Paragraph revised. Reference to "regime shift" dropped. A new reference (Wigley et al., 2005) is added.**

MacCracken CH1-16, Page 25, Lines 497-500: The authors should consider redoing this plot using an equal area projection instead of the misleading Mercator projection.

Response: **The figure has been redone with a different projection as requested.**

MacCracken CH1-17, Page 25, Line 502-508: Given the vast areas where there is virtually no coupling between the surface and the tropospheric temperature monthly anomalies, it would be helpful to have an explanation about why there should be a high correlation between changes in the surface and troposphere over longer times. It would also be particularly helpful to explain why, given the extensive areas where decoupling is evident, the analysis presented in this report should focus so much on changes in the global lapse rate, especially when the rate is apparently being based on the difference between a surface and a tropospheric temperature without consideration of where inversion are and how they might weaken or strengthen. With such a relatively small area of close coupling, one suspect that the atmospheric "smoothing" that is discussed would be causing a rather sizeable disturbance of the system, and so there would be varying patterns of change to be examined and considered. At the very least, it should be mentioned that this diagram makes it inappropriate to make local to regional comparisons of surface and tropospheric changes—there are just too many regions where the two are not connected (at least, directly).

Response: This point is accepted, and it is addressed more fully in the subsequent text. Indeed, the relevant point is that global trends from surface and tropospheric temperature records should not be expected to match even if both sets of measurements were perfect (lines 580-582). Figure 1.5 is of more relevance to this discussion than correlations, as discussed in lines 510-512, because it better illustrates differences in variability produced by the differences in physical processes at the surface and in the lower troposphere. Concerning the last sentence of the comment, this point is already made in lines 507-508.

MacCracken CH1-18, Page 26, Line 518 (Figure 1.5): Again, it would be more appropriate to be using equal area maps instead of Mercator projections (this is true throughout this report). See Figure 2.1, which is closer to equal area and gives a quite different impression than would be the case with a Mercator projection.

Response: The figure has been redone with a different projection as requested

MacCracken CH1-19, Page 28, Line 550-554: The text here needs to indicate that all of these changes in circulation are not fully understood (e.g., how they might couple to the smoothing going on, the regional patterns of forcing, etc.). Unlike other sections, there seems to be no qualification on this discussion.

Response: The text has been modified to include this point.

MacCracken CH1-20, Page 28, Lines 557-559: It is not at all clear why there needs to be mention of wind blowing "from ocean to land to ocean"—why not say "from land to ocean to land" which would seem to encompass the same sets of winds (or why not leave this phrase off entirely)? Also, change "this moderating influence of the winds contributes to less" to "these stronger winds tend to moderate" and change "tropospheric data" to "tropospheric temperatures" as it is not the data that are moderated. I also do not understand the comparison between "winds blowing from ocean to land to ocean" and "[winds blowing?] at the surface"—are not the former also at the surface, or are they in the "lower atmosphere"—and if so, at what elevation? All quite confusing.

Response: The suggested changes to the text have been adopted.

MacCracken CH1-21, Page 29, Line 571: The phrase "do not explain" seems very strong and unqualified. Does this mean does not explain to two-sigma, or is not even of the right sign, or what. And does this cover all types of explanations, or just some simple correlation that does not account for various alternative ways in which the coupling might exist. Overall, just seems too strong a statement.

Response: The text has been reworded to address this concern.

Trenberth CH1-1, Page 2, Line 51-56: The summary is dumbed down and becomes meaningless in places.

Response: The opening paragraph has been revised and slightly expanded to spell out some more details, without derailing a succinct communication of the principal message that responds to the specific question posed to this Chapter. Too much detail would detract focus from the main point viz., the variation of temperature in the vertical.

Trenberth CH1-2. Page 2, Line 54: the tropopause is also a function of longitude.

Response: Incorporated.

Trenberth CH1-3, Page 2, Line 56: The tropopause is more a dynamic phenomenon than radiative. The role of dynamics is underplayed throughout this section.

Response: The revised paragraph makes it clear that the entire thermal structure is the result of a balance between radiation, convection and dynamical heating/cooling. Dynamical processes are stated as responsible for the mixing of heat vertically and horizontally. The rate of decrease of temperature with height is mentioned as being dependent on geographical conditions and meteorological factors.

Trenberth CH1-4, Page 5, Line 108, Figure 1.1: contains major errors at both poles where the contouring program has not accounted for the interpolation across the pole value and has artificially closed the contours. The period used for Fig 1 should be specified and I hope it is only after 1979? The line of the tropopause makes little sense.

Response: The error is the plotting routine has been corrected, and a new Figure 1.1 has been drafted. Also, text has been added to the figure caption to make it clear that the tropopause pressure level was obtained from the NCEP reanalyses, and not computed directly from the plotted temperature field. The comment "The line of the tropopause makes little sense" refers to the fact that, at high latitudes for instance, the relationship between the tropopause and the plotted temperature field is not clear. Indeed the tropopause at high latitudes is ill-defined, especially during winter. Finally, the period of computation (1979-2003) is stated in the revised caption.

Trenberth CH1-5, Page 7, Line 153: delete "drastic"

Response: Done.

Trenberth CH1-6, Page 8, Line 166-168. This is grossly oversimplified. The dynamics does not "homogenize" temperatures in fact during cyclogenesis it creates cold fronts and warm fronts and increases temperature gradients. Temperatures should be linked to pressure gradient and recognize geostrophy and thermal wind balance. Generally this whole section is weak on dynamics and even wrong.

Response: Accepted in part and text is revised to bring in more of the "dynamical" sense. Disagree that the material presented is wrong. The word "homogenization" is deleted. Instead, large-scale dynamical mechanisms are mentioned as resulting in more spatially uniform temperatures above the boundary layer on monthly-mean and longer time scales. It is out of scope to discuss issues like fronts and cyclogenesis. The idea of this introductory chapter is not to wade into a lot of technical details.

Trenberth CH1-7, Page 8, Line 179-181: This is oversimplified and ignores the Hadley and Walker circulations which play a key role in the tropopause. The main variations in the atmosphere have opposite signs of temperature perturbations below and above the tropopause as divergence in the upper troposphere is compensated for by subsidence in the stratosphere and upward motion in the troposphere (see Trenberth and Smith 2005 submitted and available from our web site for great examples.)

Trenberth, K. E., and L. Smith, 2005: The vertical structure of temperature in the tropics: Different flavors of El Niño**.** *J Climate***,** Submitted.

Response: The last three paragraphs of 1.1 have been revised and rearranged. The sense that dynamics is underplayed is rectified by deleting the sentence/s that apparently gave such an impression, and by mentioning that the Hadley and Walker circulations play a key role in the atmospheric energy balance of the tropics and subtropics thereby influencing the thermal structure in those regions. It is not possible to discuss these in greater detail in the manner of the contemporary technical literature. Instead, we cite standard text books where the reader can go to acquire more details. The specific reference in the comment carries a lot of technical details that are inappropriate for an introductory chapter of this document.

Trenberth CH1-8, Page 9, Line 188: yes it is too simple to the point of being wrong.

Response: The paragraph containing the sentence has been revised to ensure that this is being discussed in a paragraph in the context of radiative-convective-dynamical balance. The decrease of lapse rate due to an increase in humidity can be obtained, to a very good approximation, from considerations of vertical motion of saturated air (e.g., Houghton, 1977). One gets this same general result even allowing for mixtures of saturated and unsaturated air. Undoubtedly, in the real atmosphere, the quantitative aspects require more detailed considerations of other factors, such as planetary-scale motions, but the principal result still holds. We have added that atmospheric circulation, which accompanies changes in humidity, also needs to be considered. The word "simple" is dropped.

Trenberth CH1-9, Page 9, Line 194: again too simple. Please should look at Trenberth and Stepaniak 2003a,b for comprehensive views of the energy budget and the overwhelming dominance of dynamics and latent heating and not radiation in the atmospheric diabatic heating.

Trenberth, K. E., and D. P. Stepaniak, 2003: Co-variability of components of poleward atmospheric energy transports on seasonal and interannual timescales. *J. Climate*, **16**, 3690-3704.

Trenberth, K. E., and D. P. Stepaniak, 2003: Seamless poleward atmospheric energy transports and implications for the Hadley circulation. *J. Climate*, **16**, 3705–3721.

Response: The word "simple" is dropped. "Radiative-convective" is replaced by "radiative-convective-dynamical". Disagree that latent heating has not been recognized. "Convection" is mentioned in quite a few places. Text is revised in a few additional places now to convey this point. Large-scale dynamics is also invoked at appropriate places. Disagree with the assertion that radiation has been stated to be dominant in the diabatic heating. In fact, the complicated interactions of solar radiation with the clouds in the Earth's atmosphere, and that of longwave exchanges between various layers of the atmosphere are not mentioned at all – basically owing to the requirement of simplicity. There is a constraint in presenting comprehensive discussions, owing to the space limitation and scope set for this chapter.

Trenberth CH1-10, Page 10, Line 208-209: This is not true in the lower stratosphere, where dynamics dominates.

Response: Accepted. Dropped in the revised section 1.1.

Trenberth CH1-11, Page 12, Line 246: Table 1.1. Some of these entries do not make sense; e.g., isn't the level of confidence very high that contrails have a small effect?

Response: Accepted. The global-mean forcing estimates are now listed in a new column in Table 1.1. The values make it clear that contrails have a small forcing and a low level of confidence. Table caption and text are revised to convey the point.

Trenberth CH1-12, Page 13, Line 278: "need not... can" should be "is not localized and is manifest…"

Response: Argument is accepted but text revised to indicate that in general it is not localized.

Trenberth CH1-13, Page 13, Line 279: "homogenize" this is not true, it tends to geostrophy whereby gradients of pressure and temperature are balanced by Coriolis effects, witness the thermal wind equation!

Response: Text revised, mentions "atmospheric motions and processes".

Trenberth CH1-14, Page 14, Line 304: None of the radiative forcings are uniform because they depend on cloud and water vapor: yes see lines 308-310.

Response: Accepted. Sentence dropped.

Trenberth CH1-15, Page 15, Line 323: aerosols are not just forcings but also feedbacks as they have short lives and depend on the flow and rainout.

Response: Accepted. "Aerosols" dropped.

Trenberth CH1-16, Page 16, Line 344: add aerosols.

Response: Amended to include "aerosol-cloud interactions".

Trenberth CH1-17, Page 28, Line 552 -553: the NAM is the Northern Annular Mode not NH Annular Mode; similarly for SAM.

Response: The text in question has been corrected.

Robock CH1-1, p. 5. Fig 1.1: Antarctica is missing and must be shown instead of extrapolated values underground.

Response: Temperature values below ground in the zonal average have not been contoured in the revised Figure 1.1.

Robock CH1-2 Chapter 1, p. 5. Fig 1.1: At 90N and 90S, the wrong values are plotted at all heights, and the contours make unrealistic bends between the poles and the next grid point. This needs to be corrected.

Response: The error is the plotting routine has been corrected, and a new Figure 1.1 has been drafted.

Robock CH1-3, p. 12, Table 1.1. It needs to be made clear that Sulfate aero. (direct), Black carbon aero. (direct), Organic carbon aero. (direct), Biomass burning aero. (direct), and Indirect aerosol all refer to tropospheric aerosols, and that Volcanic aero. refers to stratospheric aerosols. Volcanic emissions into the troposphere are sulfate aerosols and are covered by Sulfate aero. (direct) and Indirect aerosol. This is not clear from the way it is presented here.

Response: Accepted. Table caption is revised.

Robock CH1-4, p. 13, line 258: Only need one *.

Response: Done.

Robock CH1-5, p. 19, line 401: None of these are primary references to winter warming. I suggest you include a reference to "Robock (2000) and references therein," which discusses this in detail and includes all references to previous work.
Robock, Alan, 2000: Volcanic eruptions and climate. Rev. Geophys., 38, 191-219.

Response: Accepted. Incorporated.

Robock CH1-6, p. 21, line 433: The work of Lindzen and Giannitsis (1998) has been discredited by Wigley et al. (2005) and should not be referenced alone without including the fact that the climate model they used has serious problems and cannot reproduce the observed amplitude and temporal scale of climate system response to volcanic eruptions.
Wigley, T. M. L., C. M. Ammann, B. D. Santer, and S. C. B. Raper (2005), The effect of climate sensitivity on the response to volcanic forcing, J. Geophys. Res., 110, D09107, doi:10.1029/2004JD005557.

Response: Accepted. Paragraph revised and Wigley et al. reference added.

Robock CH1-7, p. 25: Fig. 1.4 is based on Christy et al data which have been found to be incorrect. The references to this and every other paper based on these incorrect data should be removed from this document (preferably) or accompanied by an explanation. It is better to exclude from this report all references to publications based on wrong data.

Response: The comment refers to the reliability of the UAH dataset for long-term trends. For the plot of the correlation between monthly anomalies (Figure 1.4), there is no significant difference if the RSS data are used in place of the UAH data.

Robock CH1-8, pp. 26-27: Fig. 1.5 is based on Christy et al data which have been found to be incorrect. The references to this and every other paper based on these incorrect data should be removed from this document (preferably) or accompanied by an explanation. It is better to exclude from this report all references to publications based on wrong data.

Response: The comment refers to the reliability of the UAH dataset for long-term trends. For the plot of the correlation between monthly anomalies (Figure 1.4), there is no significant difference if the RSS data are used in place of the UAH data.

Chapter 2 Comments and Responses:

MacCracken CH2-1, Page 6, Line 130-133: There are also changing amounts of thermal emissions around urban and suburban stations, and in concentrated areas, these can be important.
Michael MacCracken, Climate Institute

Response: Accepted "including changes in nearby thermally emitting structures"

MacCracken CH2-2, Page 6, Line 136: Insert to say "poorer temporal and spatial coverage" as both aspects matter.
Michael MacCracken, Climate Institute

Response: Accepted

MacCracken CH2-3, Page 9, Line 171: Throughout this chapter (and the report), capitalize "Earth" when referring to the planet. This is done in some chapters, but not consistently through the report. [See also lines 294 and 674]
Michael MacCracken, Climate Institute

Response: Accepted

MacCracken CH2-4, Page 10, Line 198-199: "sizes that sample sizes" does not make sense.
Michael MacCracken, Climate Institute

Response: Delete the second "sample sizes"

MacCracken CH2-5, Page 12, Line 240: It would help to say "this dilemma differently"
Michael MacCracken, Climate Institute

Response: **Accepted**

MacCracken CH2-6, Page 15, Line 305: "heating and cooling"—both affect the satellite.
Michael MacCracken, Climate Institute

Response: Accepted

MacCracken CH2-7, Page 18, Line 360: Add the following phrase to the end of the sentence: "assimilation model, which represents in a theoretical manner how the atmosphere behaves." This would help as a lead-in to the next sentence's description.
Michael MacCracken, Climate Institute

Response: Accepted

MacCracken CH2-8, Page 19, Line 399: The phrase "upper air reanalyses temperatures" is quite awkward, and I think actually not grammatically correct.
Michael MacCracken, Climate Institute

Response: "Simultaneous assimilation of radiosonde and satellite data for upper-air temperatures in reanalyses is particularly challenging …"

MacCracken CH2-9, Page 32, Line 587: "obtained from a given the climate record" makes no sense.
Michael MacCracken, Climate Institute

Response: "obtained from a given climate record"

MacCracken CH2-10, Page 36, Line 674: Change "earth" to "Earth's surface"
Michael MacCracken, Climate Institute

Response: "Earth". The relative view involves the atmosphere as well as the Earth's surface, so the term Earth is sufficient.

MacCracken CH2-11, Page 40, Line 759: Change "radiosondes is" to "radiosonde observations are"
Michael MacCracken, Climate Institute

Response: Accepted

Robock CH2-1, p. 22, Table 2.1: Format the headers so that words do not break across two rows.
Alan Robock, Rutgers University

Response: Change font. This does look bad in current form.

Robock CH2-2, pp. 25-29, Table 2.1: Format the columns as unjustified, so that they can be read much more easily.
Alan Robock, Rutgers University

Response: Accepted

Robock CH2-3, p. 32, line 577: Change to "; Vinnikov et al., 2006)"
Alan Robock, Rutgers University

Response: Accepted

Robock CH2-4, p. 38, line 707: Trends should be rounded to two decimal places only. We do not know the values as precisely as presented here.
Alan Robock, Rutgers University

Response: Accepted

Robock CH2-5, p. 38, line 710: Trends should be rounded to two decimal places only. We do not know the values as precisely as presented here.
Alan Robock, Rutgers University

Response: Accepted

Robock CH2-6, p. 50, lines 1105-1106. This paper is in press. The reference should be changed to:
Vinnikov, Konstantin Y., Norman C. Grody, Alan Robock, Ronald J. Stouffer, Philip D. Jones, and Mitchell D. Goldberg, 2006: Temperature trends at the surface and in the troposphere. J. Geophys. Res., in press, doi:10.1029/2005jd006392.

Response: Accepted

Alan Robock, Rutgers University

____

Swanson CH2-1, Page 13, Lines 250-261, Figure 2.2 - The figure caption does not note how the weighting functions for the satellite MSU based retrieval are derived. Other graphs of weighting functions mention that they are derived using the U.S. Standard Atmosphere temperature vs. pressure profile (Mears 2005, supplemental data). Given that the UAH T_{2LT} algorithm is derived by using the calculated emission profile for T_2, also based on the U.S. Standard Atmosphere, there should be some discussion of the impact of the polar profile, where the tropopause is known to appear at a lower pressure level. The right most panel shows this difference, but nowhere in chapter 2 (or elsewhere) is there any discussion of the impact of different real lapse rates on the T_{2LT} product. Spencer and Christy have never openly discussed how they arrived at their algorithm for the T_{2LT}. This issue is mentioned regarding the radiosonde data in line 268, but is not discussed regarding the T_{2LT} product, except later in discussion of the use of radiosonde data to simulate the T_{2LT} product, lines 747 and 755. Christy et al. (2000) mention the effects of differing lapse rates on their correction for orbital decay, however, they do not consider the different lapse rates and lower tropopause at polar latitudes on the validity of their static weighting function.

Response: Fig 2.2 is a cartoon that provides the essential information about what layers in the atmosphere contribute to the overall brightness temperature. Differences between tropical, mid-latitude and polar profiles would be smaller than the cartoon's capability to display and would unnecessarily complicate the purpose of the diagram.

Spencer and Christy have many publications which document the characteristics of the LT profile and the motivation for it (e.g. Spencer and Christy 1992b.) CCSP Chapter 4 discusses the construction of the LT product.

The variations of the full radiation code vs. the static weighting function were addressed in Spencer et al. 1990, Spencer and Christy 1992a and 1992b, Santer et al., 1999.) Differences in r.m.s. were on the order of 0.02 K per month per grid. This is much smaller than the other impacts described. Thus the impact of the full radiation code vs. the static weighting function on ANOMALIES is very small. Litten (PhD Thesis 2005) examined this in great detail, taking into account variations in sea ice, snow cover, moist/dry ground, wind roughening of ocean surface, liquid water amount and column water vapor. The results were extremely tiny on a global and tropical scale.

No changes introduced based on this comment.

Swanson CH2-2, Page 15, Line 292 - The satellites do not traverse the poles, as this sentence implies. For most satellites in the series, the ground track reaches a peak latitude of about 81.3 degrees. Only the extreme scan positions provide coverage of the poles and these data are not directly included in the UAH T_{2LT} algorithm, which utilizes the end scans to correct for stratospheric influence found in the raw MSU channel 2 data.

Response: Accepted …"near pole to pole"

Swanson CH2-3, Page 23, Line 460 - The polar orbiting MSU does not cover every grid box every day, as implied. Some latitudes near the Equator may not see repeat coverage for upwards of days.

Response: Accepted … "per ground location except in swath gaps between 40S and 40N."

Swanson CH2-4, Page 37, Line 682-685 - The list of surface emissivity effects does not include high altitude/mountain effects (Mears and Wentz, 2005) or possible sea-ice effects (Swanson 2003). Note that the anomalous annual cycle found by Swanson (2003) does not appear in the UAH T_2 data, thus it may be concluded that this effect is due to surface factors (see Fig 1 & 2 below). Mears and Wentz (2005, supplemental data) do not include any data for the Southern latitudes greater than 70S, nor do they include date for grid points with very high mountains, such as the Himalayas.

Figure 1. Average daily zonal data for T_{2LT}, 1979-1998, from the UAH website. These are the data which are used to compute the daily anomaly values for each year. The curves represent steps of 2.5 deg in latitude, beginning with the top curve at 55S. The lowest 2 curves are for 80S and 82.5S. These curves are the result of smoothing of the actual averages, both over time and in latitude. Data from: http://vortex.nsstc.uah.edu/data/msu/t2/tmtdayacz7998_5.1

Figure 2. Average daily zonal data for T_{2LT}, v 5.2 from the UAH website. These curves correspond to those in Figure 1. Compare these data with the unsmoothed data in Figure 1, Swanson 2003. Data from: http://vortex.nsstc.uah.edu/data/msu/t2lt/tltdayacz7998_5.2

Response: The surface emissivity effects listed in the report are those which change over time. We have added "interannual sea ice variations" to the list of examples.

This issue has been discussed with the commenter in several prior emails. RS has looked at the annual cycle of radiosonde layer temperatures (absolute values) on the Antarctica rim (not MSU-simulated brightness temperatures) and UAH LT mean annual cycle temperatures. These are two different quantities. The CCSP report is concerned with the anomaly time series. The anomalies of radiosonde-simulated LT and MT and actual LT and MT are highly correlated and have almost identical trends using the same stations reported in Swanson 2003. Additionally, the focus on the report is global and tropical spatial scales, and the tiny impact of the varying ice edge would be minuscule or irrelevant for these regions. The figures referred to by the commenter were not available.

Trenberth CH2-1, Page 2, Line 58: what about the urban heat island effect?
Kevin Trenberth, National Center for Atmospheric Research

Response: In the summary points ln 47-57, "micro-climate exposure" and "some errors reundoubtedly remain" are intended to include impacts such as the urban heat island. No changes made.

Trenberth CH2-2, Page 6, Line 128 to 133: there is no discussion of urban heat island effects or land use changes. This is a major shortcoming.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted

Trenberth CH2-3, Page 10, Line 194: "to" should be "too"
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted

Trenberth CH2-4, Page 10, Line 199: what about the FGGE buoys and follow-ons starting 1978?
Kevin Trenberth, National Center for Atmospheric Research

Response: While there were buoy observations, they were quite limited as implied by the text "Buoy observations became more plentiful …". For example, in Niño 2.4, there were only 230 ship/buoy match-ups in 1986 and fewer prior to 1986. After 1986, there were 1000 to 5000. No changes made.

Trenberth CH2-5, Page 18, Line 370: the correct reference for ERA-40 is: Uppala, S. M., P.W. Kållberg, A.J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, A.P. McNally, J.-F. Mahfouf, R. Jenne, J.-J. Morcrette, N.A Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen 2005: The ERA-40 reanalysis. Quart. J. Roy. Meteor. Soc., in press.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted

Trenberth CH2-6, Page 18, Line 372-374: bias corrections were employed in ERA-40. A lot of effort went into this (not enough and problems remain, but no worse than for other datasets).
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted…"unless flagged and corrected as ERA-40 attempts to do".

Trenberth CH2-7, Page 20, Line 404: Randel 2004 is not in references. None of these references deal with ERA-40.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted. Bengtsson et al. included. Randal reference added.

Trenberth CH2-8, Page 20, Line 407: this conclusion is not justified. No evidence is presented, no references are given, and it is based solely on the feelings of the authors. It is not acceptable. In fact the reanalyses are given a green color in Table 2.1???
Kevin Trenberth, National Center for Atmospheric Research

Response: Since stratospheric trends are wide ranging in the Reanalyses (one is most positive, the other most negative), the CCSP authors discussed this issue and feel justified in excluding such trend influences from the major comparisons and time series at this time.

Trenberth CH2-9, Page 20, Line 415-418: is written in a prejudicial way and must be reworded. In fact most of the stratospheric influence can be very effectively eliminated! The Spencer et al. 2005 publication is not available.
Kevin Trenberth, National Center for Atmospheric Research

Response: The potential for erroneous results using simple statistical retrievals has been demonstrated in Spencer et al. 2006, which was available to the authors. Stratospheric influences are not eliminated, but recast as a difference between two layers. Non-stationarity then becomes an important issue. The section has been rewritten to address this concern.

Trenberth CH2-10, Page 21, Line 427-435: Also see Kiehl et al. 2005:
Kiehl, J. T., J. M. Caron, and J. J. Hack, 2005: On using global climate model simulations to assess the accuracy of MSU retrieval methods for tropospheric warming trends. J. Climate, 18, 2533-2539.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted.

Trenberth CH2-11, Page 22, Line 455: Table 2.1; some of the colors in here are debatable, esp. radiosondes upper air temperature as green. The table is not very useful and could be abolished.
Kevin Trenberth, National Center for Atmospheric Research

Response: Table 2.2 has been revised.

Trenberth CH2-12, Page 24, Line 481: This statement is not true. The atmosphere tries to maintain temperature and pressure gradients to match the Coriolis force and thus the thermal wind equation. It is NOT smoothing horizontally!
Kevin Trenberth, National Center for Atmospheric Research

Response: Cause is not addressed, but the simple observation of large scale coherence is noted.

Trenberth CH2-13, Page 24, Lines 485-488: This sentence does not follow from the previous ones. "Thus" is not correct because the stations are not "properly spaced". The radiosondes do not give good enough coverage to do hemispheric or zonal averages. Yet this is later assumed and it is wrong!
Kevin Trenberth, National Center for Atmospheric Research

Response: Tests with the current distribution of sondes produces a very reasonable global mean value. "As a result, a given precision for the global mean value over, say, a year can be attained with fewer, if reasonably spaced, upper air measurement locations than at the surface (Hurrell et al. 2000). Thus knowledge of global, long-term changes in upper-air temperature is likely limited more by instumental errors than spatial coverage. However, for some regional changes (e.g., over sparsely observed ocean areas) sampling problems may compete with or exceed instrumental ones."

Trenberth CH2-14, Page 26, Synoptic: 1-4 days? Surely at least a week.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted (3-7 days).

Trenberth CH2-15, Page 26, Intraseasonal: "unknown cause"??? Organized convection.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted.

Trenberth CH2-16, Page 27, QBO: "nearly periodic" not so. Not so easy to remove the signal either. It aliases with ENSO.
Kevin Trenberth, National Center for Atmospheric Research

Response: Evidence indicates nearly periodic fluctuations. Studies show clear signal (e.g. Christy and Drouilhet 1994).

Trenberth CH2-17, Page 28, Decadal: "50 year PDO" huh?
Kevin Trenberth, National Center for Atmospheric Research

Response: Changed to "multi-decadal"

Trenberth CH2-18, Page 30, Line 540: three to five decades the T change is not 2 tenths per decade. Closer to 1.5 tenths since 1970.
Kevin Trenberth, National Center for Atmospheric Research

Response: Accepted.

Trenberth CH2-19, Page 30, Line 542: "few hundredths" does not follow as such small values are neither physically meaningful nor measurable. The argument here is specious.
Kevin Trenberth, National Center for Atmospheric Research

Response: The idea here is that trends of one to two tenths of a degree per decade in two difference layers will generate *differences* on the order of less than a tenth per decade. To know if these differences are real and significant, the precision needs to be on the order of a few hundredths per decade. This has significant import regarding physical relationships arising from convection and subsidence. This will also aid in detection of such layer differences in model output. No changes made.

Trenberth CH2-20, Page 32, Line 582: This omits sampling and statistical errors, a major omission.
Kevin Trenberth, National Center for Atmospheric Research

Response: Sampling and/or statistical errors are discussed more thoroughly in the appendix.

Trenberth CH2-21, Page 35, Line 642-646: This does not deal with the fit of linear trends to the data, which is generally poor. Not much variance is accounted for and it is a poor descriptor of what happened.
Kevin Trenberth, National Center for Atmospheric Research

Response: See pg 38 Line 714 KT comment #25.

Trenberth CH2-22, Page 35, Line 647-653: Fails to deal with urban heat island effects. Kevin Trenberth, National Center for Atmospheric Research

Response: Urbanization is a case of land-use change. "Land-use changes, including urbanization, …"

Trenberth CH2-23, Page 36, Line 668: This is not relevant as trends are not most of what goes on and the drift correction is important. In chapter 4, it is evident that differences between RSS and UAH occur even for the same satellite and are not due to merges, but must be due to diurnal cycle issues.
Kevin Trenberth, National Center for Atmospheric Research

Response: As new information shows, the diurnal corrections are a source of very minor differences between UAH and RSS, and in fact it is the merging procedure, with target coefficient differences, which has a greater role.

Trenberth CH2-24, Page 38, Line 706: these are linear trends?
Kevin Trenberth, National Center for Atmospheric Research

Response: Add "linear".

Trenberth CH2-25, Page 38, Lines 714-716: A linear trend is generally a bad fit to the data and has large error bars. This is a major issue in this report. It is not adequately dealt with.
Kevin Trenberth, National Center for Atmospheric Research

Response: Error bars in terms of how well a line fits a dynamical time series is one issue, but much of the report deals with the time series of differences, which have much smaller statistical error bars and for which the linear trend is an ideal measure (see Appendix). In addition, unlike raw time series, the null hypothesis of zero trend is an appropriate assumption for testing difference trends. Thus a trend is a useful metric to identify differences in various time series without necessarily trying to apply physical interpretations.

Trenberth CH2-26 Page 38, Line 718: yes it is a source of error.
Kevin Trenberth, National Center for Atmospheric Research

Response: The nuance of "error" is discussed in the Appendix.

Trenberth CH2-27, Page 38, Lines 719-720: This is a major issue. Linear trends depend enormously on end points and exhibit large changes with small changes in length. Multiple examples exist in this report, such as the 1976/77 climate shift.
Kevin Trenberth, National Center for Atmospheric Research

Response: We acknowledge the possible misinterpretations applied to linear trends.

Chapter 3 Comments and Responses:

MacCracken CH3-1, Page 2, Line 43-45: Well said.
Michael MacCracken, Climate Institute

Response: Comment noted.

MacCracken CH3-2, Page 2, Line 51-52: The indication of how well established this supposed "abrupt climate regime shift" is not at all clear—and this report does not seem to have gone back with the revised data sets and knowledge about changing spatial coverage, etc. to reconfirm this supposed shift. In terms of the terminology here, it is also not clear it is a "climate" shift, but rather seems to be a "circulation" shift, for the report indicates there is no abrupt shift in the surface temperature. It is not even clear what is meant by "regime" in the discussion—is this really a global phenomenon? It is not clear if it is natural or induced—after all, the circulation tends to smooth various anomalies and forcing gradients (from volcanic eruptions, sulfate aerosols, etc.). And it is not clear if this was not just the coincidence of two (or three) different anomalies close together—named a shift by those who arbitrarily (at least in some cases) chose a particular time to divide up taking averages. So, in my view, the phrase "a time coincident with a previously identified abrupt climate regime shift" needs to be either eliminated or significantly modified as it is not at all clear what is meant and if this is really a global shift—or may just due to the station network, etc. [A similar comment applies to lines 70-73—is a shift in circulation really a climate shift? Has this happened before, etc.?]
Michael MacCracken, Climate Institute

Response: We refer to the phenomenon as a change in "regime", following Trenberth (1990), which we cite on page 18, line 425. We have not performed any analyses to confirm the existence of this regime change because the reality of such a regime change is well-accepted, with an extensive body of literature to back this up. However, we feel it is important to note in passing any apparent associations between the regime change and features of interest discussed in this report. That this regime shift is "real", and not simply an artifact of any inherent inadequacies of the observational monitoring systems is evidenced by its appearance in many independently measured parameters, some of which are documented by two of our references, Trenberth (1990) and Trenberth and Hurrell (1994). We have added a more recent and quite comprehensive reference (Deser et al., 2004) as well. More germane to our report, an abrupt change in tropical lapse rates is common to several different analyses that we cite in section 3.7.2 (Brown et al., 2000; Gaffen et al., 2000; Hegerl and Wallace, 2002; Lanzante et al., 2003b) based on different raw datasets and approaches to the problem. The term "climate" seems more appropriate than "circulation" given that the documented changes include parameters beyond the realm of just those that measure atmospheric circulation. As to the physical causes of the regime shift, "natural or internal" in the words of the reviewer, there is no consensus in the climate community, therefore we have not commented on it. As far as our use of the phrase "a time coincident with a previously identified abrupt climate regime shift", it seems appropriate given the widely accepted nature of the 1976/77 jump in numerous climate variables.

MacCracken CH3-3, Page 3, Line 85-87: This sentence does not make sense. The change was what it was—temperature decreased, then went up following volcanic eruptions, etc. What this sentence is apparently referring to is the non-volcanically induced change due to the change in greenhouse gas concentrations and aerosol loading—but this is an attribution issue, requiring a separation of the influences. At this point in the analysis (i.e., describing what happened), all one can say is what the record shows—and that is bumps up after volcanic eruptions on a generally descending record of temperature.
Michael MacCracken, Climate Institute

Response: The purpose of this statement is to indicate that there is ambiguity in describing the evolution of the temperature of the stratosphere: was the temperature decrease more gradual or did it occur to a greater extent in the form of step-like decreases following major volcanic eruptions. In the body of the report we point to a quantitative study which has documented this ambiguity (Seidel and Lanzante, 2004). As pointed out by the reviewer, any such interpretations have relevance to attribution. Although this chapter does not deal formally with attribution, the description and interpretation of the temperature record does have relevance to attribution. Therefore it seems appropriate to point this out in the course of describing the temperature record.

MacCracken CH3-4, Page 4, Line 113-115: Just because one cannot describe one as "the best" does not mean that one cannot eliminate some as not incorporating all of the currently recognized corrections. I would hope that "credible" means that the datasets do include all recognized corrections. It would also be helpful here (as expanded upon in a general comment) to be able to refer to a table or appendix that shows the sequence of corrections that have had to be incorporated and the changes in the results that have resulted so that it is clear what 'credible" means—and to make clear that conclusions from earlier datasets were at best premature.
Michael MacCracken, Climate Institute

Response: As stated in lines 121-124 on page 4, our criteria for including datasets in this report are that the datasets are active and have been homogeneity adjusted. We feel that an attempt to eliminate the important sources of artificial inhomogeneities constitutes a "credible" dataset. Exactly what is "important" and how these are to be eliminated or reduced is determined by each dataset construction team. Some of the discussion in Chapters 3 and 4 delves into some of distinctions between approaches used by different teams. Given the complexity and highly technical nature, as well as space constraints, it would not be practical to try to present the details of the homogeneity methods. The interested and more technically competent reader can find these in the references that we have provided. As far as the suggestion of passing value judgments on the different "credible" datasets, this is a very contentious issue. The committee writing this report has representatives from most of the datasets that have been used. We have spent countless hours debating the merits of different approaches used in dataset construction and have not been able to reach a consensus. Attempts to place more or less value on a particular dataset have been met by vigorous objections by at least some committee members in all cases. Until more objective approaches to "grading" the validity of the datasets have been implemented (we make some recommendations pertinent to this in Chapter 6) we are unable to make value judgments to distinguish the various datasets.

MacCracken CH3-5, Page 5, Line 126-128: Again, this is why it would help to have a table going over the history of the corrections and their incorporation in various data sets. And why should one not consider ones excluded as an invalidation of that product; if the corrections were not made, then the product is at the least out of date and should no longer be used—and this needs to be said (will anyone go back to it?), and at least some of the findings in the papers based on these products should also be discounted (e.g., the conclusion that a disagreement between surface and tropospheric temperature records exists that invalidates the climate models). This phrasing is much too mushy given the progress that has been made—the decision makers need a clear statement that it is the results of this assessment that should be paid attention to and not the earlier work based on inadequately completed corrections.
Michael MacCracken, Climate Institute

Response: As far as an historical perspective on datasets, we do provide this in Chapter 3. We describe how certain datasets evolved from earlier ones. In this context we mention some of the earlier datasets that either were not homogeneity adjusted, or that have been superseded by datasets that incorporate more complete or more sophisticed adjustments. However, there is a limit to the amount of space that we can devote to such discussions since the purpose of this report is not historical in nature, rather it is aimed at producing an assessment based on the state of the art knowledge. We think we have made this clear by stating that our report is based on "credible" datasets that are considered "state of the art". We do not wish to denigrate any datasets that have been used in the past, out of respect to those pioneers that produced them. In many cases creation of the current "state of the art" datasets would not have been possible without the earlier datasets.

MacCracken CH3-6, Page 6, Line 161: Not only urbanization, but also the release of heat from combustion needs to be considered when looking at urban megalopolises.
Michael MacCracken, Climate Institute

Response: There is no need to make separate mention of heat from combustion since the term urbanization is all-inclusive: building more structures, replacing natural vegetation with asphalt, as well as an increase in the number of people and their machines respiring or combusting.

MacCracken CH3-7, Page 18, Line 424-426: Again, it is not at all clear that the change from one year to the next was really a "climate regime shift"—it may well have been a change in the atmospheric circulation, but it does not show up as a sudden change in any global index, as far as I am aware. And it is not at all clear this is not due to the closeness of two or three different anomalies (volcanic paired with ENSO paired with changes in aerosol forcing, etc.). This all needs to be much more closely analyzed, and put in the context of other times when seemingly sharp changes might have occurred, etc.—given it

has not done a reanalysis of this supposed shift, this report needs to be much more cautious in making such a definitive finding.
Michael MacCracken, Climate Institute

Response: See response to MacCracken CH3-2.

MacCracken CH3-8, Page 22, Line 496: Again, rather than just say "most up to date", the report should provide, in a table or appendix, a synopsis of the various corrections that have been made, why this has been necessary, when it occurred, and how it affected estimates of the trends. It should also be noted that these correction were in a number of cases larger than the suggested uncertainty in the observations, making clear that a complete uncertainty analysis had not previously been done and that instead the numbers given referred to the supposed "precision" of the measurement and not its uncertainty.
Michael MacCracken, Climate Institute

Response: In this statement we are clarifying exactly which version of the particular dataset we have used. This is necessary since many datasets evolve over time, with the most current version superseding earlier ones. Again, as stated in response to "MacCracken CH3-4", providing the technical details of dataset construction is beyond the scope of this report. The interested and technically competent reader is referred to the appropriate references that we cite. Issues pertaining to the magnitude of adjustments and uncertainties in the measurements are discussed elsewhere in this report (Chapters 2 and 4).

MacCracken CH3-9, Page 23, Line 509-511: It is fine to mention that Mears and Wentz found this methodological error, but this methodological error (and the record of other corrections and errors) needs to also be mentioned in the paragraph where the UAH data set is described—it is simply inappropriate not to indicate this where the actual data set to which the comment refers is described.
Michael MacCracken, Climate Institute

Response: We have added mention of this to footnote 11 which explains the distinction between versions 5.1 and 5.2 of the UAH datasets.

MacCracken CH3-10, Page 32, Line 688: Rephrase to say "The annual average temperature of most of the land and ocean surface increased …". The present phrasing does not really make sense.
Michael MacCracken, Climate Institute

Response: The wording has been changed as suggested.

MacCracken CH3-11, Page 32, Line 692-695: Given that the differences in the correlation of surface and tropospheric temperatures over land and ocean (Figure 1.4), why would one even expect there to be consistency of what is happening over the oceans and land areas separately? What is it that is leading to this consistency—and why is this the expectation? Some additional explanation is needed.

Michael MacCracken, Climate Institute

Response: The issue being raised related to Figure 1.4 has to do with the interannual correlation between surface and MSU temperature and how the correlation is higher over land than over ocean. The point being discussed in the chapter is not correlation but the similarity of long term trends. As a general rule, we expect that the longer the time frame considered, the larger the spatial scale of variations. So when we look at long-term trends we expect to see large-scale consistency. It doesn't seem that this needs to be written in the text, as most people would already assume that if the ocean off a coastline was cooling it would be unlikely to see the land near it warming.

MacCracken CH3-12, Page 33, Line 697-702: An additional sentence or phrase needs to be added to make clear that this discussion is not intending to suggest that this all could not be due to overall global warming.
Michael MacCracken, Climate Institute

Response: The wording has been changed as suggested.

MacCracken CH3-13, Page 33, Line 712-716: Might an alternative (or complementary) explanation be that the atmosphere adjusts far faster than the ocean conditions do (due to heat capacity and rapidity of motion) and so this may not need to be related to ENSO and the interdecadal Pacific oscillation?
Michael MacCracken, Climate Institute

Response: The portion of text discussing this aspect has been removed.

MacCracken CH3-14, Page 32, Line 718 to Page 33, Line 720: Do we really have good enough data to be able to say this in some confident manner, given the problems of data coverage and data limitations? This seems like very fine graining given that the ocean data are often averaged over 5 year periods to just try to get enough coverage. And do these conclusions apply subsequent to the most recent versions of the data sets?
Michael MacCracken, Climate Institute

Response: The portion of text going into such fine grain detail has been removed.

MacCracken CH3-15, Page 44, Footnote 19: This explanation seems to leave out mention of the chaotic nature of the climate system itself—that is, different realities may occur.
Michael MacCracken, Climate Institute

Response: The purpose of this footnote is to explain the pitfalls in examining trend maps in this report. The reviewers comment does not seem to be relevant to this point.

MacCracken CH3-16, Page 45, Line 889-893: First, it is not only theory that indicates that there should be differences by location—that the correlations of surface and troposphere are so low would seem to indicate that this is also expected based on the

observations. And this is a rather strange conclusion, given that the observational record shows that there is little correlation between surface and tropospheric monthly anomalies except over NH continents. This phrasing makes it sound strange that the models are getting this result, but should we not be expecting this behavior on an annual or decadal basis given the low correlation on a monthly basis and the presence of various other feedback processes (like ice/snow albedo feedback)? This section needs to conclude with some sort of more positive endorsement of what the models are showing—it all makes pretty good sense.
Michael MacCracken, Climate Institute

Response: The reviewer is suggesting a more detailed discussion of the issue. Such details are more appropriate elsewhere, such as Chapter 1, to which we refer the reader. The purpose of our statements here (lines 888-893 on pages 44-45) are to remind the reader that in examining the maps that there is good reason a priori to expect differences between the surface and aloft. The reviewers statement "This phrasing makes it sound strange that the models are getting this result" seems counter to what we say in line 889 "based on theory we expect the difference in trend between the surface and troposphere to vary by location" and lines 891-893 "... climate model projections ... should lead to more warming of the troposphere than the surface in the tropics, but the opposite in the Arctic and Antarctic". These two statements that we make are consistent. The reviewer states that this section should make a more positive endorsement of what the models are showing. But this would be inappropriate here since Chapter 3 is not concerned with comparing models and observations -- that is discussed in Chapter 5.

MacCracken CH3-17, Page 47, Line 904: Change "decreases" to "decrease"
Michael MacCracken, Climate Institute

Response: The change has been made as suggested.

MacCracken CH3-18, Page 51, Line 966-967: But are these conclusions based on an analysis using the latest corrections for the radiosonde data by Sherwood? Are all the corrections in all the datasets (radiosonde and satellite)?
Michael MacCracken, Climate Institute

Response: The analysis by Sherwood et al. (2005) is diagnostic in the sense that it attempts to account for one additional factor in assessing trends, but it has not actually produced a new dataset. The implications of Sherwood et al. (2005) are discussed later in Chapter 4. By its very nature Sherwood et al. (2005) is applicable to one particular dataset (the LKS dataset, from which RATPAC is derived). Its relevance to the other radiosonde dataset that we employ (HadAT) is less obvious. On the other hand, it has no relevance to the satellite datasets.

MacCracken CH3-19, Page 54, Line 1020: Again, is this really a regime shift, or just the coincidence of various anomalies, just giving that impression given a particular choice of where to divide one's analysis?
Michael MacCracken, Climate Institute

Response: See response to "MacCracken CH3-2".

MacCracken CH3-20, Page 55, Line 1034-1041: It seems a bit strange that all these results are drawn from a period when there were unresolved problems with the various datasets. Are these conclusions still justified? It also seems to me quite questionable generating estimates of lapse rate changes by comparing changes in surface and tropospheric temperatures when there is, over much of the Earth, a very low correlation between monthly surface and tropospheric anomalies, indicating that inversions are present and so the lapse rate needs to be much more carefully determined and it is not even appropriate to calculate it in the way it has apparently been done. If such an approach is being justified based on smoothing occurring, then why is it not working at one month, but would be expected to at longer times? How should we know what to be expecting—especially given that models do not yet fully resolve the PBL inversions that get created?
Michael MacCracken, Climate Institute

Response: The reviewer is correct in implying that the unresolved problems with the various datasets may be contributing to inconsistencies between the cited studies. Some of the other comments made by the reviewer point to other possible shortcomings of these studies. We are simply reporting the current state of knowledge. Trying to resolve these issues is far beyond the bounds of our report. Motivation for these studies was the changes expected based on climate models driven by historical forcings. On time scales longer than a month, and when averaging over the tropics, these studies do generate some results consistent with expectations -- a lagged response to ENSO variations and an abrupt change associated with the 1976-77 climate regime shift. The question is, how well can both the observations and models resolve the spatial details, given the shortcomings of each? This is still an open question.

____

Robock CH3-1, p. 23, line 518: Add reference: (Grody et al., 2004; Vinnikov et al., 2006)
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-2, p. 23, line 519: Remove "(M)"
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-3, p. 23, line 521: Change to: "Also, in both versions they do not adjust ..."
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-4, p. 23, line 524: Change "scheme more consistent with that of the other two groups" to "scheme that is different from that of the other two groups"
Alan Robock, Rutgers University

Response: The latter part of the sentence has been removed to avoid any confusion.

Robock CH3-5, p. 23, line 526: Change "0.17 ºC/decade" to 0.20ºC/decade.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-6, p. 23, lines 526-528. Remove "Very recently they have revised their method to produce a third version of their dataset, which we use in this report, whose trends differ only slightly with those from the second version." What you call the third version is the same as the second version.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-7, p. 23, line 530: Change to "; Vinnikov et al., 2006)"
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-8, p. 31, Table 3.2: Do we really know the confidence intervals to three decimal places? I think they should only be expressed to two decimal places.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-9, p. 36, Table 3.3: Do we really know the confidence intervals to three decimal places? I think they should only be expressed to two decimal places.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-10, p. 38, Fig. 3.4b: Change "VG" to "UMd"
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-11, p. 38, Fig. 3.4b: Change green box (for UMd) in legend to solid, to match how it is plotted in figure. Also make it larger to the same size as the other boxes.
Alan Robock, Rutgers University

Response: The change cannot be made as suggested because the fact that the symbol is plotted as solid indicates statistical significance as indicated in the figure caption. All of the symbols in the legend are open for this reason.

Robock CH3-12, pp. 40-41, Table 3.4: Do we really know the confidence intervals to three decimal places? I think they should only be expressed to two decimal places.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-13, p. 49, Table 3.5: Do we really know the confidence intervals to three decimal places? I think they should only be expressed to two decimal places.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Robock CH3-14, p. 64, lines 1385-1387. This paper is in press. The reference should be changed to:
Vinnikov, Konstantin Y., Norman C. Grody, Alan Robock, Ronald J. Stouffer, Philip D. Jones, and Mitchell D. Goldberg, 2006: Temperature trends at the surface and in the troposphere. J. Geophys. Res., in press, doi:10.1029/2005jd006392.
Alan Robock, Rutgers University

Response: The change has been made as suggested.

Trenberth CH3-1, Page 2, Lines 58-61: There are no assessments in this chapter as to which products have known flawed: should refer to chapter 4.
Kevin Trenberth, National Center for Atmospheric Research

Response: The report is structured as such that data shortcomings are discussed in Chapter 4. The preface has been modified to make this clearer.

Trenberth CH3-2, Page 2, Line 63: the changes since 1958 are not linear. Indeed lines 50-54 say so. Using a single rate or decade is misleading.
Kevin Trenberth, National Center for Atmospheric Research

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix.

Trenberth CH3-3, Page 2, Line 64: the sonde data are known to be flawed and trends too low (chapter 4). Why should the three satellite datasets be treated equally when some

have known problems. Where is the commentary on these? (Later we find it in chapter 4.)
Kevin Trenberth, National Center for Atmospheric Research

Response: The report is structured such that data shortcomings are discussed in Chapter 4. The preface has been modified to make this clearer. The committee writing this report has representatives from most of the datasets that have been used. We have spent countless hours debating the merits of different datasets and have not been able to reach a consensus. Attempts to place more or less value on a particular dataset have been met by vigorous objections by at least some committee members in all cases. Until more objective approaches to "grading" the validity of the datasets have been implemented (we make some recommendations pertinent to this in Chapter 6) we are unable to make value judgments to distinguish the various datasets.

Trenberth CH3-4, Page 2, Lines 66-69: given the dominance of ENSO, using linear trends is flawed. See also lines 70-73: it is not linear across 1976.
Kevin Trenberth, National Center for Atmospheric Research

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix. We make note of the nonlinear change across 1976 in the chapter text as well as in the key findings.

Trenberth CH3-5, Page 2, Lines 75-79: the balloon data are known to be flawed and biased in the stratosphere, c.f. Randel and Wu 2005.
Kevin Trenberth, National Center for Atmospheric Research

Response: Since the report is structured such that data shortcomings are discussed in Chapter 4, most discussion of possible data flaws is deferred until then. The purpose of Chapter 3 is to present estimates of temperature change as determined from the observed data, regardless of any shortcomings of those data. Randel and Wu (2005) is discussed in Chapter 4.

Trenberth CH3-6, Page 4, Line 114: It is still possible to critique the datasets and this is not done.
Kevin Trenberth, National Center for Atmospheric Research

Response: The report is structured such that data shortcomings are discussed in Chapter 4.

Trenberth CH3-7, Page 4, Line 119: the collective expert judgment is a function of the participants, many of whom have vested interests.
Kevin Trenberth, National Center for Atmospheric Research

Response: This is inevitable since the participants were chosen based on their expertise. Experts tend to be those persons that have invested the most effort towards a particular problem. Nevertheless, the committee of participants represents a diverse group, covering the spectrum of opinion and outlook to achieve the necessary balance.

Trenberth CH3-8, Page 18, Line 433: Why is there nothing in this chapter on the major problems known in the radiosonde datasets? See Sherwood et al. 2005 and Randel and Wu 2005. Refer to chapter 4.
Kevin Trenberth, National Center for Atmospheric Research

Response: The report is structured such that data shortcomings are discussed in Chapter 4. Chapter 4 contains a discussion of problems with the radiosonde datasets, including those presented by Sherwood et al. (2005) and Randel and Wu (2005). The Preface has been modified to make clearer the structure of the report with regards to the purpose of each chapter.

Trenberth CH3-9, Page 29, Lines 632-634: is a cop out on for use of reanalyses. It does not deal with the advantages of reanalyses, such as four dimensional assimilation and multivariate data.
Kevin Trenberth, National Center for Atmospheric Research

Response: The unsuitability of reanalyses for use in assessing long-term climate change is discussed at length in Chapter 2.

Trenberth CH3-10, Page 29, Line 647: such linear models are often not a good fit to the data, as shown in this chapter.
Kevin Trenberth, National Center for Atmospheric Research

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix. We make note of any important nonlinear changes both in the chapter text as well as in the key findings.

Trenberth CH3-11, Page 33, Lines 699-702: refer to chapter 1 for physical reasons why they differ.
Kevin Trenberth, National Center for Atmospheric Research

Response: Modified as suggested.

Trenberth CH3-12, Page 33, Lines 705-724: the discussion here is specious. During ENSO there are changes in surface fluxes of heat of order ±50 W m-2 and changes in SST vs. marine air temperature, see Trenberth et al. 2002.

Trenberth, K. E., D. P. Stepaniak and J. M. Caron, 2002: Interannual variations in the atmospheric heat budget. J. Geophys. Res., 107(D8), 4066, 10.1029/2000JD000297

And this discussion fails to recognize these.
Kevin Trenberth, National Center for Atmospheric Research

Response: The portion of this section dealing with fluxes has been removed and a more narrowly focused discussion of the differences between NMAT and SST is now presented.

Trenberth CH3-13, Page 34, Line 740: this is incorrect, as shown by Dai et al.: the DTR dependence on clouds is all through the maximum temperature being affected by clouds blocking the sun. The infrared effects occur day and night with clouds and do not affect DTR.
Kevin Trenberth, National Center for Atmospheric Research

Response: It seems unreasonable to expect that the infrared forcing from clouds that occurs during the day and night would have the same impact on surface temperature taken in a deep well mixed boundary layer during the day and a shallow stably stratified nocturnal boundary layer. The paper cited does indeed find the effect of clouds on minimum temperature to be small. In fact in Table 1, it has a negative correlation between cloud cover and minimum temperature during the summer (i.e., as clouds increase minimum temperature decreases). The paper also states that the historical cloud data used "contain inhomogeneities". Thus, the results could be an artifact of the statistics (cloud changes being associated with changes in synoptic conditions which impact temperature for other reasons).

Furthermore, the comment that the IR effect is both day and night and therefore doesn't impact DTR is interesting but misses the point. If that was true, then the IR effect would warm the nighttime temperatures and warm the daytime temperatures the same amount. The fact that clouds cool Tmax indicates that the reflection of sunlight has a greater effect than the IR effect. So clouds cool Tmax. But the IR effect still warms Tmin. So what we say is technically correct: "This makes physical sense since clouds tend to cool the surface during the day by reflecting incoming solar radiation, and warm the surface at night by absorbing and reradiating infrared radiation back to the surface." To make this accurate in keeping with the comment would simply require us to change the statement that clouds cool daytime temperatures by reflecting more incoming solar radiation than the IR that they absorb and reradiate to the surface. While that would be technically accurate it would detract from the basic correct statement that we make.

Trenberth CH3-14, Page 35, Line 750: hence the trends are not linear in the troposphere.
Kevin Trenberth, National Center for Atmospheric Research

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix.

Trenberth CH3-15, Page 35, Line 752: nor are they linear in the stratosphere.
Kevin Trenberth, National Center for Atmospheric Research

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix.

Trenberth CH3-16, Page 35, Line 766, Table 3.3: the ERA-40 values go only through August 2002.
Kevin Trenberth, National Center for Atmospheric Research

Response: Our ERA-40 data end in September 2001. This is noted in the Table 3.3 caption.

Trenberth CH3-17, Page 38, Line 779, Figure 3.4a: has no error bars which should be plotted.
Kevin Trenberth, National Center for Atmospheric Research

Response: The complexity of Figure 3.4 precludes the plotting of error bars. Furthermore, there are other reasons for not showing error bars, and these are discussed in the Appendix.

Trenberth CH3-18, Page 40, Line 828, Table 3.4: The UAH T2LT value appears to be in conflict with the surface trends.
Kevin Trenberth, National Center for Atmospheric Research

Response: The purpose of Chapter 3 is to present results based on various datasets without passing any judgment on data quality or possible flaws. These issues are then discussed in Chapter 4.

Trenberth CH3-19, Page 42, Line 839, Figure 3.5: Radiosonde trends are not reliable owing to incomplete spatial sampling.
Kevin Trenberth, National Center for Atmospheric Research

Response: Possible effects due to incomplete spatial sampling of radiosonde data are discussed at the end of section 2.1 in Chapter 4.

Trenberth CH3-20, Page 43, Lines 855-864: The sondes are not global.
Kevin Trenberth, National Center for Atmospheric Research

Response: Possible effects due to incomplete spatial sampling of radiosonde data are discussed at the end of section 2.1 in Chapter 4.

Trenberth CH3-21, Page 43, Line 864: the footnote 18 is important and no reference is made to chapter 4, where it is discussed.
Kevin Trenberth, National Center for Atmospheric Research

Response: A reference to Chapter 4 has been added to the footnote as suggested.

Trenberth CH3-22, Page 45, Line 893: where is discussion of land vs. ocean? See also Chapter 1, Figs 1.4 and 1.5.
Kevin Trenberth, National Center for Atmospheric Research

Response: A sentence has been added to indicate differences in response between land and ocean.

Trenberth CH3-23, Page 48, Lines 929-963: This material ought to be purged as the sondes are known to have negative trend biases and are not global (see chapter 4).
Kevin Trenberth, National Center for Atmospheric Research

Response: The report is structured as such that data shortcomings are discussed in Chapter 4. The purpose of Chapter 3 is to present results based on observed datasets, without regard to any possible shortcomings. Possible effects due to incomplete spatial sampling of radiosonde data are discussed at the end of section 2.1 in Chapter 4.

Trenberth CH3-24, Page 54, Line 1020: note the comments here that linear trends are not an appropriate fit.
Kevin Trenberth, National Center for Atmospheric Research

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix.

Trenberth CH3-25, Page 54, Line 1023: these are not trends.
Kevin Trenberth, National Center for Atmospheric Research

Response: This comment does not appear to be relevant. The cited study (Christy et al., 2001) does indeed present linear trends of differences between air and sea temperatures.

<u>Chapter 4 Comments and Responses:</u>

MacCracken CH4-1, Page 3, Line 86: Is not this difference now the major difference between the RSS and the UAH datasets? It may once have been a secondary contribution, but is it not now a major one?
Michael MacCracken, Climate Institute

Response: We disagree with this comment. In fact, the opposite is true for globally averaged data. Before the latest change in the UAH data (v5.1 to v5.2) the diurnal correction was a primary reason for the difference. While we have not yet performed a detailed analysis of the new UAH diurnal correction for TLT, it is in good agreement with the RSS correction of tropical land regions, despite the very different methods used to generate the two corrections.

MacCracken CH4-2, Page 4, Line 92-96: This seems like much too much a caving in to trying to be inclusive rather than to really be pointing out the corrections that have had to

be made and to critically be evaluating what seems to be most consistent with thorough consideration of each methodology. As indicated in the general comments, it would be helpful to have an appendix or table laying out the various corrections that have had to be made to the various versions of each data set, and what the effect of this has been on trends, etc. This is a serious issue, and such nice and polite puffery does not do the scientific advances justice (nor point to the mistakes that were made and the overly narrow claims about uncertainty in the past).
Michael MacCracken, Climate Institute

Response: The question of which satellite dataset is the most accurate, (and, in addition, whether or not recently discussed problems with the radiosonde record can explain the apparent discrepancies in the tropics) is still an open question subject to several different points of view that were represented on the author team. Different conclusions about the satellite data are reached depending on, in addition to other factors, one's assessment of the accuracy of radiosonde trends, and of the degree to which current models accurately reflect vertical transport of energy in the atmosphere. We are unable to make unambiguous, consensus statements at this time regarding the relative accuracy of the satellite data. The evolution of the satellite datasets is discussed briefly in Chapter 2.

MacCracken CH4-3, Page 5, Line 117-120: Of course there needs to be continuing work on the various data sets, but this opening statement also needs to make clear that the extensive testing and investigation that has gone on have made it so that the available data sets are quite useful. For example, the surface temperature record has been extensively examined and continues to show very strong surface warming over the past few decades, etc. This opening text almost makes it seem as if we do not yet have any useful datasets.
Michael MacCracken, Climate Institute

Response: The following text has been added: " have undergone extensive testing and analysis in an effort to make them useful tools for investigating Earth's climate during the recent past. In order to further increase our confidence in their use as climate diagnostics, they… "

MacCracken CH4-4, Page 5, Line 122-137: It would really be more useful to not only say that work needs to be done, but to provide some perspective on how important this type of effort would be compared to other investments of money. To a large extent, the supposed contradiction between surface and tropospheric temperature changes has been resolved, and so it would seem likely that other investments of funds would be more important (like working to better understand how extremes have and should be projected to change). The text here provides no context for making a judgment about how investment here might change the overall sense of what has and is projected to happen, and what benefit would come from doing what is suggested. Will it really matter?
Michael MacCracken, Climate Institute

Response: Given the charge of our report (to discuss problems associated with the vertical structure of temperature trends), we should provide recommendations that are important for solving this problem. It is outside our purview to compare the importance of our suggested solutions to these problems with work in other areas of climate research.

MacCracken CH4-5, Page 6, Line 149-151: It would also help to do some synoptic analyses over past periods to get a sense if the data are self-consistent. For example, during WW II, there were all sorts of problems with the taking the observations (for quite legitimate reasons), changing spatial coverage, etc.—and it would really be beneficial to determine the confidence that can be placed in our estimates of what happened during this period as it is a crucial tipping point in some data sets (from warming to cooling, etc.). Was all this real, or are there still problems with the data sets?
Michael MacCracken, Climate Institute

Response: the phrase " and from efforts to assess the self-consistency of historical data" has been added. The time period during WWII is outside of the time period that is the focus of this report.

MacCracken CH4-6, Page 10, Line 227-229: Given this uncertainty, how can there be much confidence in the notion of a well-defined regime shift in the mid-1970s (specifically 1976-77)? Might this all have been a confluence of normal fluctuations and the shift is all an artifact of how we are looking at it?
Michael MacCracken, Climate Institute

Response: The discussion here focuses on the data from a single radiosonde, not on the combined data that is used to argue for the 1976-1977 "climate regime change".

MacCracken CH4-7, Page 15, Line 336-338: I would think this should be I subjunctive tense—so say, "effects had on average" and "would have introduced".
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-8, Page 15, Line 339-341: Are these uses of the word "likely" really appropriate, especially in the second case? This is a word that IPCC has imbued with a special meaning, and it is not at all clear to me that following all these efforts we have more than 67% confidence that large biases remain. I would think the chances are a good bit lower. And see lines 915-923 which seem to suggest that most of the uncertainties are out of the datasets.
Michael MacCracken, Climate Institute

Response: The statement referred to here addresses the likelihood of large biased remaining in the records of individual radiosonde stations. We stand by our conclusion that it is likely that such biases remain. The comments in lines 915-923 refer to gridded surface temperature data, so there is no contradiction implied.

MacCracken CH4-9, Page 15, Footnote 4: "source of data" and "that has not yet"
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-10, Page 16, Line 361: Change to "the global means of the two radiosonde datasets are"
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-11, Page 16, Line 366-367: Change "are" to "have been" in two spots.
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-12, Page 18, Line 413-414: The use of "model" twice here is quite confusing. I would suggest saying "microwave radiative transfer algorithm"
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-13, Page 18, Line 418: How "accurately"—what does this mean—given some indication of the degree of agreement or disagreement.
Michael MacCracken, Climate Institute

Response: A footnote has been added to more completely describe the findings of Dai and Trenberth.

MacCracken CH4-14, Page 18, Line 420: Which model—two were mentioned above?
Michael MacCracken, Climate Institute

Response: The text has been changed to "atmospheric model" to "atmospheric component of the climate model"

MacCracken CH4-15, Page 19, Line 433-435: Awkward phrasing, having a "However" and a "but"
Michael MacCracken, Climate Institute

Response: The sentence has been changed to read "Although the removal of the diurnal cycle before merging may also introduce some error into UAH and RSS merging procedures if the assumed diurnal cycle is inaccurate, the removal of the diurnal harmonics before merging seems to be a more logical approach as the diurnal harmonics will tend to add noise unless removed."

MacCracken CH4-16, Page 21, Line 468: Change to "groups now remove"
Michael MacCracken, Climate Institute

Response: done

MacCracken CH4-17, Page 24, Line 531-532: Change to read "any overall assessment of uncertainties in the estimates of tropospheric"
Michael MacCracken, Climate Institute

Response: done

MacCracken CH4-18, Page 24, Line 536 and 538: Change "difference" to "differences"
Michael MacCracken, Climate Institute

Response: Line 536 -- Done; Line 538 -- added "a" before "difference"

MacCracken CH4-19, Page 23, Footnote 8: On line 6, change to "Earth". Also, page 29, line 622
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-20, Page 29, Line 624: The report needs to indicate which year the UAH group added in this correction to make clear that it was not done in their early data sets and so those papers should not be relied upon.
Michael MacCracken, Climate Institute

Response: A footnote has been added to make it clear that this adjustment was not performed prior to version D of the UAH data set. Also, a table that described the dates of changes to the various MSU datasets has been added to chapter 2.

MacCracken CH4-21, Page 31, Line 672: The word "now" needs to be changed to indicate when this change was made, so say "since 200?" this inconsistency has been addressed or something.
Michael MacCracken, Climate Institute

Response: The version number and date of introduction are now noted in this sentence.

MacCracken CH4-22, Page 36, Line 759: This is not really a "NASA" data set—it is from some particular scientists who should be cited.
Michael MacCracken, Climate Institute

Response: "NASA" has been changed to "the NASA group" -- the citation is at the end of the sentence

MacCracken CH4-23, Page 36, Line 764: In some megalopolises, the thermal emissions may also be large enough to be having an effect.
Michael MacCracken, Climate Institute

Response: To the extent that such changes affect the entire region in question, they are not a source of error, but part of the signal that should be modeled using land use change input in models. No changes made.

MacCracken CH4-24, Page 39, Line 820-822: Given the latest corrections and adjustments, this sentence should be turned around, indicating that the RSS data set is likely the most accurate—this voting technique that includes data sets that are not the most up-to-date seems really flawed.
Michael MacCracken, Climate Institute

Response: The sentence was inverted as suggested.

MacCracken CH4-25, Page 39, Line 824-826: This statement seems not to have taken into account the issue of surface-troposphere correlations being high mainly over NH continents and not elsewhere. So, why the "However"—is that result not just what one would expect?
Michael MacCracken, Climate Institute

Response: The paragraph was rewritten in response to both of the above comments. Here is the new paragraph:

On a global scale, one satellite dataset ($T_{2LT\text{-}RSS}$) suggests that the troposphere has warmed more than the surface, while both radiosonde datasets and one of the satellite datasets ($T_{2LT\text{-}UAH}$) indicate the opposite. The magnitude of these differences is less than the uncertainty estimates for any one data record. The situation is similar in the tropics. Both global and tropical averages of the radiosonde data contain many stations with less reliable data and metadata, which may be part of the cause for the surface-tropospheric differences. In contrast, in North America and Europe, where the most reliable radiosonde stations are located, the warming in the surface and lower troposphere appears to be very similar in all datasets.

MacCracken CH4-26, Page 39, Line 832: This use of the term "structural uncertainty" is really quite jargony—in simpler terms, it is saying that there are large uncertainties in going from the supposed observations to a validated dataset—and hiding this important finding in such terminology is not really very helpful to understanding the report's findings.
Michael MacCracken, Climate Institute

Response: The term "structural uncertainty" is introduced in several places in the report. The term is used to simplify wording -- we are not trying to hide anything. One of the main conclusions of the report is that the uncertainty in upper-air trends is dominated by this type of uncertainty. No changes made.

MacCracken CH4-27, Page 39, Line 833-835: The report should be saying which dataset is out of line—this is all a bit cryptic. Also make clear which has the latest and most widely accepted (published) corrections.
Michael MacCracken, Climate Institute

Response: Information has been added so that it is obvious which datasets are show warming/cooling relative to the surface.

MacCracken CH4-28, Page 40, Line 851-852: This treatment of "add datasets are equal" really has not been much of a service to the reader, for it does not clarify how much advance in understanding has occurred. It is thus helpful to have the discussions starting on line 857, and I would encourage more of that more critical type of analysis.
Michael MacCracken, Climate Institute

Response: The type of analysis proposed in lines 857 ff has not yet been performed. No changes made.

MacCracken CH4-29, Page 40, Line 860: Change "unsurprising" to "not surprising"
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH4-30, Page 42, Line 888: Change to "the apparent tropical" as it is really no longer real.
Michael MacCracken, Climate Institute

Response: Done

Swanson CH4-1, Page 28, Line 606-607 - The UAH group now uses a different diurnal correction method for their T_{2LT} product than the swath difference approach previously applied. UAH has switched to a grid point based diurnal correction instead of the zonal correction for the latest version of that product. Further discussion of the differences between the UAH approach and that of RSS would be useful.

Response: This is incorrect -- the new UAH correction is still a zone-by-zone correction, and independent of longitude within each latitude band. The latest UAH correction is not yet in the public domain, so a detailed discussion of the differences is not yet possible.

Swanson CH4-2, Page 30, Line 632 - The comparisons performed by UAH between radiosonde data and their T_{2LT} product make use of sonde data to simulate the output of their T_{2LT} algorithm (Spencer and Christy, 1992). Since this process uses the same algorithm on both sides of the comparison, there is no test of the validity of the algorithm itself. It should be noted that Christy and Spencer claimed good agreement between their older versions of the T_{2LT} and simulated sonde data, but now have produced a new version of their product, after a major correction. If the comparison was actually a valid test of the accuracy of the earlier versions T_{2LT}, why didn't the earlier data fail the test?

Response: The comparison to radiosonde data uses the deduced TLT vertical weighting function to weight the radiosonde results. In this sense, the comparison with radiosondes DOES test the TLT algorithm. In any case, errors in the atmospheric radiative transfer calculations are unlikely to have much effect on these results.

The major changes between UAH V5.1 and UAH V5.2 occur in the tropics, where there are not many sonde stations.

Robock CH4-1, p. 5, lines 124-125: Models cannot be considered as a reliable source of information about the diurnal cycle of air temperature. And such information should not be used to correct observed data.
Alan Robock, Rutgers University

Response: We agree that in a perfect world, we would not have to use a modeled diurnal cycle to adjust observed data. However, there is no current method that has been shown to do any better. The UAH V5.1 method that uses cross-track information has been shown to be very sensitive to satellite attitude errors. The new UAH v5.2 method is basically a very simple model constrained by observed data. Thus we see no evil in the use of a modeled diurnal cycle, as long as it is validated to the extent possible. The problems with the current RSS model-based method are well documented in the main text, and improving the specification of the diurnal cycle is among our recommendations. No changes.

Robock CH4-2, p. 5, line 127: Change "a satellite-borne sounder" to "satellite-borne sounders"
Alan Robock, Rutgers University

Response: Done

Robock CH4-3, p. 19, line 430: Change to "; Vinnikov et al., 2006)"
Alan Robock, Rutgers University

Response: Done

Robock CH4-4, p. 22, line 506: Change to "; Vinnikov et al., 2006)"
Alan Robock, Rutgers University

Response: Done

Robock CH4-5, p. 27, line 575: use 0.19 and 0.12 instead of 0.189 and 0.115.
Alan Robock, Rutgers University

Response: Done

Robock CH4-6, p. 30, line 630: At end of sentence add the following sentence: "The decay of orbital height does not affect any results of the Maryland group because they use nadir-only observations."
Alan Robock, Rutgers University

Response: The discussion here is for a "2LT" product. The Maryland group does not yet produce a 2LT product. None of the "T2" products are affected significantly by orbital decay. Added a footnote to this section to make this point.

Robock CH4-7, p, 47, lines 1089-1090: This paper is in press. The reference should be changed to:
Vinnikov, Konstantin Y., Norman C. Grody, Alan Robock, Ronald J. Stouffer, Philip D. Jones, and Mitchell D. Goldberg, 2006: Temperature trends at the surface and in the troposphere. J. Geophys. Res., in press, doi:10.1029/2005jd006392.
Alan Robock, Rutgers University

Response: Done

Swanson CH4-1, Page 28, Line 606-607 - The UAH group now uses a different diurnal correction method for their T_{2LT} product than the swath difference approach previously applied. UAH has switched to a grid point based diurnal correction instead of the zonal correction for the latest version of that product. Further discussion of the differences between the UAH approach and that of RSS would be useful.

Response: This is incorrect -- the new UAH correction is still a zone by zone correction, and independent of longitude within each latitude band. The latest UAH correction is not yet in the public domain, so a detailed discussion of the differences is not yet possible.

Swanson CH4-2, Page 30, Line 632 - The comparisons performed by UAH between radiosonde data and their T_{2LT} product make use of sonde data to simulate the output of their T_{2LT} algorithm (Spencer and Christy, 1992). Since this process uses the same algorithm on both sides of the comparison, there is no test of the validity of the algorithm itself. It should be noted that Christy and Spencer claimed good agreement between their older versions of the T_{2LT} and simulated sonde data, but now have produced a new version of their product, after a major correction. If the comparison was actually a valid test of the accuracy of the earlier versions T_{2LT}, why didn't the earlier data fail the test?

Response: The comparison to radiosonde data uses the deduced TLT vertical weighting function to weight the radiosonde results. In this sense, the comparison with radiosondes DOES test the TLT algorithm. In any case, errors in the atmospheric radiative transfer calculations are unlikely to have much effect on these results.

The major changes between UAH V5.1 and UAH V5.2 occur in the tropics, where there are not many sonde stations.

Trenberth CH4-1, Page 2, Line 62: nothing here on urban heat island effects.
Kevin Trenberth, National Center for Atmospheric Research

Response: Added parenthetical statement to make it clear the urban heat island effects have been considered.

Trenberth CH4-2, Page 2, Line 67: for SST the main issue are the adjustments about 1940-44.
Kevin Trenberth, National Center for Atmospheric Research

Response: This time period is outside the time period "the radiosonde era" that is the focus of this report.

Trenberth CH4-3, Page 3, Line 76: if they are homogenized then how come they have remaining errors? I.e. they are not homogenized. Also Page, 11, Line 252.
Kevin Trenberth, National Center for Atmospheric Research

Response: changed "homogenized" to "adjusted"

Trenberth CH4-4, Page 4, Lines 97-98: this is not merely "very likely" but certain.
Kevin Trenberth, National Center for Atmospheric Research.

Response: Changed to "It is virtually certain that most of the satellite-sonde discrepancy arises from uncorrected errors in the radiosonde data."

Trenberth CH4-5, Page 6, Line 159: no recommendation on reanalyses.
Kevin Trenberth, National Center for Atmospheric Research

Response: Reanalysis is de-emphasized in this report as it will be covered in another CCSP report (Product 1.3)

Trenberth CH4-6, Page 7, Line 178: Box 2.2? Where is it?
Kevin Trenberth, National Center for Atmospheric Research

Response: This now refers to Box 2.1, Chapter 2.

Trenberth CH4-7, Page 8, Line 196: "discussed in chapter 2" does not seem to be.
Kevin Trenberth, National Center for Atmospheric Research

Response: Removed "As discussed in Chapter 2"

Trenberth CH4-8, Page 11, Line 264: also do not cover zones, especially in southern hemisphere.

Kevin Trenberth, National Center for Atmospheric Research

Response: Added a footnote to note this fact.

Trenberth CH4-9, Page 11, Line 267: Please see Hurrell et al. (2000): errors in trends were found of up to 0.03ºC but individual monthly means could be in error by 0.2ºC. These numbers could be larger.
Kevin Trenberth, National Center for Atmospheric Research

Response: Changed value to 0.03 and added the Hurrell et al reference.

Trenberth CH4-10, Page 15, Line 350: also Hurrell et al. 2000.
Kevin Trenberth, National Center for Atmospheric Research

Response: Added reference

Trenberth CH4-11, Page 15, Line 355: see comment on Line 267.
Kevin Trenberth, National Center for Atmospheric Research

Response: Added reference

Trenberth CH4-12, Page 20, Line 455: The following section shows these conclusions are false. In Fig 4.1 there remain trends even when satellites are stable and not changing. Also land vs. ocean issues are not adequately addressed, especially for Africa (see Fig 4.3).
Kevin Trenberth, National Center for Atmospheric Research

Response: Reasons for remaining difference trends addressed in lines 481-486. Added a phrase at this location ", which in addition to their direct effect on the diurnal correction, also lead to large changes in the temperature of the calibration target."

Trenberth CH4-13, Page 20, Line 461: Here the figure 3 in appendix could go.
Kevin Trenberth, National Center for Atmospheric Research

Response: Difference plots are available in Fig 4.1. No changes

Trenberth CH4-14, Page 23, Line 511: what about the use by Univ. of MD. of only nadir soundings?
Kevin Trenberth, National Center for Atmospheric Research

Response: Added a sentence to describe this difference.

Trenberth CH4-15, Page 27, Line 584: yes, see especially Africa.
Kevin Trenberth, National Center for Atmospheric Research

Response: No changes made.

Trenberth CH4-16, Page 29, Line 608: This is important. These problems with UAH were identified by Hurrell and Trenberth (1998) over Africa and they still remain in UAH. Surface emissions are important and the diurnal cycle can be 30ºC.
Kevin Trenberth, National Center for Atmospheric Research

Response: Added a reference to Hurrell and Trenberth.

Trenberth CH4-17, Page 31, Line 672: "now resolved" not clear that this is true.
Kevin Trenberth, National Center for Atmospheric Research.

Response: Within error bars, this is true. Both this statement and the statement that the RSS TLT results are consistent are unpublished, but can be clearly seen by looking at the tropical trends. No changes.

Trenberth CH4-18, Page 34, Line 718: what is base period for Figure plots?
Kevin Trenberth, National Center for Atmospheric Research

Response: Added this information to the figure caption.

Trenberth CH4-19, Page 35, Line 736, and Page 36, Lines 753-760: urban heat island issues not adequately addressed.
Kevin Trenberth, National Center for Atmospheric Research

Response: Added additional material about the relative magnitude of urbanization effects on large spatial scale averages.

Trenberth CH4-20, Page 39, Line 818: why no figures here?
Kevin Trenberth, National Center for Atmospheric Research

Response: We added figure 4.5, which shows maps of trend differences between the surface and the two satellite-derived T_{2LT} datasets.

Trenberth CH4-21, Page 40 Line 851: Strongly disagree with this philosophy. The datasets are not all equal.
Kevin Trenberth, National Center for Atmospheric Research

Response: We also agree that it not ideal. Currently we lack tools and methods to make an unambiguous statement about which datasets are closer to reality. We make a number of suggestions about how to make progress in the following section.

<u>Chapter 5 Comments and Responses:</u>

Douglass CH5-1, P6, L116-121, **Quote from report:** "A second explanation is that remaining errors in some of the observed tropospheric data sets adversely affect their long-term temperature trends. The second explanation is more likely in view of the model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming." **Comment:** The "Uncertainties" are among the observations and are not large enough to include the mean of the models. Choosing the 2nd explanation is not convincing.

Response: The Reviewer's comment is incorrect. We assume that he is referring to the mean of the sampling distribution of model-estimated trends in tropical lapse rates (defined here as T_S minus T_{2LT}). The mean value of the model sampling distribution is given as –0.06°C/decade in Table 5.4B [Page 112]. This mean value was calculated as described in the caption of Table 5.4A [Page 109]. The RSS tropical T_{2LT} trend over 1979 to 1999 (the same period over which model T_S and T_{2LT} trends were calculated) is +0.128°C/decade. The observed tropical T_S trends in the NOAA, NASA, and UKMO datasets are 0.125°C/decade, 0.125°C/decade, and 0.137°C/decade, respectively. This yields observed tropical lapse-rate trends involving the RSS T_{2LT} dataset that range from –0.003°C/decade to +0.009°C/decade, depending on the choice of the observed T_S dataset. Observed tropical lapse-rate trends involving the other three primary T_{2LT} datasets (the UAH satellite data, and the HadAT2 and RATPAC radiosonde data) are invariably positive, as is mentioned in the Chapter [Page 111, column 2, para. 2] and shown in Figure 5.4G.

The Reviewer's comment does not account for parametric and structural uncertainties in the individual datasets. For example, as is now mentioned in the new footnote 60 [Page 111], the RSS group claims a 2σ uncertainty of ±0.09°C/decade on their tropical T_{2LT} trend. This uncertainty arises from statistical uncertainty in the RSS regression approach, from uncertainty in the choice of target factor, and from uncertainty in the diurnal cycle correction. Accounting for these uncertainties in the RSS tropical T_{2LT} trend (while keeping the observed T_S trends unchanged) leads to RSS-based T_S minus T_{2LT} trends that range from –0.093°C/decade to +0.18°C/decade. This range does incorporate the mean value of the model sampling distribution (–0.06°C/decade). So the Reviewer's assertion is incorrect.

Furthermore, recent research by *Sherwood et al.* (2005) and *Randel and Wu* (2006) suggests that previous work (and the present report!) may have underestimated the true magnitude of structural uncertainties in radiosonde-derived tropical T_{2LT} trends. Both studies provide evidence of a residual cooling bias in tropical radiosonde data. Removal of this bias yields observed tropical T_{2LT} trends that are larger than the surface trends. Such behavior is consistent with the RSS T_S minus T_{2LT} trends, and would likely expand the range of observational uncertainty shown in Figure 5.4G.

Finally, we point out in the new footnote 59 [Page 111] that the UMd group does not produce either a T_{2LT} or T_4 product. Because of this, UMd results could not be used in comparisons of modeled and observed trends in T_S minus T_{2LT} or T_S minus T^*_T.

Assuming that the relationships between the UMd T_2, T_{2LT} and T^*_T trends were similar to those for the UAH and RSS data, the UMd data would yield T_{2LT} and T^*_T trends that were larger than in RSS. Once again, this would expand the range of observational uncertainty for tropical lapse-rate trends in Figures 5.4F and G.

We also note that we have slightly changed the language in the text cited by the Reviewer. The revised text [Page 90, Key Finding 6, bullet 5] now reads:

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational data sets, leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report (model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and the independent physical evidence supporting substantial tropospheric warming) favors the second explanation".

Instead of "favors the second explanation", the public review version stated that the second explanation was "more likely". Use of the new phrase "favors the second expression" is a simple, factual description of the majority opinion of the Lead Authors of this Report, and does not express any value judgment regarding the relative likelihood of the two posited explanations.

There is now a new, sixth bullet of Key Finding 6 [Page 90]. This new bullet injects a note of caution by pointing out that the reasons for discrepancies between model and observed tropical lapse-rate trends are not fully understood at present.

Bottom line: The Reviewer is incorrect in stating that observational uncertainties "are not large enough to include the mean of the models". We have clarified this point with the addition of footnotes 59 and 60. The changes to Key Finding 6 are also a direct response to the Reviewer's concerns.

Douglass CH5-2, P40, L797, **Quote from report:** "***New Comparisons of Modeled and Observed Temperature Changes".*** **Comment:** This is unpublished and not reviewed. Work for IPCC-AR4. Inappropriate.

Response: This comment is incorrect. Most of the work discussed in Section 5 is in the peer-reviewed literature, in a paper by *Santer et al*. published in *Science* in 2005.[1] For example, Figures 5.2A, B, and C in Section 5 are modified versions of Figures 1A, B, and C in *Santer et al*. (2005). Figure 5.4 in Section 5 is a modified version of Figure 2 in *Santer et al*. (2005). Figure 5.6 in Section 5 is a modified version of Figure 4 in *Santer et al*. (2005). Spatial maps of the temperature difference between T_S and T_{2LT} in various model and observational datasets (Figure 5.5 in Section 5) are not earth-shattering new results requiring independent corroboration! Indeed, the observational results in panels E and F of Figure 5.5 simply replicate information that has been published previously in

[1]Santer, B.D., *et al*., 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 284-287.

Figures 3D and E of the *Mears and Wentz* (2005) *Science* paper[2]. Likewise, zonal-mean profiles of simulated and observed temperature changes in the free atmosphere have been published in many different peer-reviewed sources, such as *Tett et al.* (1996)[3] and *Hansen et al.* (2005)[4]. It is entirely appropriate for Section 5 to comment on updated versions of previously-published material.

Douglass CH5-3, P45, L883-888, **Quote from report:** "The model ensemble encapsulates uncertainties in climate forcings and model responses ... The observational range characterizes current structural uncertainties in historical changes. ... Our goal here is to determine where model results are qualitatively consistent with observations, and where serious inconsistencies are likely to exist." **Comment:** Uncertainties and ranges are not carefully defined -- max minus min?; 1- sigma; 2- sigma?.

Response: This comment is incorrect. The "structural uncertainties" in the observed surface and atmospheric temperature changes are clearly and carefully defined in Chapter 2. These uncertainties arise from the different methods that analysts employ in their attempts to generate homogeneous Climate Data Records (CDRs) from raw observational data. The observed structural uncertainties are clearly shown in Figures 5.3, 5.4, and 5.6, where there is one discrete point for each individual observational dataset (or for each pair of T_S and upper-air datasets in the case of the observed lapse-rate trends shown in Figures 5.3F,G and 5.4F,G). It is immediately obvious, upon even cursory inspection of these Figures, that the "observational range" referred to by the Reviewer is indeed a range, and not a standard deviation!

Likewise, the derivation of the model histograms in Figures 5.3 and 5.4 is clearly explained in the caption of Figure 5.3. We provide some simple statistics of "the model ensemble" in Tables 5.4A and B. The derivation of these statistics (mean, median, standard deviation, maximum, and minimum) is clearly explained in the caption of Table 5.4A [Page 109].

As stated on Page 105 [column 1, para. 2], the model results analyzed here constitute an "ensemble of opportunity". This a finite sample. It is not clear whether it is also a representative sample, and whether it can be used to make rigorous statistical inferences. This issue is now addressed in the new footnote 45, which has been added to the text at the point referred to by the Reviewer's comment [Page 107]:

"The 49 20CEN realizations analyzed here are a very small sample from the large population of results that could have been generated by accounting for existing

[2]Mears, C.A., and F.W. Wentz, 2005: The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science*, **309**, 1548-1551.

[3]Tett, S.F.B., *et al*., 1996: Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science*, **274**, 1170-1173.

[4]Hansen, J., *et al*., 2005a: Efficacy of climate forcings. *Journal of Geophysical Research (Atmospheres)*, **110**, D18104, doi:10.1029/2005JD005776.

uncertainties in physics parameterizations and historical forcings (*e.g.*, Allen, 1999; Stainforth *et al*., 2005). Likewise, the observational datasets that we consider in this report probably only capture part of the true "construction uncertainty" inherent in the development of homogeneous climate records from raw temperature measurements. We do not know *a priori* whether temperature changes inferred from these small samples are representative of the true temperature changes that would be estimated from the much larger (but unknown) populations of model and observational results. This is another reason why we are cautious about making formal assessments of the statistical significance of differences between modeled and observed temperature trends. We do, however, attempt to characterize some basic statistical properties of the model results (see Tables 5.4A,B)".

Douglass CH5-4, P47, L916, **Quote from report:** "simple weighting function approach (Box 2.2)." **Comment:** No definition of the "simple" weighting function. There is no Box 2.2

Response: This box is in Chapter 2. "Chapter 2" has now been added to avoid confusion [Page 105, para 3].

Douglass CH5-5, P48 L936, **Quote from report:** "Figure 5.3:" **Comment:** Values of the data that was used to make these histograms are not available.

Response: This is a serious criticism. It is also an invalid criticism. Dr. Douglass first wrote to the CLA of Chapter 5 (Dr. Ben Santer) on December 2, 2005. In an email to Dr. Santer, Douglass requested:

"…the data table from which the histograms were made".

The reference here was to the histograms displayed in Figures 5.3 and 5.4 of Chapter 5.

Dr. Santer replied by email on the same date (December 2, 2005). He noted that:

"The IPCC data that I've used in generating figures 5.3 and 5.4 are freely available to scientific researchers. You are welcome to request these data from the IPCC and independently repeat my calculations of synthetic MSU temperatures, *etc*."

In an email to Tom Karl dated December 15, 2005, Dr. Douglass complained about Dr. Santer's email reply of December 2, 2005. Dr. Douglass wrote that:

"The essence of scientific research is verifiability by other scientists of scientific claims. Santer should be willing and eager to have me examine his claims. To be secretive or to hold back supporting material invites suspicions as to the validity of his claims."

In fact, Dr. Santer was quite willing for Dr. Douglass to perform independent verification of the calculations on which Figures 5.3 and 5.4 were based. In an email to Dr. Douglass dated December 20, 2005, Dr. Santer noted that:

"If you wish to independently verify the calculations that I made in order to generate Figures 5.3 and 5.4 of the CCSP Report, you will need to start with the raw surface and atmospheric temperature data. Those data are freely available to you. Algorithms for generating synthetic MSU temperatures, or for calculating T_{Fu}, are freely available in the published literature.[5] As a competent climate scientist, calculations of synthetic MSU temperatures or T_{Fu} should be well within your capabilities. These calculations should require weeks rather than months to complete (assuming basic competency in atmospheric and computational science)."

"You have all of the information you need in order to reproduce the results shown in Figures 5.3 and 5.4. You know which models were used. You know which set of forcings was used by each modeling group. You know exactly which realizations were used (see below). You seem to be laboring under the misapprehension that – for any given model – different forcings are used for different realizations of the 20th century experiment. This is not the case. For a given model's 20th century experiment, forcings do not vary from realization to realization. The only variation between realizations is in the initial conditions of the coupled atmosphere-ocean system."

"You are now in possession of all the information you need to independently verify the results in Figures 5.3 and 5.4. Personally, I would welcome independent verification of my calculations. I don't see what else there is to "verify"."

Bottom line: The model runs from which Figures 5.3 and 5.4 were derived are publicly available via the IPCC model data archive held at Lawrence Livermore National Laboratory (see http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php). Dr. Douglass had full access to this data. Dr. Santer provided Douglass with full details of the 49 model 20CEN runs used for generating the model histograms in Figures 5.3 and 5.4. Dr. Douglass also had full access to the published algorithms used to calculate synthetic Microwave Sounding Unit temperatures and T_{Fu} temperatures from model data. He could, with a modicum of effort on his part, have attempted an independent verification of the results given in Figures 5.3 and 5.4. Dr. Douglass did not do so. The criticism is invalid.

Douglass CH5-6, P54 L1027, **Quote from report:** "Figure 5.4:" **Comment:** Values of the data that was used to make these histograms are not available.

Response: See **Response to Douglass CH5-5**.

Douglass CH5-7, P55, L1036-1038, **Quote from report:** "The RSS trends are just within the range of model solutions. Tropical lapse-rate trends in both radiosonde datasets and in the UAH satellite data are always positive (larger warming at the surface than aloft), and lie outside the range of model results." **Comment:** All the observations including RSS are 2 sigma or more away from the mean of the models. If you choose the

[5]Methods for calculating synthetic MSU temperatures from model or reanalysis data are discussed in Chapter 2, Box 2.1. Calculation of "T_{Fu}" temperatures (*i.e.*, what our report refers to as T^*_T and T^*_G) is described in Chapter 2.

range to mean 2-sigma, then you can catch RSS. However, if more than 49 simulations were chosen [82 are available], then the sigma value of the models would likely be less and RSS would be more than 2-sigma away.

Response: The first two sentences of this comment contradict each other! If all the "observations including RSS are 2 sigma or more away from the mean of the models", how can you "catch RSS" if "you choose the range to mean 2-sigma"?

In fact, the RSS-derived tropical lapse-rate trends shown in Figures 5.4F and G are (in 3 out of 4 cases) within 2σ of the model average result. This is stated in footnote 60 [Page 111]. Consider first the results for trends[6] in T_S minus T^*_T in Figure 5.4F:

RSS T^*_T trend:	+0.155°C/decade
HadCRUT2v T_S trend:	+0.137°C/decade
NOAA T_S trend:	+0.125°C/decade
RSS T_S minus T^*_T (HadCRUT2v T_S):	–0.018°C/decade
RSS T_S minus T^*_T (NOAA T_S):	–0.030°C/decade
Model average T_S minus T^*_T (from Table 5.4B) :	–0.080°C/decade
Model 1σ T_S minus T^*_T (from Table 5.4B):	0.040°C/decade

So the "range" spanned by the model average trend, $\pm 2\sigma$, extends from –0.160°C/decade to 0.0°C/decade. This range encompasses both RSS T_S minus T^*_T trends.

For tropical lapse-rate trends based on T_{2LT}, the situation is as follows:

RSS T_{2LT} trend:	+0.128°C/decade
HadCRUT2v T_S trend:	+0.137°C/decade
NOAA T_S trend:	+0.125°C/decade
RSS T_S minus T_{2LT} (HadCRUT2v T_S):	+0.009°C/decade
RSS T_S minus T_{2LT} (NOAA T_S):	–0.003°C/decade
Model average T_S minus T_{2LT} (from Table 5.4B):	–0.060°C/decade
Model 1σ T_S minus T_{2LT} (from Table 5.4B):	0.030°C/decade

So the "range" spanned by the model average trend, $\pm 2\sigma$, extends from –0.120°C/decade to 0.0°C/decade. This range encompasses one of the two RSS T_S minus T^*_{2LT} trends.

Irrespective of the Reviewer's erroneous statement that "All the observations including RSS are 2 sigma or more away from the mean of the models", we note that the Reviewer's comments fail to account for the large structural uncertainty in the RSS tropical T_{2LT} trend. This issue has already been addressed in detail in the **Response to Douglass CH5-1**. Furthermore, the observed lapse-rate trends did not include information from UMd, an issue also dealt with in the **Response to Douglass CH5-1**. Inclusion of UMd data would likely expand the range of observational uncertainty.

[6] All trends were calculated over 1979 to 1999.

As the Reviewer points out, there are now 82 realizations of 20CEN runs in the IPCC AR4 archive held at the Program for Climate Model Diagnosis and Intercomparison (PCMDI). At the time of preparation of this report, only 49 realizations were available. Using the 82 runs would involve data that was "unpublished and not reviewed", a concern that this reviewer expressed earlier (**Douglass CH5-2)**. The issue of data availability is now addressed in the new footnote 42 [Page 105].

Finally, we note that the Reviewer either misread or overlooked information on how we calculated standard deviations from the model results. This information is provided in the caption of Table 5.4 [Page 109]. The standard deviations provided in Tables 5.4A and B are based on sample sizes of $n = 19$ (the number of climate models available), not $n = 49$ (the total number of 20CEN realizations available)! This avoids placing too much weight on a single model with a large number of realizations. Had we had access to all 82 20CEN realizations that are currently in the IPCC archive (at the time of writing this report), standard deviation estimates would have been based on sample sizes of $n = 23$, not $n = 82$! So the Reviewer's musings regarding the effect of a large increase in sample size on standard deviation values are incorrect.

Douglass CH5-8, P64, L1186, **Quote from report:** "Fig 5.7" **Comment:** Values of the data that was used to make these maps are not available.

Response: See **Response to Douglass CH5-5**. The "maps" referred to by the Reviewer are actually zonal-mean profiles of atmospheric temperature change (not maps!) The data used for calculating these zonal-mean trend profiles were part of the IPCC AR4 archive held at PCMDI, and were readily available to Dr. Douglass.

Kheshgi CH5-1, Page 6, Line 120: It is not clear what is meant by "independent physical evidence"? Suggest that whatever is meant by this be referred to here (e.g. the section of this report where it is discussed). -- *Haroon Kheshgi, ExxonMobil Research & Engineering Company*

Response: This independent physical evidence is discussed within the Chapter (in Section 6). It includes recent increases in tropospheric water vapor and tropopause height, and accelerated retreat of high-elevation tropical glaciers.

We would prefer not to provide details of this independent evidence in the "Key Findings and Recommendations" Section, which is supposed to be brief. Nor do we think it is appropriate to provide (in Key Finding 6) an explicit reference to Section 6. If we did this, then other "Key Findings and Recommendations" would also have to refer forwards to relevant portions of the underlying Chapter. In our opinion, this would detract from the principal results we are trying to convey in the "Key Findings and Recommendations".

Kheshgi CH5-2, Pages 10-25, Lines 498-796: Section 4 (chapter 5) seems to switch back and forth between a general assessment of detection and attribution and specific assessment on reconciling trends with observations. While detection and attribution generally is an important topic for assessment, it is not the topic of this assessment and should not be assessed here. It may also be useful to consider the roles of this SAP in assessing detection and attribution Vs. SAP1.3 which includes attribution in its title, and is clearly in its scope. A general challenge for this section is covering all the identified uncertainties and gaps in our understanding, and seeing how comparisons between models and data may indicate issues to be reconciled. The conclusions of this section seem in contrast with those of section 7 of this chapter which indicates difficulties. -- *Haroon Kheshgi, ExxonMobil Research & Engineering Company*

Response: We disagree with the Reviewer's comment. Our charge was to consider the causes of recent temperature changes at the Earth's surface and in the free atmosphere. We have tried to evaluate and assess the scientific literature relevant to this charge. Clearly, detection and attribution studies – which use rigorous statistical methods to investigate the causes of climate change – are highly relevant to this Chapter, and an integral part of it. We do not understand how or why the Reviewer can claim that detection and attribution work "…is not the topic of this assessment".

The Synthesis and Assessment Product referred to by the Reviewer (SAP1.3) deals with reanalysis products only, and thus will not cover most of the detection and attribution studies that are assessed here. The majority of the detection and attribution studies that we consider seek to understand the causes of temperature changes in observational satellite, radiosonde, and surface temperature data.

We do not understand what point the Reviewer is trying to make in the sentence beginning "A general challenge for this section…" It is indeed challenging to perform a comprehensive assessment of the many studies that have attempted to understand the nature and causes of recent surface and atmospheric temperature changes. We have tried hard to identify "uncertainties and gaps in our understanding", to identify what we know and what we do not know, and to be fair and balanced in our assessment.

We disagree with the Reviewer's comment that Sections 4 and 7 seem to reach contradictory conclusions.

Kheshgi CH5-3, Page 37, Line 733: Simply looking at Figure 1 of the Exec Summary shows a clear correlation between volcanoes and stratospheric warming. This should be noted here, since the existing paragraph taken out of the full context of the reference (which is not given) suggests this effect is unclear. -- *Haroon Kheshgi, ExxonMobil Research & Engineering Company*

Response: The study referred to by the Reviewer (*Thorne et al.*, 2003) deals with identification of volcanic effects in tropospheric temperatures – not in stratospheric temperatures! It would be inappropriate, therefore, to mention volcanic effects on stratospheric temperatures at this point in the text. We discuss volcanically-induced warming of the stratosphere in a number of places in Chapter 5. Examples include Key

Finding 1, bullet 5 [Page 89]; Page 109 [column 1, para. 2, and column 2, para. 1]; and footnote 49.

We note that while volcanic effects on stratospheric temperatures are immediately obvious, volcanic effects on tropospheric temperatures are less easily identifiable in fingerprint detection studies. This is in part because of the effects of ENSO variability, which obscures much of the tropospheric cooling signal associated with the 1982 El Chichón eruption, and some of the tropospheric cooling caused by the 1991 Pinatubo eruption (see footnotes 2 and 52). A further complication is the need to reduce dimensionality in detection and attribution studies (see Box 5.5). This means that decadal averages are often used, which "smear out" short-term (3-5 year) volcanic effects on tropospheric temperature.

We do not understand why the Reviewer states that the reference "is not given". It is very clear that the paper by *Thorne et al*. (2003) is being referred to throughout this paragraph.

No changes made.

Kheshgi CH5-4, Page 39, Lines 789-791: The conclusion given that radiosonde records give "strong" evidence for attribution raises some questions as to what this statement means. Strong is a relative term. Is evidence from radiosondes stronger than evidence from surface temperature? Nearly all detection and attribution studies focus on surface temperature records. Also, given all the considerations that go into the real uncertainty of radiosonde-based estimates of atmospheric temperature changes, how strong can any attribution conclusion be (that is based just on radiosondes)? Finally, suggest that summary conclusions be grouped so that there are not multiple sets of conclusions that may decrease transparency (*e.g*. at the front of this chapter in key findings, here in the middle, and at the end in the concluding section 7. -- *Haroon Kheshgi, ExxonMobil Research & Engineering Company*

Response: We feel that use of the word **"**strong" is justifiable here. D&A analysts have used a variety of different fingerprint techniques, methodological choices, and model and observational datasets. Despite these differences, the finding of a statistical significant anthropogenic signal remains robust. This is at least partly due to the fact that different radiosonde datasets all show a qualitatively similar pattern of tropospheric warming and stratospheric cooling from the 1960s to the present.

The Reviewer's comment suggests that attribution conclusions (involving temperature changes in the free atmosphere) are based solely on studies that have searched for model-predicted climate-change "fingerprints" in observational radiosonde data. This is not the case. Some D&A studies have successfully identified anthropogenic fingerprints in observational satellite data (*e.g.*, *Santer et al*., 2003b).

As we point out in Recommendation 4 [Page 91], the D&A studies that are assessed in Chapter 5 need to be repeated "with the new generation of model and observational data sets" described in the present Report.

Summary conclusions are necessary both within the text and up-front. They are internally consistent (we have checked).

MacCracken CH5-1, Page 2, Line 44-45: Although the burdens of aerosols are regional, their cooling influence can be experienced more than regionally, and were aerosols the only forcing, would cause global cooling, though most strongly in the region of the aerosols. With multiple forcings, sulfates may only cause cooling in the region of the aerosols, but that does not seem to be what the statement is about. Thus, I would urge a rewrite indicating their regional forcing, but wider scale cooling influence as the rest of the world's atmosphere responds.

Response: While it is true that maximum cooling generally occurs nearest the source regions, this is a complex issue, which we would prefer not to get into in the limited space available to us in the Key Findings. For example, the surface temperature changes at sea-ice margins (due to changes in sea-ice extent that may arise from natural internal variability alone, and are unrelated to regional changes in sulfate aerosol forcing) may be as large or larger than the temperature changes in aerosol source regions. We prefer to keep the original text of the bullet.

MacCracken CH5-2, Page 2, Line 47-48: Volcanic eruptions do not cool the surface for all seasons in all locations and for all injection latitudes and times. A more nuanced statement is needed.
Michael MacCracken, Climate Institute

Response: We disagree. We feel that the changes requested by the Reviewer are too specific for the "Key Findings and Recommendations" section. More detailed discussion of the surface and atmospheric temperature response to volcanic eruptions is given in the main text, and in the references cited therein [*e.g.*, in footnote 2 on page 94, and in the first two paragraphs of Section 6 on Pages 116 and 177]. No change made.

MacCracken CH5-3, Page 3, Line 49-50: This statement should say that the warming influence is also global.
Michael MacCracken, Climate Institute

Response: Done. This now reads: "Increases in solar irradiance warm globally throughout the atmospheric column (from the surface to the stratosphere)." [Page 89, Key Finding 1, bullet 5]

MacCracken CH5-4, Page 3, Line 52-54: I would suggest a stronger phrasing: Results from many different fingerprint studies convincingly indicate that the best explanation for the observed changes over the second half of the 20th century is that there has been a strong human influence on the three-dimensional structure of atmospheric temperature.
Michael MacCracken, Climate Institute

Response: Given the currently-large uncertainties in the upper-air observations (discussed in the preceding Chapters) we feel that the statement is fair as originally written. No changes made.

MacCracken CH5-5, Page 3, Line 65-66: This phrasing is quite misleading, implying that natural factors could have played up to an almost full explanation of the warming when they seem unable to explain virtually any of it. Thus, get rid of this artful statistical jargon, and say this more clearly, something like: Although natural factors have likely had modest influences on surface and atmospheric temperatures over the past 50 years, their influence are not nearly adequate to explain the observed changes.
Michael MacCracken, Climate Institute

Response: We feel that the existing phrasing is appropriate, particularly in view of the currently-large uncertainties in the upper-air observations. We are not using "artful statistical jargon". No changes made.

MacCracken CH5-6, Page 4, Line 75-77: Because of the time varying influences of both natural and human influences, it is really not clear to me why support should be given to linear trend comparisons. The forcings have varying time and space patterns, and their interactions will lead to changes that are nonlinear and not even monotonic. Linear analyses have been seriously abused in some studies (*e.g.*, Pat Michaels' trend extrapolation, and his subdividing the record in 1976-77, etc.) and really should not be encouraged. Also, there dependence on end points and sometimes a few outliers can be misleading. I would urge a statement saying that great care must be taken with such simple analyses.
Michael MacCracken, Climate Institute

Response: There is an entire Appendix devoted to the issues raised by the Reviewer. We do not think it is necessary to go into these issues here. To address the Reviewer's concerns, we have modified the text slightly. The sentence immediately before Key Finding 4 now reads:

"Linear trend comparisons are less powerful than "fingerprinting" for studying cause-effect relationships, but when treated with caution can highlight important differences (and similarities) between models and observations." [Page 90]

We believe that this statement is entirely justifiable based on the trend comparisons presented in Section 5. These trend comparisons <u>have</u> been helpful in identifying consistencies and inconsistencies between modeled and observed temperature changes. We also note that the subject of important temporal variations in anthropogenic signal patterns – and the failure of linear trends to capture such variations – is discussed in some detail in Chapter 5 [Page 103, column 2, paragraphs 2 and 3; Page 104, column 1, para. 1]

MacCracken CH5-7, Page 5, Line 102-103: Is the effect not more correctly stated as being due to the Clausius-Clapeyron relationship—so water content is very non-linear with temperature?
Michael MacCracken, Climate Institute

Response: The current text is correct "as is". The nonlinear relationship between temperature and atmospheric water vapor is discussed elsewhere within the Chapter [Page 117, column 2, para. 1]. No changes made.

MacCracken CH5-8, Page 5, Line 109-111: Science is not really a voting proposition—it should be based on the best representation and account of the various important physical relationships, etc. And when a distribution of results is given, only the most up-to-date datasets should be included after all the best attempts at corrections have been made. If others are to be mentioned, their shortcomings should also be indicated.
Michael MacCracken, Climate Institute

Response: It is not clear what point the Reviewer is trying to make here. There is no "voting" in the statement referred to by the Reviewer: "For longer-timescale temperature changes over 1979 to 1999, only one of four upper-air data sets has larger tropical warming aloft than in the surface records. All model runs with surface warming over this period show amplified warming aloft." [Page 90, Key Finding 6, bullet 4]

We are simply reporting on results here, with no value judgment on our part. We compared model-estimated surface and atmospheric temperature changes with results from four state-of-the-art observational datasets. These are the same observational datasets that have been used in earlier Chapters. Their limitations are discussed in depth in Chapter 4. The expert judgment of our group is given in 5th bullet of Key Finding 6. This states that:

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational data sets, leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report (model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and the independent physical evidence supporting substantial tropospheric warming) favors the second explanation".

MacCracken CH5-9, Page 9, Line 200: These references seem a bit out of date, given advances since then. Are there not any more recent references?
Michael MacCracken, Climate Institute

Response: The sentence referred to by the Reviewer is the following one: "However, models also have systematic errors that can diminish their usefulness as a tool for interpretation of observations (*Gates et al.*, 1999; *McAvaney et al.*, 2001)." [Page 92, column 2, para. 1] Both papers cited here are entirely appropriate references. Many of the systematic errors that they discuss (*e.g.*, model cold biases in the vicinity of the polar night jet, split ITCZ, *etc.*) are still manifest in current models. No change made.

MacCracken CH5-10, Page 13, Line 275-284: The report, as near as I could find, does not really provide enough background on the chaotic nature of the climate and the potential for there to be multiple realizations. This paragraph provides just a hint at this, but the report (and this paragraph) does not really explain in enough detail and generality that the real world set of observations is only one realization, and we run the models multiple times to get a possible distribution, and so the comparisons will not be exact, etc. [An indication of this not being sufficiently explained occurred in the Preface where it was implied that the models should "replicate" the observations. No indication of how well this should be expected to be done was provided.
Michael MacCracken, Climate Institute

Response: We disagree. We believe that this issue is adequately covered throughout Chapter 5. Here are a few examples:

⇒ "Because the climate system is chaotic, fully coupled models of the atmosphere and ocean cannot simulate exactly the same sequence of individual weather events that occurred in the real world (see Section 2). Such models can, however, capture many of the statistical characteristics of observed weather and climate variability…" [Page 92, Box 5.1, para. 2]

⇒ "We refer to these subsequently as "20CEN" experiments. Since the true state of the climate system is never fully known, the same forcing changes are applied n times, each time starting from a slightly different initial climate state. This procedure yield *n* different realizations of climate change. All of these realizations contain some underlying "signal" (the climate response to the imposed forcing changes) upon which are superimposed *n* different manifestations of "noise" (natural internal climate variability)." [Page 94, column 1, first complete paragraph]

⇒ "This illustrates the need for caution in comparisons of modeled and observed atmospheric temperature change. The differences evident in such comparisons have multiple interpretations… They may also be due to different manifestations of natural variability noise in the observations and a given CGCM realization." [Page 96, column 2, para. 4; Page 97, column 1, para. 1]

⇒ "In addition to model forcing and response uncertainty, the 20CEN ensemble also encompasses uncertainties arising from inherently unpredictable climate variability (Boxes 5.1, 5.2). Roughly half of the modeling groups that submitted 20CEN data performed multiple realizations of their historical forcing experiment (See Section 2 and Table 5.1)… Such multi-member ensembles provide valuable information on the relative sizes of signal and noise." [Page 105, column 2, para. 1]

⇒ "The model ensemble encapsulates uncertainties in climate forcings and model responses, as well as the effects of climate noise on trends." [Page 106, column 2, para. 1]

Note that the first example given above [para. 2 in Box 5.1] is new, and was included in order to address the Reviewer's concerns.

MacCracken CH5-11, Page 19, Line 372-375: The IPCC's use of low and very low levels of scientific confidence were really quite confusing in that there was really no separation out of when this mattered and when it did not. For example, while IPCC says that the level of confidence for solar is very low, we actually have quite useful observations covering two decades indicating the relative size of the influence, and only very limited indications that the influence was ever much bigger—and the influence is relatively small compared to human influences. Similarly for contrails, etc.—so while LOSU may be very low, it is not clear that this matters. Somehow, the text here needs to be identifying when the level of uncertainty in scientific understanding really makes a difference to the situation at hand.
Michael MacCracken, Climate Institute

Response: We do not have the expertise to evaluate whether or not forcings for which we currently have a low "Level Of Scientific Understanding" (LOSU) are important or unimportant for the specific scientific problem we are considering. We are very careful and circumspect in what we say in Section 3. We prefer not to make value judgments on the relative importance of different forcings, particularly since we currently lack comprehensive single-forcing experiments with such factors indirect aerosol effects, carbonaceous aerosols, *etc*.

MacCracken CH5-12, Page 25, Footnotes: It is not clear why the various footnotes have the same number.
Michael MacCracken, Climate Institute

Response: Microsoft word bug. Now fixed.

MacCracken CH5-13, Page 26, Line 511: It is not out of the realm of possibility that the ENSO variations have a human influence—this might at least be footnoted as a possibility.
Michael MacCracken, Climate Institute

Response: We have added a footnote to Box 5.1 [Page 92]. This footnote states that "There is some evidence that human-induced climate change may modulate the statistical behavior of existing modes of climate variability (*Hasselmann*, 1999)." We prefer to make this general statement, and not to venture into the more contentious issue of whether anthropogenic forcing has altered the frequency and/or intensity of ENSO events.

MacCracken CH5-14, Page 26, Footnote 19: It is not clear to me why these early studies with defective sets of corrections are any longer being quoted. It may be fine to indicate that different studies give different results, but mention should at least be made that these studies were using what are now considered defective datasets.
Michael MacCracken, Climate Institute

Response: At the end of Section 4.3, there is a summary paragraph that addresses the Reviewer's concern [Page 100, column 1, para.2]:

"It should be emphasized that all of the studies reported on to date in Section 4 relied on satellite data from one group only (UAH), on early versions of the radiosonde data[25], and on experiments performed with earlier model "vintages". It is likely, therefore, that this work may have underestimated the structural uncertainties in observed and simulated estimates of lapse rate changes."

Footnote 25 elaborates on problems with the radiosonde data:

"These radiosonde data sets were either unadjusted for inhomogeneities, or had not been subjected to the rigorous adjustment procedures used in more recent work (*Lanzante et al*., 2003; *Thorne et al*., 2005)."

We feel it important for an assessment report to provide some sense of the evolution of the field. This includes discussion of earlier research conducted with potentially flawed datasets. We are very open about the possible deficiencies in this earlier work. No changes made.

MacCracken CH5-15, Page 27, Line 524-534: Again, fine to say that regression techniques have been used, but why be citing the results of these studies when they are based on defective datasets—at the very least, the problems in them should be mentioned. I have elsewhere suggested that a table or appendix is needed that gives a timeline of the corrections that were made (mostly to the UAH datasets, but also the radiosonde ones) and the effects that not accounting for later corrections had on the results. So, if a trend in an earlier paper is quoted here, the value that would result using the corrected/improved datasets should be given in parentheses.
Michael MacCracken, Climate Institute

Response: See **Response to MacCracken CH5-14**. Note that Table 2.3 in Chapter 2 does now provide a timeline of adjustments made to the UAH data.

MacCracken CH5-16, Page 28, Line 536 and 539-540, *etc*.: It is not clear here which spatial dimension is being referred to—there are different mechanisms affecting the vertical and horizontal dimensions, so it seems to me essential to be indicating which ones are being referred to in each sentence, and in headings.
Michael MacCracken, Climate Institute

Response: We disagree. We explicitly note that regression can be "…performed "locally" at individual grid-points and/or atmospheric levels." [Page 99, column 1, para. 1]. The *Free and Angell* (2002) paper cited in this Section relied on radiosonde data in the form of zonal means at individual pressure levels. The study by *Hegerl and Wallace* (2002) used "…gridded fields of surface temperature data, UAH T_{2LT}, and "synthetic" T_{2LT} calculated from radiosonde data." [Page 99, column 1, para. 2] We do not believe that additional technical detail is necessary at this point in the text. Interested readers can refer to the peer-reviewed literature for further methodological details. No changes made.

MacCracken CH5-17, Page 28, Line 542-545: Although tied to this sentence, this comment is more general. As I understand what is being done in a number of these analyses, the lapse rate is being determined by taking the difference between a tropospheric temperature (at some level) and the surface temperature, and this seems to me a seriously flawed approach. Given that the surface and troposphere are essentially disconnected over much of the Earth (when looking, at least, at monthly anomalies—see Figure 1.4), how can one be confident that one is really determining a lapse rate change rather than a change in the intensity of an inversion (a possibility that needs to be mentioned here in the text)? While a lapse rate can be determined from a radiosonde, there is a real need to make sure this is being done with a data set that has been fully corrected, and this has taken until very recently, and may not even be good enough yet. I just do not see how a satellite-derived temperature can be used along with a surface temperature to determine that the lapse rate has changed, and use of satellite-derived temperatures at different levels seems to me also fraught with problems (e.g., stratospheric contamination, etc.). Now, one might be able to say that the surface and atmosphere are warming at different rates, but calling this a lapse rate change does not seem justified to me given that the temperature profile could well include an inversion. Also, in comparing models and observations, it seems to me that the limitations in model representations of the near surface PBL might well lead to differences in estimates of changes across this interface, and this might well have nothing to do with the suggestions of model physics aloft having shortcomings, *etc*.
Michael MacCracken, Climate Institute

Response: The Reviewer's criticism would be valid if such "temperature differencing" were being performed for very small regions and very short timescales. However, when averages are taken over very large spatial scales (such as the entire tropics) and long periods of time (months to decades), the changes in lapse-rate determined by simple differencing of temperatures at the surface and aloft are very similar to temperatures inferred from "true" lapse-rate calculations. For example, as described in Section 4.3 [Page 99], the study by *Gaffen et al*. (2000) explicitly calculated tropical lower tropospheric lapse rates from radiosonde data, whereas *Brown et al*. (2000) differenced observed T_S and synthetic T_{2LT} data calculated from radiosondes in order to obtain approximate changes in tropical lapse rates. The two studies yielded very similar decadal changes in tropical lower tropospheric lapse rates.

As is noted in both Chapter 5 [Page 113, footnote 61] and Chapter 1, planetary waves and synoptic scale disturbances rapidly smooth out tropospheric temperature anomalies (*e.g.*, between convecting and non-convective regions). We deliberately restrict our analysis to very large spatial scales and to longer time scales in order to minimize the problems referred to by the Reviewer.

MacCracken CH5-18, Page 28, Line 552: Change "it had" to "their analyses indicated that it had"—it is their analyses that find this; we do not yet know this is absolutely true.
Michael MacCracken, Climate Institute

Response: Done. [Page 99, column 1, para. 2]

MacCracken CH5-19, Page 28, Line 553-554: Again, it is simply not clear to me that what is being done here gives an indication of a lapse rate change (which might more generally result from a change in the intensity of the circulation, *etc*.). What it seems to me is being said is that the surface and tropospheric temperature difference is changing, and somehow calling this a lapse rate seems to me to stretch the definition of lapse rate given that changing the strength of an inversion might be involved (certainly, Figure 1.4 shows vast areas where there is a disconnect). This is a very important issue as quite different processes are involved, and expectations from models would be quite different.
Michael MacCracken, Climate Institute

Response: See **Response to MacCracken CH5-17**.

MacCracken CH5-20, Page 29, Line 556-557: That models fail to reproduce this result ["replicate" is too strong an expectation, it seems to me] might well be due to inadequacies in their representation of the surface boundary layer and inversions that are present.
Michael MacCracken, Climate Institute

Response: Changed "replicate" to "adequately reproduce" [Page 99, column 1, para. 2]. See also **Response to MacCracken CH5-17**.

MacCracken CH5-21, Page 29, Line 561-568: Again, it is not clear that the term "lapse rate" is being appropriately used, except perhaps for some of the radiosonde analyses. Instead, it may be that these studies are addressing the issue of surface-troposphere coupling, the changing strength of the inversion, *etc*.
Michael MacCracken, Climate Institute

Response: See **Response to MacCracken CH5-17**.

MacCracken CH5-22, Page 29, Line 571: It should be indicated what levels were analyzed so it is clear how the lapse rate was determined.
Michael MacCracken, Climate Institute

Response: These details can be found in the literature being referred to. Each study has used slightly different data sets, methods of calculating actual or approximate lapse-rate, atmospheric levels, *etc*. In our judgment, outlining all of these technical issues for each study referred to is well beyond the scope of this Chapter. The interested reader can find this technical information in the cited papers.

MacCracken CH5-23, Page 30, Line 576: Again, what levels were analyzed to get at the lapse rate?
Michael MacCracken, Climate Institute

Response: See **Response to MacCracken CH5-22**.

MacCracken CH5-24, Page 30, Line 581: Again, it may be weaknesses in the model representations of the PBL that are causing the problem—not something more serious.
Michael MacCracken, Climate Institute

Response: See **Response to MacCracken CH5-17**.

MacCracken CH5-25, Page 32, Line 637-646: Somewhere here, it would be helpful to say that the observational record is not long enough to get at the correlations structure of natural variability using observations alone.
Michael MacCracken, Climate Institute

Response: In Box 5.5 [Page 101, para. 3], it is explicitly stated that: "A number of choices must be made in applying D&A methods to real-world problems. One of the most important decisions relates to "reduction of dimensionality". D&A methods require some knowledge of the correlation structure of natural climate variability. This structure is difficult to estimate reliably, even from long model control runs, because the number of time samples available to estimate correlation behavior is typically much smaller than the number of spatial points in the field."

The fact that observational data are generally of insufficient length to reliably estimate this correlation structure is implicit in the text quoted above. No changes made.

MacCracken CH5-26, Page 33, Line 658-659: It might be noted that some fingerprints can be sought from the observations. For example, superposed epoch analysis (or a similarly named technique) is used to get at the fingerprint of volcanic eruptions from observations alone.
Michael MacCracken, Climate Institute

Response: Superposed epoch analysis has not been used in any of the D&A studies reported on here. We do not think it is necessary to mention this.

MacCracken CH5-27, Page 33, Footnote 39: It should be mentioned that the failure of models to generate the QBO has been found (by Mahlman) to be due to insufficient vertical resolution in the area of the tropopause, so that this is really a shortcoming that results from inadequate computer resources to do the full problem (also the case, quite likely for simulating surface inversions), and is not some fundamental physical flaw with the models or their sets of equations.
Michael MacCracken, Climate Institute

Response: The issue of why models fail to produce QBO variability is not of central interest to this Report. Factors other than vertical resolution may also play a role (*e.g.*, treatment of upper boundary condition and gravity wave drag). We do not believe it is appropriate to discuss this issue in more detail.

MacCracken CH5-28, Page 34, Line 681-682: This is a much more informative way of presenting present understanding than was expressed on lines 65-66.

Michael MacCracken, Climate Institute

Response: The IPCC TAR statement that "There is new and stronger evidence that most of the warming observed over the past 50 years is due to human activities" relates to D&A results obtained with near-surface air temperature changes. In contrast, our Key Finding 2 ("Results from many different fingerprint studies provide consistent evidence of a human influence on the three-dimensional structure of atmospheric temperature over the second half of the 20th century") is based on D&A studies that consider both surface and upper-air temperature changes. Our conclusion is a bit more cautious than the IPCC statement, since (as shown in this Report) current observational uncertainties are larger for upper-air data than for surface data. No changes made.

MacCracken CH5-29, Page 35, Line 696: Why is the word "claimed" used here, implying some doubt about this result, and not used in describing quite a number of the earlier results of Christy and Spencer, for example, where the findings (e.g., of the early data sets being highly accurate and not having biases still needing to be corrected, etc.) have since been found not to be the case?
Michael MacCracken, Climate Institute

Response: Done. Changed "have claimed" to "have reported". [Page 102, column 1, para. 2]

MacCracken CH5-30, Page 40, Line 797: I am a bit surprised that the new results are not featured in this chapter rather than deferring presentation of them to section 5 (and page 40—and chapter 5). These are the results that the readers will want to know and that are most useful to them—so why are they hidden way back here? This needs to be changed.
Michael MacCracken, Climate Institute

Response: We do not believe that the new results are being "hidden". Readers are being presented with information in a logical way:

⇒ An introduction to the physical climate system;
⇒ An introduction to temperature measurement systems;
⇒ Detailed discussion of the observed changes;
⇒ Discussion of possible explanations for differences between observed changes in different datasets;
⇒ Discussion of the historical evolution of studies seeking to understand and explain "differential warming" of the surface and troposphere;
⇒ Discussion of the latest research on differential warming, involving new model and observational datasets.

Chapter 5 draws heavily on information provided in previous Chapters regarding structural uncertainties in the observations. We therefore feel that the current ordering is optimal for the purposes of assessing the science.

MacCracken CH5-31, Page 44, Line 861: I would suggest changing "have" to "may have". What has become clear in the new findings is that it is the observations that have had the problems, not the models, so making this statement without qualification is really misleading. Also, as noted earlier, this notion of lapse rate problems may instead be a problem with model resolution of the inversions created in the PBL (that is, in the intensity of the disconnects shown in Figure 1.4). Thus, it seems to me the sentence on lines 859-861 needs to be revised to not be so one-sided about models and so limited to the notion of lapse rate (which may itself be a misleading naming of the problem).
Michael MacCracken, Climate Institute

Response: Sentence has been changed to: "Our primary focus is on the tropics, since previous work by *Gaffen et al*. (2000) and *Hegerl and Wallace* (2002) suggests that this is where any differences between observations and models are most critical." [Page 105, column 2, para. 2]

MacCracken CH5-32, Page 45, Footnote 55: Stopping the analysis in 1999 seems really unfortunate. Not being able to present results up to the present has previously been the subject of misleading complaints, and it would be a shame for that to happen again. Effort should be put into carrying forward the analysis through 2005, which would also get one away from the potential bias of being near to a major El Nino event. [Comment also applies to Figure 5.3.]
Michael MacCracken, Climate Institute

Response: Unfortunately, this is not possible. Only a small number of the models analyzed here had 20CEN runs that extended beyond 1999. [See Page 107, footnote 46]

MacCracken CH5-33, Page 56, Line 1043-1044: Is it really so difficult to reach a more definitive conclusion now that one has the Sherwood fixes to the radiosondes and the Wentz-Mears improvements to the satellite data record? With the most up to date records (most carefully corrected records), is it not possible to indicate that there is no real inconsistency between models and observations, except perhaps due to treatments of the surface inversion and QBO, which would be possible with more highly, resolved models?
Michael MacCracken, Climate Institute

Response: See Key Finding 6, bullet 5 [Page 90]. We do in fact make an explicit statement that:

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational data sets, leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report (model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and the independent physical evidence supporting substantial tropospheric warming) favors the second explanation".

We believe that this Key Finding, together with the discussion in the final two paragraphs of Section 5.4 [Page 115], is a reasonable summary and assessment of the current state of the science.

McDonald CH5-1, Page 5, Line 113: In a rhetorical question to Dr Watson, Sherlock Holmes asked: "How often have I said to you that when you have eliminated the impossible, whatever remains, HOWEVER IMPROBABLE, must be the truth?", Sir Arthur Conan Doyle, The Sign of Four (1890) ch. 6.

On line 113 it is stated that there are several possible explanations why the climate models and the observations differ with regard to decadal temperature changes. However, only two explanations are given which can be paraphrased as that the models are wrong and that the observations are wrong. Since this covers all the possibilities, I suggest 'several' is changed to 'two'.

Response: Done. Text has been modified. The modified text [Page 90, Key Finding 6, bullet 5] now reads as follows:

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational data sets, leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report (model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and the independent physical evidence supporting substantial tropospheric warming) favors the second explanation".

Hopefully this would lead to more attention being paid to the first explanation, which seems to have been ignored both in the past by investigators and now by writers of this report, on the grounds that the second explanation is "more likely." In fact the idea that all the radiosonde and MSU measurements are wrong, despite the intense activity over the last ten years to prove them so, seems to me to be impossible, so the idea that the computer models are wrong, although improbable, must be true!

Response: Sherlock Holmes also commented that "It is a capital mistake to theorize without data." We now have the hard data that Sherlock Holmes would have wanted (had he been in charge of this assessment Report!) On the basis of the scientific evidence presented in Chapters 3 and 4, Holmes would have reached the inescapable conclusion that structural uncertainties in satellite- and radiosonde-based estimates of tropospheric temperature change are much larger than hitherto believed. The science clearly shows that the choices made by different data analysts (in adjusting raw data for known inhomogeneities) can have a significant impact on estimated large-scale temperature trends. The structural uncertainties in the observations encompass current model-based estimates of the tropospheric temperature changes. These are statements of fact, not value judgments on our part.

In the expert judgment of our group, "the observational error" explanation is a better fit to the available scientific evidence than the "model error" explanation. We believe that the "observational error" explanation is the most logical one based on the model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and the independent physical evidence supporting substantial tropospheric warming (tropospheric water vapor increases, accelerated retreat of high-altitude tropical glaciers, *etc*.)

Bottom line**:** Our Key Finding 6 (bullet 5) is consistent with the available scientific evidence. We are very careful not to state categorically that our finding is "truth" – as new evidence becomes available, our conclusions will be reassessed. This caution is reflected in the bullet 6 of Key Finding 6.

In fact, as is postulated in the draft, line 114, diurnal forcing and decadal forcing ARE driven by "different physical mechanisms." The diurnal forcing is due to changes in solar radiation, and the decadal trend is driven by the increase in greenhouse gases [IPCC TAR]. This points to the model error lying in the treatment of outgoing longwave radiation.

I have, in fact, identified where the models are going wrong. They are using Planck's function to calculate the effect of greenhouse gases in the atmosphere. Planck's function is correct for continuous radiation such as that emitted by a blackbody e.g. the surface of the Earth. It is not valid for line radiation such as that absorbed and emitted by greenhouse gas molecules. In the "real world" line radiation is broadened according to the Voigt profile. It is not amplified by Planck's function as the models assume.

The formula used for calculating the effect of greenhouse radiation is known as Schwarzschild's equation. It was developed by him to model the radiation in the Sun.

It is valid for the photosphere which does radiate as a blackbody, but it is not appropriate for the Sun's chromosphere, nor for the Earth's atmosphere both of which are composed of low pressure gases that do not act as blackbody radiators. Just as the chromosphere creates lines in the blackbody radiation from the Sun, so the atmosphere creates lines, which merge into bands, in the blackbody radiation from the Earth's surface. This erroneous use of Schwarzschild's equation was first applied to the Earth's atmosphere by Robert Emden in 1913, long before the true quantum mechanical explanation of line emission was known. Later, Chandrasekhar, extolled the Schuster-Schwarzschild method, but he was an astrophysicist, and he correctly applied it to radiation within stars.

The reason that the tropical diurnal cycle does fit with the Schuster- Schwarzschild model is because the tropical climate is dominated by the evaporation and condensation of water vapour. The water aerosols (clouds) which forms when the vapour condenses do emit blackbody radiation based on their temperature because they have a surface being liquid. Thus Schwarzschild's equation does provide a reasonable approximation in case where there is a column of cloud.

There are two reasons I have not published these ideas. The first is that they have not been fully developed. It is easy to see why the Schuster-Schwarzschild method is wrong. It is not quite so easy to build a new model which is correct.

The second reason is that I feel there is little chance of such an "improbable" idea being published. My first attempt, which can be seen here;
http://www.abmcdonald.freeserve.co.uk/brief/brief.pdf
did not get past the editor, far less receive rejection from a disbelieving peer reviewer. Hence my attempt now to bypass the middle man and speak directly to the scientists concerned. (Alastair B McDonald, The Open University)

Response: This is a synthesis and assessment report. It is not the correct forum for consideration of new hypotheses not accepted in the literature. We suggest that Mr. McDonald continues to pursue recognition and peer-review of his hypothesis through the traditional channels.

Robock CH5-1, p. 13, footnote 2. Add at the end: "However, Mao and Robock (1998) used this fact to isolate the volcanic effect on surface air temperature." Mao, Jianping and Alan Robock, 1998: Surface air temperature simulations by AMIP general circulation models: Volcanic and ENSO signals and systematic errors. *J. Climate*, **11**, 1538-1552.
Alan Robock, Rutgers University

Response: This reference has now been added, but at an earlier point in the text. [Page 93, column 2, para. 2]

Robock CH5-2, p. 24, Fig. 5.1: This was done previously by Vinnikov *et al*. (1996), and their work should be acknowledged and referenced. Vinnikov, Konstantin Ya., Alan Robock, Ronald J. Stouffer, and Syukuro Manabe, 1996: Vertical patterns of free and forced climate variations. *Geophys. Res. Lett*., **23**, 1801-1804.
Alan Robock, Rutgers University

Response: The point of Figure 5.1 is to show that different external forcings have different characteristic signatures in vertical profiles of atmospheric temperature change. The Figure contrasts results from "single forcing" runs performed with individual changes in well-mixed GHGs, sulfate aerosol direct effects, tropospheric and stratospheric ozone, solar irradiance, and volcanic aerosols. It also shows the temperature response to combined changes in all five of these forcings. While *Vinnikov et al*. (1996) did present a similar vertical profile (their Figure 2), the only external forcing that they considered was a change in atmospheric CO_2. It is not appropriate, therefore, to reference the *Vinnikov et al*. (1996) paper at a point in the text where the fingerprints of different forcings are being described.

Robock CH5-3, p. 25, footnote 17: The formatting of this footnote is all messed up. The first paragraph stops erroneously at "...atmosp" The following should be inserted there:

heric CO2 levels. This is often referred to as ΔT2•CO2. Estimates of ΔT2•CO2 have been obtained by studying Earth's temperature response to "fast", "intermediate", and "slow" forcing of the climate system. Examples include the "fast" (<10-year) response of surface and

Everything else in the footnote after the first paragraph, starting with "17It is useful to mention one technical issue..." should be deleted.
Alan Robock, Rutgers University

Response: This relates to a bug in Microsoft word and has been rectified. [Page 94]

Robock CH5-4, p. 48, Fig. 5.3: Change "VG" to "UMd"
Alan Robock, Rutgers University

Response: Done. [Page 110]

Robock CH5-5, p. 49, line 964: This line is part of the text and should be moved after the table and should be the same font size as line 966.
Alan Robock, Rutgers University

Response: Done. [Page 109, column 1, para. 2]

Robock CH5-6, p. 49, Table 5.4A: All estimates should be rounded to two decimal places. Use "<0.01" if necessary.
Alan Robock, Rutgers University

Response: Done. [Page 109]

Robock CH5-7, p. 55, Table 5.4B: All estimates should be rounded to two decimal places. Use "<0.01" if necessary.
Alan Robock, Rutgers University

Response: Done. [Page 112]

Robock CH5-8, p. 55, Footnote 68. Use "UMd" instead of "VG"
Alan Robock, Rutgers University

Response: Done. [Page 111, footnote 59]

Robock CH5-9, p. 80, lines 1764-1766: This paper has now been published. The reference is: Robock, Alan, 2005: Comment on "Climate forcing by the volcanic eruption of Mount Pinatubo" by David H. Douglass and Robert S. Knox. *Geophys. Res. Lett.*, **32**, L20711, doi:10.1029/2005GL023287.
Alan Robock, Rutgers University

Response: Done.

Singer CH5-1, P3 line 52-54. I strongly dispute this claim. While there must clearly be SOME effect on climate from the increased level of anthropogenic forcing, it is not evident from the climate records presented here. Clearly, the human component is still quite small in comparison to natural climate fluctuations. [*Singer*]

Response: The Reviewer is referring to Key Finding 2. This finding states that "Results from many different fingerprint studies provide consistent evidence of a human influence on the three-dimensional structure of atmospheric temperature over the second half of the 20^{th} century" [Page 89].

The scientific underpinning for Key Finding 2 is provided in Section 4.4 of our Chapter. This Section evaluates evidence from literally dozens of pattern-based "fingerprint" studies, which have used rigorous statistical methods to compare modeled and observed surface and atmospheric temperature changes. This work has been conducted by research groups around the world (*e.g.*, at Oxford University, The Hadley Centre for Climate Prediction and Research, Lawrence Livermore National Laboratory, Scripps Institution of Oceanography, Texas A&M University, Duke University, the Max-Planck Institute for Meteorology, the National Center for Atmospheric Research, the Geophysical Fluid Dynamics Laboratory, and the Canadian Climate Center). These groups have used different statistical methods, and different sets of model and observational data. The common denominator in all of this research is that:

⇒ Human-caused greenhouse-gas and sulfate aerosol signals are identifiable in observed surface temperature records.

⇒ A human-induced ozone depletion signal is identifiable in stratospheric temperature records.

⇒ The combined effects of greenhouse gases, sulfate aerosols, and ozone depletion are identifiable in the vertical structure of atmospheric temperature changes (from the surface to the stratosphere).

⇒ Natural factors have influenced surface and atmospheric temperatures, but cannot fully explain their changes over the past 50 years.

The Reviewer may find these conclusions unpalatable, but they are extensively documented in the peer-reviewed literature. Extraordinary claims demand extraordinary proof. Claims of a substantial human effect on global climate <u>have</u> been subjected to tremendous scrutiny, and the "extraordinary proof" of these claims has been presented not only in this assessment, but also in previously-published assessments by the Intergovernmental Panel on Climate Change and the U.S. National Academy of Sciences.

In contrast, the Reviewer engages in "science by assertion". He asserts that "…the human component is still quite small in comparison to natural climate fluctuations", but provides absolutely no scientific evidence to support this assertion.

Our assessment relies on the analysis of the peer-reviewed literature, not on unsupported assertions. No changes made.

Singer CH5-2, P3 line 58: The sulfate aerosol signal is NOT seen in the observed record. In fact, it is contradicted by the observed NH/SH temp differences. See Fig 5.7 on p. 64, line 1185. [*Singer*]

Response: A signal of sulfate aerosol effects on surface and atmospheric temperatures has been statistically identified in numerous fingerprint studies. These studies are discussed at length in Section 4.4 of our Chapter.

As is pointed out on Page 103 [column 2, paragraphs 2 and 3] and Page 104 [column 1, paragraph 1], it is necessary to use so-called "space-time" fingerprint methods for identifying sulfate aerosol effects on climate. Such methods explicitly account for important changes over time in the spatial pattern of both the sulfate aerosol signal and the observed temperatures. As Section 4.4 explains, because the forcing from both greenhouse gases and sulfate aerosols has changed over the 20th century, and because each of these factors is expected to have different temperature effects in the Northern and Southern Hemispheres (NH and SH), we expect that that NH/SH temperature differences should change with time! As the recent *Stott et al.* (2006)[7] paper shows [see Page 104], model simulations with combined changes in greenhouse gases and sulfate aerosols are capable of capturing observed changes in NH/SH temperature differences.

Bottom line: The Reviewer's unsupported assertion is incorrect. Figure 5.7 does not contradict claims of an identifiable sulfate aerosol effect on climate. No changes made.

Singer CH5-3, P3 line 60: The observed stratospheric temp decrease is difficult to explain by ozone depletion. There has been no ozone depletion in the tropics at all, and an increase in ozone levels in NH mid-latitudes since 1992. These are not reflected in the strat. temp obs. See Fig 5.7 on p. 64, line 1185 [*Singer*]

Response: Again, the Reviewer simply makes unsupported assertions. A number of peer-reviewed studies have rigorously compared simulated and observed stratospheric temperature changes (*e.g.*, *Ramaswamy et al*. 1996[8]; *Santer et al*., 2003[9]; *Ramaswamy et al*., 2006[10]). These studies – which are cited in Chapter 5 – find hard scientific evidence for a pronounced effect of stratospheric ozone depletion on stratospheric temperatures.

---

[7]Stott, P.A., *et al*., 2006: Robustness of estimates of greenhouse attribution and observationally constrained predictions of global warming. *Journal of Climate* (in press).

[8]Ramaswamy, V., M.D. Schwarzkopf and W.J. Randel, 1996: Fingerprint of ozone depletion in the spatial and temporal pattern of recent lower-stratospheric cooling. *Nature*, **382**, 616-618.

[9]Santer, B.D., *et al*., 2003: Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science*, **301**, 479-483.

[10]Ramaswamy, V., *et al*., 2006: Anthropogenic and natural influences in the evolution of lower stratospheric cooling. *Science*, **311**, 1138-1141.

Changes in solar and volcanic forcing alone cannot explain the observed lower stratospheric temperature changes over the satellite era (*Ramaswamy et al*., 2006).

The Reviewer correctly notes that "There has been no ozone depletion in the tropics at all, and an increase in ozone levels in NH mid-latitudes since 1992". But this comment is disingenuous. It fails to note that since 1980, there has been a substantial decrease in total column ozone poleward of 30°S (see *Chipperfield et al*., 2003[11], their Figure 4-7; *Fahey*, 2003[12], their Figure Q13-1). Furthermore, even in the NH mid-latitude region highlighted by the Reviewer, there has been an overall decrease in total column ozone since 1980 (see *Fahey*, 2003, their Figure Q13-1). The increase since 1992 in NH mid-latitudes arises in part because of "recovery" from the Pinatubo-induced depletion of stratospheric ozone. In the tropics, the SAGE I/II data do show a significant decrease in ozone above 35 km (*Chipperfield et al*., 2003, their Figure 4-9).

Bottom line: Observed stratospheric temperature changes are difficult to explain without stratospheric ozone depletion. The Reviewer's comments regarding stratospheric ozone loss in the tropics and NH mid-latitudes are highly selective and disingenuous. No changes made.

Singer CH5-4, P4 line 82-83: I agree that temp and temp trend comparisons in the Tropics would provide the clearest test of GH theory, unaffected by sea ice and snow feedbacks, etc [*Singer*]

Response: No response required.

Singer CH5-5, P4 line 91-93: This discrepancy between obs and models is crucial to the conclusion that the GH effect is still quite small compared to natural climate variations. Evidently, the models overestimate the importance of GH warming. [*Singer*]

Response: The Reviewer is referring to the following sentence: "In the tropics, most observational datasets show more warming at the surface than in the troposphere, while most model runs have larger warming aloft than at the surface" [Key Finding 5, bullet 3, Page 90]. He ignores all of the evidence – presented in this chapter and throughout the Report – of significant uncertainty in observationally-based estimates of tropospheric temperatures trends. This uncertainty is particularly serious in the tropics. New satellite- and radiosonde-based estimates of tropical T_{2LT} trends suggest that that there is no fundamental discrepancy between modeled and observed trends in lower tropospheric lapse rates (see **Response to Douglass CH5-1**).

---

[11]Chipperfield, M.P., *et al*., 2003: Global ozone: Past and future. Chapter 4 in *Scientific Assessment of Ozone Depletion: 2002*. Global Ozone Research and Monitoring Project. Report No. 47, World Meteorological Organization, Geneva, 498 pp.

[12]Fahey, D.W., 2003: *Twenty Questions and Answers About the Ozone Layer: Scientific Assessment of Ozone Depletion: 2002*. World Meteorological Organization, Geneva, 42 pp.

The Reviewer's interpretation of the apparent discrepancy between modeled and observed tropical lapse-rate trends is that "…the models overestimate the importance of GH warming". We admit the possibility of model error ("These results could arise due to errors common to all models") in Key Finding 6, bullet 5 [Page 90]. However, in the expert judgment of most of the authors of this Report, the more likely interpretation of the "discrepancy" mentioned by the Reviewer is the existence of "significant non-climatic influences remaining within some or all of the observational data sets, leading to biased long-term trend estimates" [Page 90]. This interpretation was favored because of the "model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and independent physical evidence supporting substantial tropospheric warming" [Page 90].

Clearly, not all observational upper-air datasets can be correct. The currently-large range of observational uncertainty encompasses model-based estimates of recent trends in tropical lower-tropospheric lapse rates. The Reviewer may not like this conclusion, but it is inarguable (see **Responses to Douglass CH5-1, CH5-7**). The fact that important cooling biases have been identified – as recently as last year – in commonly-used satellite and radiosonde climate data records should give the Reviewer pause for thought. From our perspective, we must reduce current uncertainties in observed upper-air temperature records before we can reach definitive conclusions regarding the reality (let alone the causes) of putative "discrepancies" between modeled and observed lapse-rate changes.

Bottom line: The existing text is suitably cautious and circumspect on the point raised by the Reviewer. The Reviewer presents a conclusion that is not cautious and circumspect ("…the models overestimate the importance of GH warming"), and which he does not attempt to justify. No changes necessary or made.

Singer CH5-6, P5, line 98-100: This result on amplification on monthly and inter-annual time scales confirms my conclusion that a moist convective atmosphere is in accord with theory; however, the absence of such amplification on a decadal time scale shows that the models overestimate GH warming [*Singer*]

Response: See **Response to Singer CH5-5.** As noted above, amplification is not "absent on a decadal scale". It is actually present on a decadal scale in some observational datasets (see **Responses to Douglass CH5-1, CH5-7**). The Reviewer's preferred conclusion ("…models overestimate GH warming") implies that in the real world, different physical mechanisms must control amplification behavior on short (month-to-month and year-to-year) and on long (decade-to-decade) timescales.[13] What are these different physical mechanisms? Unfortunately, the Reviewer's comments do not enlighten us on this key point.

Bottom line: We are suitably cautious in our conclusions regarding simulated and observed amplification behavior [see Key Finding 6, bullets 5 and 6, Page 90]. We mention both possible explanations (model error and observational error) for the

[13] At least according to those observational datasets which show tropospheric damping of tropical surface temperature changes.

amplification results presented in Chapter 5. We note that these explanations are not mutually exclusive. The Reviewer favors the "model error" interpretation of our results, without providing any scientific justification for his preference. No changes made or necessary.

Singer CH5-7, P5 line 113-116: The simplest explanation is one not mentioned. Namely: amplification on monthly and inter-annual time scales confirms merely that a moist convective atmosphere is in accord with theory; however, the absence of such amplification on a decadal time scale shows that the models overestimate GH warming [*Singer*]

Response: See **Response to Singer CH5-6**.

Singer CH5-8, P6 line 117: This alternative explanation, which simply blames any disagreement between data and model results on errors and uncertainties, is unsatisfactory. It appears to be more ideological than scientific. [*Singer*]

Response: Disagree strongly. See **Response to Singer CH5-6**. It is undeniable that there are large uncertainties in observed estimates of tropospheric temperature changes over the past 2-3 decades. These uncertainties make it difficult to reach definitive conclusions regarding the reality of a significant discrepancy between modeled and observed tropical lapse-rate changes. These issues are discussed in a fair and balanced way in Chapter 5 [see, *e.g*., Key Finding 6 on Page 90, and the final two paragraphs of Section 5.4 on Page 115].

In contrast, the Reviewer favors a "model error" interpretation of the Chapter 5 results, but provides absolutely no scientific justification for this interpretation. Once again, the Reviewer is engaging in "science by assertion". The charge of ideological bias is unjustified and offensive. We provide a detailed scientific rationale for our expert judgments. The Reviewer does not. Perhaps he should consider whether his own criticism is ideologically motivated. No changes necessary or made.

Singer CH5-9, P54 Line 1027: The crucial evidence for disagreement between data and models comes from Fig 5.4G. It cannot be just explained away by errors and uncertainties [*Singer*]

Response: See **Response to Singer CH5-6, CH5-8**.

Singer CH5-10, P55 Line 1036-1038: As stated, the radiosonde data and UAH satellite result lie outside the range of the results from 49 model runs. The RSS satellite result is barely consistent with the model results used here. [No explanation is given as to why UAH and RSS disagree.] In any case, it is more than likely that if more than 49 model runs had been used, the dispersion would have been reduced, and the RSS result would then also be inconsistent with models. [*Singer*]

Response: On the issue of "overlap" between modeled and observed tropical lapse-rate trends, please see **Responses to Douglass CH5-1 and Douglass CH5-7**. Possible explanations as to why RSS and UAH T_{2LT} results disagree are discussed in Chapter 4. The issue of a whether a larger number of model runs would have led to a reduction in the "dispersion" of the model results is discussed in the **Response to Douglass CH5-7**. No changes made or necessary.

Singer CH5-11, P64 line 1185: Fig. 5.7 clearly shows the disagreement between modeled and observed (Fig. 5.7E) temp trends vs altitude. The radiosonde data show even a slight mid-troposphere cooling trend in the equatorial zone. These results confirm the findings of Douglass, Pearson, Singer GRL 2004. Note also that the strat cooling trend is rather uniform as a function of latitude, in disagreement with measured ozone depletion. [*Singer*]

Response: As discussed in Chapter 5 and elsewhere in the report (particularly in Chapter 4), there are significant uncertainties in current radiosonde-based estimates of atmospheric temperature change. Recent reanalyses of radiosonde records by *Sherwood et al*. (2005) and *Randel and Wu* (2006) suggest that the observed trends shown in Figure 5.7E may contain important residual cooling biases, particularly in the lower stratosphere and the tropical troposphere. Both *Sherwood et al*. (2005) and *Randel and Wu* (2006) show that such biases can translate to large uncertainties in the observed vertical profile of recent atmospheric temperature change. The paper referred to by the Reviewer (*Douglass et al*., 2004) did not consider such observational uncertainties, and is of limited usefulness here.

<u>Bottom line:</u> Although the Reviewer may not like existing observational uncertainties, they are undeniably real and important for our ability to evaluate climate models. No changes made or necessary.

Trenberth CH5-1, This chapter is pretty good but I only skimmed it.
Kevin Trenberth, National Center for Atmospheric Research

Response: Thanks! No change required.

Trenberth CH5-2, Page 14, Line 290: this footnote 5 assumes that ENSO is well simulated in models, but it isn't in any, even though it has improved.
Kevin Trenberth, National Center for Atmospheric Research

Response: We have updated Box 5.1 [Page 92] in response to this and several other comments. We cite a paper that documents (at least in certain models) demonstrable improvement in simulation of certain aspects of ENSO behavior.

Trenberth CH5-3, Page 14, Line 292: footnote 6: not "may remain" but certainly do remain.

Kevin Trenberth, National Center for Atmospheric Research

Response: Changed "may remain" to "will remain". [Page 94, footnote 6]

Trenberth CH5-4, Page 48, Line 935, Figure 5.3: should have error bars on the observations.
Kevin Trenberth, National Center for Atmospheric Research

Response: We disagree. If we had only a single realization of a tropical T_{2LT} trend from a single CGCM, it would indeed be necessary to provide appropriate statistical error bars for the model trend and the observational trend. In our case, however, we have a large, multi-model, multi-realization ensemble of tropical T_{2LT} trends. Each of these realizations has a different manifestation of ENSO variability superimposed on the underlying model response to the imposed forcing changes. It is meaningful to ask – even without explicit consideration of statistical error bars – whether the observational T_{2LT} trend is contained within multi-model, multi-realization "envelope" of T_{2LT} trends. This is what we do in Chapter 5, as is explained in the paragraph immediately before Section 5.1 [Page 106 and 107], in the first paragraph of Section 5.2 [Page 111], and in the new footnote 45 [Page 107]. There is also further discussion of this issue in Section 8 of the Statistical Appendix.

Trenberth CH5-5, Page 54, Line 1027, Figure 5.4: should have error bars on the observations.
Kevin Trenberth, National Center for Atmospheric Research

Response: See **Response to Trenberth CH5-5**.

Winstanley ES-1, Page 2, Lines 25-26; and **Winstanley CH5-1:** In the Executive Summary, the focus of the report is broadened from that stated in the Preface (to understand the causes of differences between independently produced data sets) to also include understanding of the causes of the temperature changes themselves, which are addressed in Chapter 5. Whereas much attention is given in the report to addressing the strengths and weaknesses of different observed temperature trends, little attention is paid to documenting the strengths and weaknesses of the models whose outputs are compared with observations. The models also are used to understand causes of the differences among the observed trends and to understand the causes of the trends. Since there is considerable reliance on models in comparing observations with theoretical expectations and in evaluating the causes of observed changes, similar critique of the strengths and weaknesses of models should be included in the report as is given to the critique of the strengths and weaknesses of observations.

Response: Model evaluation is not the subject of the present Report. An in-depth critique of "the strengths and weakness of models" will be provided in CCSP Synthesis and Assessment Product 3.1: Climate Models: An Assessment of Strengths and Limitations for User Applications. CCSP Synthesis and Assessment Product 3.1 is now explicitly mentioned in Box 5.1 [Page 92].

Climate model experiments, and the forcings that are included in "20CEN" simulations, are discussed in some detail in Sections 2 and 3 of Chapter 5. Section 2 gives a fair and balanced discussion of the advantages and disadvantages of different experimental configurations (*e.g.*, "AMIP-style" runs versus CGCM experiments). Section 3 discusses uncertainties in natural and anthropogenic climate forcings, and in how these forcings are applied in 20CEN experiments. Boxes 5.1 and 5.2 give the reader useful background information on "Climate Models" and "Uncertainties in Simulated Temperature Changes".

Throughout Chapter 5, there is explicit mention of some of both the strengths and weakness of climate models. For example, Section 4.4 synthesizes information from many different "fingerprint" studies, and illustrates that some models have demonstrable skill in simulating important aspects of historical climate change. These are rigorous tests of model performance. The fact that a number of models pass these tests is undeniably a "strength" of climate models.

Nor are model "weaknesses" glossed over. Here are few examples of the discussion of model deficiencies:

⇒ "However, models also have systematic errors that can diminish their usefulness as a tool for interpretation of observations (Gates *et al*., 1999; McAvaney *et al*., 2001)." [Page 92, column 2, para. 1].

⇒ "Most models undergo some adjustment of poorly-known parameters which directly affect key physical processes, such as convection and rainfall… The aim of this procedure is to reduce the size of systematic model errors…" [Page 94, column 2, para. 3].

⇒ "This illustrates the need for caution in comparisons of modeled and observed atmospheric temperature change. The differences evident in such comparisons have multiple interpretations. They may be due to real errors in the models, errors in the forcings used to drive the models, the neglect of important forcings…" [Page 97, column 2, para. 4; page 97, column 1, para. 1].

⇒ "These (model errors) may lie in the physics, parameterizations, inadequate horizontal or vertical resolution, *etc*." [Page 97, footnote 10].

⇒ "For example, current CGCMs fail to simulate the stratospheric temperature variability associated with the QBO or with solar-induced changes in stratospheric ozone (Haigh, 1994)." [Page 100, footnote 28].

⇒ "Model errors in internal variability can bias detection results, although most detection work tries to guard against this possibility by performing "consistency checks" on modeled and observed variability…" [Page 100, column 2, para. 1].

⇒ "One possible interpretation of these results is that in the real world, different physical mechanisms govern amplification processes on short and on long timescales, and models have some common deficiency in simulating such behavior." [Page 115, column 1, para. 2].

Bottom line: Model evaluation will be covered in a separate Report. The Reviewer's claim that the current Report does not discuss model strengths and weaknesses is incorrect. The focus here is on those model strengths and weaknesses that are most relevant to the specific charge of this Report.

Winstanley ES-4 and **Winstanley CH5-2a:** Due to the fundamental climatological importance of lapse rates, the Executive Summary should contain a summary of what we know about lapse rates regionally and globally and how well regional and global climate models simulate actual temperatures and lapse rates. The draft Executive Summary says nothing about the fundamental subject of lapse rates. **Chapter 2, page 30, lines 541-543** state that explaining atmospheric and surface trends demands relative accuracies of a few hundredths of a degree per decade in global time series of both surface and upper-air observations and **Chapter 3, Section 7.2**, contains limited information on lapse rates. **Chapter 3, lines 986-988** acknowledges that "Most of the observational work to date has not examined lapse rates themselves, but instead has used an approximation in the form of a vertical temperature difference." In **Chapter 3**, with a summary in the Executive Summary, there needs to be discussion of the implications for climate studies of not reporting actual temperatures and lapse rates, and not comparing observed lapse rates with modeled lapse rates. Also, there should be discussion of the implications for the questions posed of using a surrogate lapse-rate approximation in climate studies. As a focus of the report is to compare observed and modeled vertical temperature variations, **Chapter 5** should include a statement about the accuracy of models in simulating decadal lapse rates, as well as changes in lapse rates.

Response: Most comparisons between modeled and observed lapse-rate changes have used what the Reviewer refers to as "a surrogate lapse-rate approximation" (*i.e.*, a difference between temperature trends at the surface and in some weighted average atmospheric layer, such as T_2 or T_{2LT}). Very few studies explicitly calculate a true lapse rate. There are some notable exceptions, such as the *Gaffen et al*. (2000) study discussed on pages 99 and 100. Lapse-rates are also explicitly calculated in some comparisons of modeled and observed changes in tropopause height (*Santer et al*., 2003a, 2004) [Page 118, footnote 75].

Bottom line: We can only assess the relevant studies that are available in the peer-reviewed literature, and most of these rely on a lapse-rate approximation rather than an explicit calculation. In our judgment, it is highly unlikely that this approximation will yield significantly different estimates of slow, large-scale lapse-rate changes (which are the primary focus of this Report) than explicit calculations of lapse-rate changes. This is supported by the similarity of the decadal-timescale lapse-rate changes in *Brown et al*. (2000) and *Gaffen et al*. (2000), which use (respectively) approximate and explicit lapse-rate calculations [Page 99, column 2, paragraphs 1 and 2].

The Reviewer requests information about "about the accuracy of models in simulating decadal lapse rates, as well as changes in lapse rates". We are not sure what this request means. Model performance in simulating changes in lapse rates is discussed extensively in Section 5 of Chapter 5 (see, *e.g.*, discussion of Figures 5.3F,G and 5.4F,G). A comprehensive assessment of model skill in simulating climatological mean lapse rates (which may or may not be what the Reviewer is trying to articulate in the phrase "simulating decadal lapse rates") has not yet been performed, and could not be assessed here.

Winstanley CH5-2b: The global climate system is a composite of regional climates and more discussion of regional lapse rates and changes in lapse rates would give readers more confidence that global analyses represent the composite of regional conditions accurately. That comprehensive regional-scale analyses of lapse rates have not been conducted is recognized in **Chapter 5, lines 862-866**. The Executive Summary should incorporate recognition of the importance of comprehensive regional analyses of lapse rates and state that they have not been conducted, if this is an accurate statement.

The report also should discuss the implications for the climate system (*e.g.*, stability and precipitation) of reported spatial and temporal variations in vertical temperature differences and lapse rates.

Derek Winstanley, Illinois State Water Survey

Response: See Response To Pielke Sr., GEN-3d,e. While we agree with the Reviewer that regional-scale evaluation of climate models is an important exercise, it was not an exercise central to this Report. The question at the core of our Report relates to a problem manifest at very large spatial scales. The large-scale nature of the discrepancy between observed surface and tropospheric temperature changes (and between modeled and observed tropospheric temperature changes) was what initially attracted the attention of scientists and policymakers.

As the Reviewer points out, Chapter 5 notes that:

"Our primary focus is on the tropics, since previous work by *Gaffen et al*. (2000) and *Hegerl and Wallace* (2002) suggests that this is where any differences between observations and models are most critical… We do not discount the importance of comparing modeled and observed lapse-rate changes at much smaller scales (particularly in view of the incorporation of regional-scale forcing changes in many of the runs analyzed here), but no comprehensive regional-scale comparisons were available for us to assess." [Page 105, column 2, para. 2].

<u>Bottom line:</u> We do not think it is necessary to expand on the discussion of this point in Chapter 5. While evaluation of model skill on regional scales is a useful exercise, we note that uncertainties in the observed tropical T_{2LT} trends over the satellite era are as large or larger than the expected signal arising from external forcing. In our judgment, the task of constraining the large uncertainties in observed upper-air datasets should be the highest-priority activity. These uncertainties "…make it difficult to determine whether models still have common, fundamental errors in their representation of the vertical structure of

atmospheric temperature change." [Page 90, Key Finding 6, bullet 6]. This holds for any evaluation of model skill, be it at regional, continental, or global scales.

Winstanley ES-5 and **Winstanley CH5-3a:** All major climate reports (e.g., IPCC, NRC, CCSP) adopt the approach of examining only temperature differences, either from one time period to another or between the surface and some height above the Earth's surface. This approach, adopted in reporting both observed temperature changes and modeled temperature changes, excludes explicit reporting of actual temperatures. A differential approach is appropriate in addressing many aspects of climate change, but also has limitations, which need to be addressed.

Particularly when discussing lapse rates or vertical temperature differences, actual temperatures and changes in actual temperatures are of great importance in evaluating the stability of the atmosphere and precipitation. By focusing only on temperature differences and avoiding actual temperatures conceals some important issues relating to model limitations, which are important in comparing differences between observed temperature changes and modeled temperature changes, and in evaluating the causes of temperature changes.

Response: As in the case of issue of 'regional evaluation of model skill', we can only assess what is actually available in the peer-reviewed literature. To our knowledge, comprehensive assessments of the type requested by the Reviewer are not available. The focus on anomalies rather than on actual temperatures arises because observational uncertainties are larger for the latter than for the former. This is why observational datasets considered in this report are generally expressed in anomaly form.

Winstanley CH5-3b: Kunkel *et al.* ("Can CGCMs simulate the Twentieth Century "Warming Hole" in the central United States?", in press, *Journal of Climate*, and attached with these comments*)* show major differences between the observed evolution of mean annual 20th Century temperature in Central North America (CAN) (*sic*) and mean annual temperature simulated by global climate models. There are significant differences between the observed and modeled temperature changes, and large differences between observed and modeled temperatures. The models simulate CNA mean annual temperature to an accuracy of only +/- 3^{o}C. This raises the question as to the credibility of models in simulating regional changes in temperature of a few tenths of a degree when the accuracy of the models in simulating mean annual temperature of the region spans a range of 6^{o}C.

Response: The Reviewer's comment implicitly assumes that there is a clear relationship between model biases in simulating the mean state and model errors in simulating time-evolving temperature changes. It is not obvious that such a relationship exists. Model skill in simulating the CNA's time-evolving surface temperature changes over the 20th century must also be related to the fidelity with which slow changes in external forcings are specified. Furthermore, meaningful skill assessments for such small regions are difficult owing to the large, chaotic variability of the climate system. Because of this variability, models cannot be expected to exactly reproduce observed regional patterns of temperature trends, even with hypothetical "perfect" models and complete knowledge of radiative forcing changes [see comments on Page 111, column 1, first complete paragraph, and footnote 56].

Detailed studies of regional hindcast skill were not available for all of the models discussed in Section 5 of Chapter 5, and so could not be provided. However, several of the models presented in Chapter 5 have been subjected to regional-scale assessments of model skill. Such work suggests that at least some current climate models do have skill in simulating observed, regional-scale surface temperature changes over the 20th century [see page 102, column 1, paragraphs 1 and 2]. One of these investigations (*Karoly et al*., 2003)[14] was for North America, and includes the CNA region analyzed by *Kunkel et al*. (2006).

Winstanley CH5-3c: This is consistent with the finding in the Third Assessment Report of the Intergovernmental Panel on Climate Change that "Nearly all regional temperature biases are within the range of +/- 4°C " (Giorgi and Hewitson, 2001, p.592 and figure 10.2(a)).

The draft Chapter 5 concludes that "When run with natural and human-caused forcings, model global-mean temperature trends for individual atmospheric layers are consistent with observations" (page 4, lines 79-80). The knowledge that there are large discrepancies between observed temperatures and modeled temperatures at the regional scale should be incorporated in Chapter 5 and the Executive Summary and the significance of these biases for global syntheses discussed.

Response: See **Response to Winstanley CH5-3b**.

Winstanley CH5-3d: Also, it must be asked what is the significance of these model limitations when evaluating lapse rates and changes in lapse rates? A bias in simulating surface temperature of +/- 3 °C must have major implications for understanding the stability of the atmosphere and precipitation regionally. When climate models simulate mean annual temperature across a range of 6°C or more, how well do they simulate lapse rates and changes in lapse rates? Is it only surface temperature values that are inaccurate, or do the inaccuracies extend into the atmosphere above? What are the implications of such inaccuracies when evaluating the causes of observed temperature changes of a fraction of a degree? How accurately do global climate models simulate actual temperatures in other regions of the world and globally? What does it mean to conclude that "there is no inconsistency between models and observations at the global scale" when studying vertical variations in temperature and temperature changes? The CCSP report needs to address these issues.

Derek Winstanley, Illinois State Water Survey

Response: See **Response to Winstanley CH5-2a,b; CH5-3a,b**.

[14]Karoly, D.J., *et al*., 2003: Detection of a human influence on North American climate. *Science*, **302**, 1200-1203.

Winstanley, ES-7 and **Winstanley, CH5-4a:** The discussion on models includes consideration of internal and external forcings as drivers of climate variations and change. There is no explicit recognition that natural internal variations of the climate system can bring about climate variations and change, and that internal variability needs to be considered as a factor when attributing causes of observed or modeled change.

Response: This is incorrect – there is “explicit recognition that natural internal variations of the climate system can bring about climate variations and change, and that internal variability needs to be considered as a factor when attributing causes of observed or modeled change.”

We provide below some examples of the discussion of natural internal variability in Chapter 5:

⇒ “In both observations and climate models, variations in the El Niño-Southern Oscillation (ENSO) have pronounced effects on surface and tropospheric temperatures.” [page 93, column 2, para. 2]

⇒ “Even with the specification of observed ocean boundary conditions, the time evolution of modes of variability that are forced by both the ocean and the atmosphere (such as the North Atlantic Oscillation; see Rodwell et al., 1999) will not be the same in the model and in the real world (except by chance).” [page 93, footnote 1]

⇒ “All of these realizations contain some underlying “signal” (the climate response to the imposed forcing changes) upon which are superimposed *n* different manifestations of “noise” (natural internal climate variability).” [Page 94, column 1, first complete paragraph]

⇒ “In a CGCM, ocean temperatures are fully predicted rather than prescribed. This means that even a (hypothetical) CGCM which perfectly captured all important aspects of ENSO physics would not have the same timing of El Niño and La Niña events as the real world (except by chance).” [page 94, column 1, second complete paragraph]

⇒ “In the real world and in “AMIP-style” experiments, this slow, volcanically induced cooling of the troposphere and surface is sometimes masked by the warming effects of El Niño events…” [Page 94, footnote 2]

⇒ “This illustrates the need for caution in comparisons of modeled and observed atmospheric temperature change. The differences evident in such comparisons have multiple interpretations… They may also be due to different manifestations of natural variability noise in the observations and a given CGCM realization.” [Page 96, column 2, para. 4; Page 97, column 1, para. 1]

⇒ Section 4.1 contains numerous examples of the use of regression-based methods for estimating the effects of ENSO variability on observed and simulated atmospheric temperature changes!

⇒ “While ENSO and COWL variability made significant contributions to the month-to-month and year-to-year variability of temperature differences between the surface and T_{2LT}…” [Page 99, column 1, para. 2]

⇒ "To evaluate whether natural climate variability could explain these slow variations…" [Page 99, column 2, para. 2]

⇒ "Fingerprints are also compared with patterns of climate change in model control runs. This helps to determine whether the correspondence between the fingerprint and observations is truly significant, or could arise through internal variability alone." [Page 100, column 2, para. 1]

⇒ "D&A methods have some limitations… They make at least two important assumptions: that model-based estimates of natural climate variability are a reliable representation of "real-world" variability…" [Page 101, Box 5.5, para. 4]

Bottom line: The Reviewer's claim is incorrect. Natural climate variability is discussed in detail throughout the text of Chapter 5. Note also that a paragraph relevant to this issue has been added to Box 5.1 [Page 92]

Winstanley, CH5-4b: Kunkel *et al.* ("Can CGCMs simulate the Twentieth Century "Warming Hole" in the central United States?" in press, *Journal of Climate,* and attached to these comments) demonstrate that "…the warming hole is not a robust response of contemporary CGCMs to the estimated external forcings. A more likely explanation based on these models is that the observed warming hole involves external forcings combined with internal dynamic variability that is much larger than typically simulated." The models produce substantially less variability of critical north Atlantic sea surface temperature than observed. From this, I conclude that the deficiencies of models to represent the internal dynamics of the climate system adequately can lead to erroneous attribution of climate variations and change to internal (*sic*) and external forcing factors.

Response: Some – but not all – models do indeed "produce substantially less variability of critical north Atlantic sea surface temperature than observed", at least on decadal time scales. Other models, such as HadCM3 (*Knight et al*., 2005)[15], have been shown to capture many of the salient features of the observed "Atlantic Multidecadal Oscillation (AMO)". In other regions, such as the tropical Pacific, there is credible scientific evidence that many current models actually overestimate observed decadal-timescale SST variability (*AchutaRao and Sperber*, 2006).

Even if climate models seriously underestimated internal variability for some limited spatial region, this would not affect any of the conclusions drawn in this Report, which relate to similarities between modeled and observed temperature changes at large spatial scales. As we explicitly point out in Section 4.4 [Page 100, column 2, para. 1]:

"Model errors in internal variability can bias detection results, although most detection work tries to guard against this possibility by performing "consistency checks" on modeled and observed variability (*Allen and Tett*, 1999), and by using variability estimates from multiple models (*Hegerl et al*., 1997; *Santer et al*., 2003a,b)."

[15]Knight, J.R., R.J. Allan, C.K. Folland, M. Vellinga, and M.E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophysical Research Letters*, **32**. L20708, doi:10.1029/2005GL024233.

Winstanley, CH5-4c: Chapter 1, page 11, lines 230-231 recognizes that "unforced variability could be substantial" and states that "Chapter 5 provides more details on models and their limitations (see particularly Box 5.1 and 5.2)". However, **Chapter 5** does not incorporate recognition of the importance of internal variations in its discussions of the causes of reported changes in vertical temperature profiles. It should do so.

Response: It already does so! See **Response to Winstanley, CH5-4a**.

Winstanley, CH5-4d: Chapter 2, page 31, lines 556-560, recognizes the importance of internal modes of climate variability on regional scales and states that identifying the patterns and separating the influences of such modes from the warming signal is required.

The extent to which the report is able to identify the internal modes of climate behavior and separate these from internal and external forcings should be addressed in **Chapter 5** and summarized in the **Executive Summary**.

Response: See **Response to Winstanley, CH5-4a**. Chapter 5 does address the problem of separating externally-forced signals from internally-generated climate variability. This problem is at the core of all detection and attribution work, as discussed at length in Section 4.4 and Box 5.5.

Winstanley, CH5-4e: Kunkel *et al.* (("Can CGCMs simulate the Twentieth Century "Warming Hole" in the central United States?", in press, *Journal of Climate*, and attached to these comments) demonstrate that model simulations, even simulations from the same model, are highly sensitive to initial conditions. **Chapter 5** should incorporate this reference on page 14 and include as a Key Finding on model limitations (section to be added) the fact that noticeably different regional simulations of changes in atmospheric temperature profiles probably can result from model simulations that employ the same atmospheric model and the same climate forcings.

Response: Sensitivity to initial conditions is discussed throughout Chapter 5. Here are a few examples:

- ⇒ "We refer to these subsequently as "20CEN" experiments. Since the true state of the climate system is never fully known, the same forcing changes are applied *n* times, each time starting from a slightly different initial climate state. This procedure yields *n* different realizations of climate change. All of these realizations contain some underlying "signal" (the climate response to the imposed forcing changes) upon which are superimposed *n* different manifestations of "noise" (natural internal climate variability)." [Page 94, column 1, first complete paragraph]
- ⇒ "This illustrates the need for caution in comparisons of modeled and observed atmospheric temperature change. The differences evident in such comparisons have multiple interpretations… They may also be due to different manifestations of natural variability noise in the observations and a given CGCM realization." [Page 96, column 2, para. 4; Page 97, column 1, para. 1]

⇒ "In addition to model forcing and response uncertainty, the 20CEN ensemble also encompasses uncertainties arising from inherently unpredictable climate variability (Boxes 5.1, 5.2). Roughly half of the modeling groups that submitted 20CEN data performed multiple realizations of their historical forcing experiment (See Section 2 and Table 5.1)… Such multi-member ensembles provide valuable information on the relative sizes of signal and noise." [Page 105, column 2, para. 1]

⇒ "The model ensemble encapsulates uncertainties in climate forcings and model responses, as well as the effects of climate noise on trends." [Page 106, column 2, para. 1]

Note also that Recommendation 1 (page 91) now explicitly mentions initial condition differences as a contributory factor to differences in simulations of 20th century climate change.

Winstanley, CH5-4f: Chapter 5, part of a much needed discussion on model limitations (parallel to the extensive discussions on the limitations of observational data throughout the draft report) should be discussion of the implications of a lack of explicit treatment of internal variability as a cause of climate variability and change and the lack of explicit treatment of model initialization. Also, different treatment of internal variations of the climate system and initial conditions should be included in the list on Page 7 of Chapter 5 of the reasons why climate simulations differ.

Response: There is no "lack of explicit treatment of internal variability as a cause of climate variability and change." See **Response to Winstanley, CH5-4a,d,e**. Differences in initialization procedures are a highly technical issue that is best dealt with in the Synthesis and Assessment Product on climate modeling. See **Response to Winstanley, CH5-1**.

Winstanley, CH5-4g: A key finding of **Chapter 5** should be that it is important to account for model uncertainty and limitations in comparisons between modeled and observed temperature changes. In the present draft, it is recognized only that observational uncertainty should be accounted for (page 6, lines 128-130).

Response: The Reviewer is incorrect. Model uncertainties and limitations are prominently discussed in interpreting the results of comparisons with observations. Here are a few examples:

⇒ "This illustrates the need for caution in comparisons of modeled and observed atmospheric temperature change. The differences evident in such comparisons have multiple interpretations. They may be due to real errors in the models, errors in the forcings used to drive the models, the neglect of important forcings, and residual inhomogeneities in the observations themselves." [Page 96, column 2, para. 4; Page 97, column 1, para. 1]

⇒ "One possible interpretation of these results is that in the real world, different physical mechanisms govern amplification processes on short and on long timescales, and models have some common deficiency in simulating such behavior." [Page 115, column 1, para. 2]

⇒ " 'Model error' and 'observational error' are not mutually exclusive explanations for the amplification results shown in Figures 5.6C and D." [Page 115, column 1, para. 4]

⇒ "These results could arise due to errors common to all models…" [Page 90, Key Finding 6, bullet 5]

Chapter 6 Comments and Responses:

MacCracken CH6-1, Page 9, Line 229: Indeed, it would be appropriate to go back to relook at this apparent finding of a climate regime shift (perhaps more appropriately named a shift in the atmospheric circulation) to see if it is as significant as is suggested by the phrasing in this report once the data are fully corrected and considerations are given to: how the shift was sampled by the existing network; whether this was a chance confluence of opposing anomalies; whether such shifts are rare or common in the longer record; etc. In my view, this report gives too much credence to this really being a shift, given that it had no substantial influence on surface temperature, etc.
Michael MacCracken, Climate Institute

Response: Inserted the word apparent before regime shift within this recommendation.

MacCracken CH6-2, Page 12, Line 293: Change "will inevitably lead to better future reanalyses" to "will in the future inevitably lead to better reanalyses" as we are not doing reanalyses of the future.
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH6-3, Page 15, Line 358-360: This sentence seems very poorly phrased, seeming to imply that in the future our reassessments might lead us to reconsider if there has been a human influence. It is fine to call for further detection and attribution studies, and hopefully these will be able to better apportion the changes to various influences, but a rephrasing is needed to make clear that there is no expectation that this will make the human influence disappear.
Michael MacCracken, Climate Institute

Response: Sentence has been modified to: "Finally, detection and attribution studies should be undertaken using this new range of observations and model-based estimates to refine our understanding of human-induced influences on climate (C5)."

MacCracken CH6-4, Page 18, Line 418: Change "satellite" to "key instrument" as it is not normally the satellite that failed.
Michael MacCracken, Climate Institute

Response: Done

MacCracken CH6-5, Page 19, Line 444: Change "would" to "need to" to make the point more strongly.
Michael MacCracken, Climate Institute

Response: Done

Trenberth CH6-1, Amen to most of this. This is the most important chapter in the whole document. Unfortunately the document is long and it is near the end and less likely to be read. It has no figures to make it punchy. I strongly urge some form of diagram, figure or table be used to summarize and make for an attractive finale.

Response: *See New Figure 6.1 which summarises the recommendations and their interlinkages.*

The other major change I would make is to add a major recommendation for reprocessing of many data, including satellite data. This might come under Section 3, line 266, as an addition and this would deal with things like water vapor, precipitation, clouds, radiation, surface winds, sea ice, etc. These are all single variables and all have problems but enough is known to reprocess these and produce better results. There is also a need to then bring them together and make sure they are physically consistent. These are all then fed into reanalyses. Please see the WCRP Observation and Assimilation Panel (WOAP) web pages.
Kevin Trenberth, National Center for Atmospheric Research

Response: Recommendation 3 has been expanded to include these ideas and an explicit reference added to the WCRP plans to reflect this concern.

Trenberth CH6-2, Page 13, Lines 303-309: I don't endorse these suggestions especially given known sonde problems. Please do not make these mandatory. A key ingredient is the use of OSEs to calibrate the impact of new or different observations on the analyses. Please emphasize these much more.
Kevin Trenberth, National Center for Atmospheric Research

Response: These have been re-ordered, only one raobs recommendation retained, and it has been re-emphasized more strongly in the redraft that they are a far from exhaustive list.

Changes made by the authors

1. We have numbered all the recommendations as agreed before Public Review. This helps greatly with Fig 6.1.

2. We have added a new sentence to introduce Fig6.1 at line 76 of the current draft.

3. We have modified recommendation 1 to include the idea of several independent research teams.

____

Statistical Appendix Comments and Responses:

MacCracken App. A-1, Page 5, Line 74: I would encourage revising and inserting a phrase so this reads: "be strictly linear, so the results can be quite misleading, but the linear trend can sometimes provide a simple way …" There is a lot of abuse of linear trends (like through the 20th century), and the report should be making clear that misleading results can occur.

Response: We disagree that a linear change representation can be "quite misleading". Over the study period (1958 onwards) the expected anthropogenic changes are near linear, so a linear representation is just the opposite of "misleading". The text has been modified to clarify this.

MacCracken App. A-2, Page 6, Line 90-92: It would be helpful to have a graph showing this PDO switching (and using the newly revised data sets). I think it much appropriate to be calling this shift the PDO to emphasize that it is most apparent in the Pacific (so not globally) rather than how this is referred to elsewhere in the text, calling it a previously identified climate regime shift. Here, the phrasing is about changes in variability—and it should be added, mainly in the atmospheric circulation and not in the surface temperature.

Response: The reviewer's criticism of the wording "previously identified regime switch" is understandable. We do not agree that this is "mainly in the atmospheric circulation". Minor text changes have been made to cover these items. It is not possible to add a new Figure. In any event, we do not consider this necessary, since this is a minor point. Further, the data are illustrated elsewhere in the Report, and also in the Executive Summary. As a guide to the reader, we have added a reference to Fig. 3.2a where the apparent step is shown.

MacCracken App. A-3, Page 7, Line 107-110: This seems to me a serious misuse of the word "trend." What is being referred to is a difference and not a trend (which is a rate). I think it very unscientific to use the word trend as used here.

Response: This refers to the use of "total trend" on line 107. This has been deleted, and only "total change" is now used.

MacCracken App. A-4, Page 18, Line 327: I would insert the phrase so this reads: "observed data are reliable, which is not always the case), we …" Given the problems that are reported in this assessment with datasets, indicating that they are not always reliable would seem very appropriate.

Response: The suggested text change has been made.

Trenberth App. A-1 Much of this appendix is basic text book material and should not be included. It should be shortened by 80%. I agree with most of it except that it misses one vital point on how autocorrelations are computed, and the material on trends of differences ought to be in the main report. But I strongly disagree with the conclusions to omit error bars.

Response: This Appendix was added to the original Report in response to comments from the NAS review panel. It is true that most of the theory (but, of course, not the examples) can be obtained from textbooks (although not from any single textbook). However, the reason for including a comprehensive Appendix was to make this material available in a self-contained form for readers not familiar with Statistics.

The concern regarding missing details on the calculation of autocorrelations arises because the reviewer apparently did not notice the reference to footnote 5 on line 400, where the requested information is given.

The reviewer suggests that the material on trends of differences should be in the main Report. In the opinion of the author team, the main issue is to make this material available somewhere in the Report, in a way that does not upset the flow of the exposition. Our judgment is that this is best achieved by putting this material (as now) in this Appendix.

The omission of error bars was a decision made by the author team and applies to the whole Report. This was partly an issue of ensuring that the Figures were not too "messy" – in most cases the inclusion of error bars would have made the Figures much more complex and difficult to understand. Error bar information is given in Tables – see Chapter 3, Tables 3.2, 3.3, 3.4 and 3.5.

Trenberth App. A-2, Page 6, Line 79, Figure 1: is useful but citing trends to 3 figures and not giving error bars is absurd, also Line 105 below.

Response: Often, in pedagogical texts, extra precision is required for numerical reasons – where precision should not be confused with accuracy. This is the reason for using 3 decimals in some places, but the reviewer is correct in noting that 3 decimals should not be used everywhere. The point is now clarified in the text. We have replaced all 3 decimal results in Figures 2 and 3 with 2 decimals, and added the 95% C.I. values (from Table 3.3).

Trenberth App. A-3, Page 7, Line 105. Try computing trends with 1 or 2 years removed from each end to see how stable they are.

Response: If the reviewer's suggestion is followed, the trends are stable. It is true that one could select end points to give noticeably different trend values. For example, the trend in the surface data over early 1985 to mid 1998 is larger than the trend over the full data period. However, trends are only given here over the full data period, and these values are robust to 1 or 2 year omissions from either or both ends of the record

Trenberth App. A-4, Page 7, Line 102: What about a sine curve with one cycle, which can be fitted with a linear trend that has a correlation 0.71 and is highly statistically significant. Add half a cycle at each end and the linear trend is again statistically significant but now with reversed sign!

Response: This is an elegant, but purely artificial example, of no relevance to the data sets used in this Report. The possible inadequacies of linear trends as a data descriptor are clearly stated in the preceding paragraphs of the text, and elsewhere.

Trenberth App. A-5, Page 10, Lines 159-160: this is not true if data prior to 1979 are added.

Response: If earlier data are added, the trend value and the total change value both change. If one goes back to pre-1976, especially for tropospheric data, the choice of start point has a more noticeable effect on the trend. However, the example here concerns data from 1979 only, so it makes no sense to start earlier.

Trenberth App. A-6, Page 17, Line 301: The issue is not statistical noise in dataset but the appropriateness of the linear trend as a model for the data.

Response: The reviewer claims that the "issue is not statistical noise". In fact, this is precisely the issue that this part of the text is addressing. Whether or not a linear trend is an appropriate descriptor for the data is a separate issue. This second issue is addressed elsewhere in the Appendix, on a number of occasions. Some new text on this has been added in the revised text noting that, over the study period (1958 onwards) the expected anthropogenic changes are near linear. This further justifies the use of linear trends in the present context.

Trenberth App. A-7, Page 21, Line 403: It does not say how the autocorrelation at lag 1 is computed. In particular given a trend, the r1 value can be large but not indicative of an AR1 process at all. The data should be detrended before computing r1.

Response: The concern regarding missing details on the calculation of autocorrelations arises because the reviewer apparently did not notice the reference to footnote 5, which covers this point.

Trenberth App. A-8, Page 26, Line 469: Why isn't Fig. 3 in the main report? The message here is important.

Response: The reviewer needs to address this question to the appropriate Chapter author. This Figure and the accompanying text on differences in trends is included in the Appendix largely because the material was not covered in any detail in any earlier Chapter. (See also response to A-1 above.)

Trenberth App. A-9, Page 29, Lines 516-518: I strongly disagree with this statement.

Response: The statement referred to here relates to the omission of error bars, and the statement that individual error bars can be misleading when (as here) the primary concern is with the comparison of time series. The reasons for this are explained in the text. For more on this point, see response to A-1 above and to A-11 below.

Trenberth App. A-10, Page 33, Lines 573-574: It is the model used as in line 574, not the noise that is the issue.

Response: There are two issues here, the choice of model and the uncertainty in fitting the chosen model. We have chosen to describe all data sets used in the Report with a linear model. The reasons for this are explained in many places (see also A-6 above).

Trenberth App. A-11, Page 35, Lines 627-630: And vice versa even more so. It is always misleading not to show the error bars.

Response: As noted in the text (see response to comment A-9), it can also be misleading to show error bars on individual trends when the primary concern is the differences in trends between data sets. The key point here is not <u>how</u> the uncertainty information is illustrated (e.g. as "error bars"), but <u>whether</u> the information is given. This information <u>is</u> given for all observed data trends given in this Report: see the Tables in Chapter 3. For completeness, we have added this information to Figures 1, 2 and 3 of this Appendix.

<u>Responses to comments on Appendix B and Glossary</u>

MacCracken App. B-1, Page 1: As indicated more fully in the general comments, this ordering of the authors of this report seems to me seriously flawed. For this report to be credible, it MUST be clear that the authors are the scientists who wrote it and that they are in charge of it and that they are the ones who should be listed when the report is being referred to. To be listing the various directors, the technical editor, the graphic designer, and the technical support person ahead of the scientists who wrote the report is totally inappropriate. This is a report by Thomas Karl and other scientists and they should be receiving the prominent billing—the others should not even be listed as members of this "Assessment/Synthesis Product Team"—the two directors can be separately referred to as representatives of the sponsoring CCSP or something and the others, after the listing of authors, as support for this particular product, but the listing here is totally inappropriate.

Michael MacCracken, Climate Institute

Response: Appendix B has been removed from the Report. The Author Team is now listed on a separate page immediately following the Table of Contents.

MacCracken App. B-2, Page 1: This Appendix needs to make clear how this product meets the revised guidelines for these assessment products, so indicating how this meets FACA if it is to be published as a report of the CCSP set of agencies. If instead, it is to be presented as solely a report of these authors as a scientific team, then it might instead be put out as a NOAA NCDC report or something similar—so clearly identifying this as a scientific report/article and not some agency approved product. In either case, this appendix needs to make the case given the guidelines, FACA, and how it was prepared.
Michael MacCracken, Climate Institute

Response: The report was prepared under NOAA leadership on behalf of the CCSP, in full accordance with FACA guidelines. No text modification is necessary since, as noted in the previous response, Appendix B has been removed from the Report.

MacCracken Glossary-1, Line 62: This definition of "latent heat of water" needs to be amplified to be the latent heat of fusion and the latent heat of vaporization/condensation and the phase change for each identified. This definition seems to me too incomplete.
Michael MacCracken, Climate Institute

Response: The definition of "latent heat of water" has been amplified as suggested.

MacCracken Glossary-2, Page 3, Line 97: The term "uncertainty" should be defined, making clear that it represents the range of the likely value—so does not mean something is totally unknown, but is known to within some specified value with some likelihood.
Michael MacCracken, Climate Institute

Response: A definition of the term "uncertainty" has been added to the Glossary.

General Comments and Responses:

Douglas GEN-1: Comment: "...The report however is flawed, perhaps fatally. It is not *policy-neutral* – as required by NRC. Because of this it cannot *provide the best possible scientific information.* The prime example is from the Executive Summary

> *Given this range of results, there is no conflict between observed changes and the results from climate models.*

This statement is an assertion and is the subject of vigorous current research. It is obvious that whoever wrote it must believe it. However, there are many -- including some authors of this report -- who would certainly disagree. In any case, it is not *policy-neutral*. In addition, it is a violation of NRC's point 2 above: *If any recommendations are based on value judgments or the collective opinions of the authors, is this acknowledged?"*

Response: This document is policy neutral. The phrase "policy neutral" means that no specific policy actions are stated. This does not mean that the text should not be "policy relevant". Indeed, the whole point of this Report is to provide a review of the state of the science in order to guide policy. The quoted statement regarding "no conflict" between observations and models is a scientific statement, assessing the state of the science. It does not recommend any specific policy action. The evidence in support of this statement is clearly stated both in the Executive Summary and the body of the Report (primarily in Chapter 5).

Douglas GEN-2 Comment: I believe that this report may be fatally flawed because of the known views and agendas of some of the authors, and esp. the Lead Authors of Chapter 5 and the Exec Summary. These individuals would naturally favor their particular view in this report [Four of the lead authors have among themselves 197 citations to themselves in the report]. Again -- not neutral.

Response: The large number of publications of the authors for this report is testimony to the fact that the authors are indeed experts in the areas of emphasis for this report. Peer-reviewed articles are the basis for scientific assessments. All authors were vetted prior to the assignments including a period of open public comment.

Douglas GEN-3 Comment: The report also fails on the point: *Are the findings and recommendations adequately supported by evidence and analysis.* Some of the major results in Chapter 5 are based upon on unpublished work: Figs 5.3, 5.4. 5.7 and tables 5.3 and 5.4. Because these results have not been peer reviewed, they must be considered as only a scientific hypothesis. I wanted to test this hypothesis so I requested the numerical values from which these figures and tables were made. I was flatly refused by one of the lead authors (Santer). All of the others that I contacted referred me to the editor of the report. My requests to him were not answered. Until I or someone receives this data for review the results of Chap 5 should be considered only an unproven assertion.

Response: All of the data and model results used in the figures and Tables are openly available. This Assessment/Synthesis Report has in a number of figures and tables aggregated the data in ways that have not been previously presented. This is a common and expected practice in conducting Assessments e.g., WMO/UNEP Assessments, the US National Assessment of Climate Variability and Change, etc. In response to his request, Dr. Douglas was informed that, the output of the model simulations in Chapter 5 can be obtained at the following web site: http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php This site hosts the IPCC models that are being used in the 2007 assessment and these were used in CCSP Synthesis and Assessment Product 1.1. To ensure that he could select the correct model simulations we used in this report, he was provided with the specific information needed to select the appropriate model simulation.

MacCracken GEN-1 Comment: Overall, from a purely scientific perspective, this assessment provides a very well done scientific overview of the topic. However, this draft does seem to underplay the significance of the most recent papers in helping to resolve the key issues under investigation, specifically in identifying why some of the datasets developed over recent years are very likely to have flaws. Issues of science are not something one votes over, and this review in some of its analyses seems to present results in terms of how many datasets find one result or another without critically reviewing whether all of the datasets being mentioned still merit being considered fully credible.

Response: This refers to the comparison of model and observed temperatures in the tropics, and the warming in the troposphere relative to the surface. The primary result, as explained and summarized in the Report, is that models show a greater warming trend aloft, while most of the observations do not show this amplification. This discrepancy is most clearly illustrated in Chapter 5, Fig. 5.6C. Evidence is also given in the Report that some observed radiosonde data show a cooling bias in the tropics, which, if this applied to the data sets used in the Report, would bring these data sets into closer accord with the model results. The considered view of the expert author team is that these issues are not yet resolved. Nevertheless, the author team does conclude that this difference between models and observations is more likely to reflect errors in the observed data than in the models. The following quote from the Executive Summary summarizes the author team's assessment:

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational datasets leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report favors the second explanation." Given the current state of the science, a stronger statement cannot be justified.

MacCracken GEN-2 Comment: In that the CCSP assessments are intended to provide information for policymakers [given that they are said to be in response to the relevant section of the US Global Change Research Act], this draft of this assessment seems to me seriously deficient in providing a historical perspective of this issue and a critical evaluation of past claims that have been made about the supposed accuracy of the early versions of the datasets and what the available data were purported to indicate about scientific understanding of climate change. For more than a decade, some of the datasets have been purported to be highly accurate and to indicate that the model simulations must have serious shortcomings. This report shows that those claims, which were made not only in the scientific community but were picked up and loudly exclaimed by some politicians and a number of industrial organizations, were based on a seriously flawed analysis because of flaws in the satellite record. I would urge that, at the least, a table or an appendix be added that gives a timeline of the history of the corrections that have had to be made to the satellite record and that indicates the past claims that should therefore be discounted (and that the IPCC rightly did not accept at the time—leading to some misdirected criticism of their careful approach). Such a historical review of changes that had to be made as understanding developed and its effect on the conclusions was presented in the case of the stratospheric ozone assessments, making clear what the effect of each advance was in improving understanding and estimates of change. This issue of the supposed disagreement between surface and tropospheric observations and model results has been at the scientific heart of much of the political discussion, and this assessment needs to recognize this and deal with it, and not simply present the current understanding as if the near sordid past criticism arising in regard to this issue did not exist. Were this all going on in the biomedical field, I rather suspect that a number of the past papers would have been withdrawn or would now have notices attached indicating that they are no longer valid. In that this was not done, it seems to me that this assessment needs to provide a historical perspective that makes clear that the criticisms made in the past of the general scientific understanding of climate change and of the evaluations done by the IPCC are not justified by what has proven to be improved understanding and that the skepticism regarding the early presentations of the MSU data and associated conclusions was justified.

Response: A table showing the adjustments made to the MSU data has been added to Chapter 2.

MacCracken GEN-3, The issue of who produced this report is not clarified by Appendix B. The preface, in lines 240-242 indicates that the Appendix presents a "fill list of this Reports' [sic] authoring team" but when one goes to the appendix one gets instead a list of something called "Members of the Assessment/Synthesis Product Team" that, except for the chief and associate editors are not the authors of the report at all (there is almost the presumably misleading implication by the formatting here that this team will be responsible for all such reports rather than just this one). It almost seems as if the citation to this report, given the ordering of the listing in the appendix, would be to Mahoney and Moss, yet they are not the authors and listing them as the main people associated with the report would frankly reduce the report's credibility. To rectify this situation, the listing of those who helped make the report happen as part of the "Assessment/Synthesis Product Team" should be listed somewhere else (in an Acknowledgments section, on a separate page dealing with availability of the report, or something); it is simply not appropriate to be listing the graphic designer, technical support, and other staff people ahead of the scientific authors of the report. Indeed, the real authors of the report should be given the prominent recognition that they deserve and that will help to provide credibility to this report! In addition, a preferred citation for the report needs to be provided, being something like: (a) Karl, T. R., C. D. Miller, and W. L. Murray, et al., 2006: etc. for the report as a whole; and (b) suggesting that reference should be made to the individual chapters and their authors as appropriate. It is absolutely vital to the credibility of the process that scientists be the lead individuals associated with the assessment, and not a political appointee and a member of his staff, no matter how pure their efforts. And this appendix also needs to make clear how this particular structuring of everyone involved meets the FACA requirements for generation of this report—clearly, those listed as the "Members of the Assessment/Synthesis Product Team" are not all a federal advisory committee.

Response: The reviewer comments are consistent with the intent of the Report. The final version of the report will have the appropriate credits on separate pages and will highlight the Science Team as authors of the report. In the Word document this was not clear, but it will be corrected in the final lay-out of the report.

MacCracken GEN-4 Comment: Throughout the report there is an almost reverent referral to the "previously identified climate regime shift" of the mid-1970s. Yet, the text also indicates that, while this shift is apparently evident in the radiosonde record, it is not evident in the surface temperature record. Given the various shortcomings that have been found with the radiosonde record over the past few decades (i.e., instrumentation problems, biases, coverage issues, etc.), it would really seem as if this supposed regime shift should be reexamined and reconfirmed with the new data. It also would seem to be appropriate to determine if this is really a shift, or was a response to, for example, a volcanic eruption or two, to cleaning up of sulfate emissions (and consequent effects on the circulation pattern), or to some other factor. It is also not at all clear that it should be labeled a "climate regime shift" as opposed to a shift in the atmospheric circulation pattern—it really gets down to what is meant by "climate." If one is going to have a change in circulation, predominantly in one region of the world, be called a "climate regime shift" then one really needs to decide how many others of those have occurred and then be consistent. In my view, the report would have a better chance of standing the test of time if it did not even raise the issue of the supposed "climate regime shift"—for then one should also be saying something about there being earlier shifts, etc., and one gets into what is a variation, fluctuation, and shift—if they are different at all. In addition, it is not at all clear whether this shift, if it occurred, was natural or human-induced. So, in my view, the whole question should be avoided by dropping that reference to a "previously identified climate regime shift."

Response: There is much published evidence for coupled atmospheric-oceanic regime shifts focused in the North Pacific region but with a wider influence (e.g. Deser et al., *J Climate*, **17,** 3109-3124 (2004).

Meyer GEN -1 Comment: The geothermal flux is being excluded from this study. The geothermal flux is the cause of major climate changes over long time periods and it is a very important detail even though it is not regarded as such by modelers. Ignoring the geothermal flux will introduce errors in the model that could be avoided. I hope these comments are taken as constructive observation and some new factor is invented to introduce the geothermal flux to the models of climate change. Doing this will improve the result.

Response: We know of no published evidence of systematic, large-scale changes in geothermal flux over recent decades.

Pielke, Sr., GEN-1, Surface temperature data. One of the examples of the lack of balance in the Report is the acceptance of the trends of surface temperature data as robust (e.g., see pages 6-8 in the CCSP Chapter 3). This is an example of accepting observations where they agree with the models, without investigating the data further. The NRC Review commented on this in one of their comments:

"It should also be clearly emphasized that data is being used to test models and not vice-versa."

An example of where the Committee failed to investigate other explanations for surface temperature trends is the following;

"Most of the recent warming has been in winter over the high mid-latitudes of the Northern Hemisphere continents, between 40 and 70° N (Nicholls et al., 1996). There has also been a general trend toward reduced diurnal temperature range, mostly because nights have warmed more than days. Since 1950, minimum temperatures on land have increased about twice as fast as maximum temperatures (Easterling et al., 1997). This may be attributable in part to increasing cloudiness, which reduces daytime warming by reflection of sunlight and retards the nighttime loss of heat (Karl et al., 1997)……."[16]

Thus it is in the higher latitudes over land in the winter where "most of the recent warming" has occurred. However, as shown in a new paper[17], any nighttime warming within the boundary layer will result in an amplified near-surface positive temperature trend. An increase in cloudiness as reported in Karl et al. (1997) is one way in which nocturnal boundary layer cooling is reduced. Since night at higher latitudes in the winter frequently have stably stratified boundary layers, this issue should have been discussed in the Report. It was not (even though an earlier version of the paper was distributed to the Committee), apparently because this was a geographic area where the existing observations agree with the models.

To use these nocturnal surface temperature trends as part of the calculation of recent global warming, therefore, overstates that warming.

The major issues with the surface temperature trend data that have not been addressed satisfactorily in the CCSP Report are summarized below:

1. The temperature trend near the surface is not height invariant[18].

The influences of different lapse rates, heights of observations, and surface roughness have not been quantified. For example, windy and light wind nights should not have the same trends at most levels in the surface layer, even if the surface-layer averaged temperature trend was the same. This raises questions regarding the conclusions of the Parker (2004) and Peterson et al. (1999) papers that are specifically cited in Chapter 3 of the CCSP Report as supporting the justification of the robustness of the surface temperature data.

[16] http://www.ucsusa.org/global_warming/science/early-warning-signs-of-global-warming-heat-waves.html
[17] Pielke Sr., R.A., and T. Matsui, 2005: Should light wind and windy nights have the same temperature trends at individual levels even if the boundary layer averaged heat content change is the same? Geophys. Res. Letts., 32, No. 21, L21813, 10.1029/2005GL024407.
[18] Pielke Sr., R.A., and T. Matsui, 2005: Should light wind and windy nights have the same temperature trends at individual levels even if the boundary layer averaged heat content change is the same? Geophys. Res. Letts., 32, No. 21, L21813, 10.1029/2005GL024407.

Question: What is the bias in degrees Celsius introduced as a result of aggregating temperature data from different measurement heights, aerodynamic roughnesses, and thermodynamic stability?

Response: Since 1979, T_{min} has not warmed relative to T_{max} globally: see Vose et al., Geophysical Research Letters, 32, doi: 10.1029/2005GL024379 (2005). In the tropics Vose et al. do not make explicit calculations but scrutiny of their global map (their Figure 4) shows no evidence of relative warming of T_{min} relative to T_{max} in the tropics or extratropics separately since 1979.

The trend for HadCRUT3 global annual anomalies from 1979-2004 was 1.80 degrees/century. Halving the trend from Eurasia >45N in October-March reduces the global annual trend from 1979-2004 to 1.76 degrees/century. Removing the trend entirely from Eurasia >45N in October-March reduces the global annual trend from 1979-2004 to 1.72 degrees/century. The reason for this result is that warming over the period 1979-2004 is almost ubiquitous globally with the exception of most of Antarctica and a little of the Southern Ocean adjacent to it.

The heights of the surface temperature observations are largely fixed, so an observed warming trend is not invalidated by any variation of trend with height.

The cited paper by Pielke and Matsui appears to be an idealized calculation for some unspecified extreme nocturnal condition e.g. that might occur over the Prairies or Siberia. Any attempt to quantify this effect globally or over the tropics requires a full assessment of the real mix of weather events that have occurred. This can only be approximately achieved by very carefully running a full climate model with a high-resolution boundary layer. Furthermore Pielke and Matsui do not take account of the fact that the radiative imbalance driving global warming is fundamentally at the tropopause rather than at the surface. The long-term average radiative imbalance at the land surface is very small when greenhouse gases are increasing, because increasing downward longwave radiation from the warming atmosphere balances increased upward longwave radiation from the warming surface. So Pielke and Matsui's paper may have limited application.

Pielke, Sr., GEN-2: The quantitative uncertainty associated with each step in homogeneity adjustments needs to be provided[19]:

Time of observation, instrument changes, and urban effects have been recognized as important adjustments that are required to revise temperature trend information in order to produce improved temporal and spatial homogeneity. However, the quantitative magnitudes of each step in the adjustments are not reported in the final homogenized temperature anomalies. Thus the statistical uncertainty that is associated with each step in the homogenization process is unknown. This needs to be completed on a grid point basis

[19] Pielke Sr., R.A., T. Stohlgren, L. Schell, W. Parton, N. Doesken, K. Redmond, J. Moeny, T. McKee, and T.G.F. Kittel, 2002: Problems in evaluating regional and local trends in temperature: An example from eastern Colorado, USA. Int. J. Climatol., 22, 421-434.

and then summed regional and globally to provide an overall confidence level in the uncertainty. This approach is ignored in the Report.

Question: What is the quantitative uncertainty in degrees Celsius that are associated with each of the steps in the homogenization of the surface temperature data?

There are several other issues that are mentioned in the Report as being issues, but are dismissed as unimportant on the larger scales, but without quantitative assessment of their importance. These effects include the role of poor microclimate exposure[20] and the effect of temporal trends in surface air water vapor in the interpretation of the surface temperature trends[21].

Response: Maps of surface temperature trends show strong coherence between adjacent grid-boxes, even in the tropics (Figure 3.6d), demonstrating that the statistical uncertainty of trends for individual grid boxes is small compared with the magnitude of the trends. Many of the uncertainties mentioned are structural and therefore likely differ somewhat between different datasets. A complete analysis of uncertainty does require the suggested work to be done but there is no published literature. Our assessment is that on the large space and time scales which are the subject of this report, these extra uncertainties are small. In future, when smaller space and time scales are investigated, the uncertainties are likely to be greater.

Surface and atmospheric water vapor is very important for the full understanding of temperature trends but this is not directly relevant to this report. The Recommendations for the future in Chapter 6 reflect the importance of water vapor for greater understanding.

Pielke, Sr., GEN-3: There is also the question of the independence of the data from which the three main groups create global data analyses (page 8 Chapter 3). Figure 3.1 presents the plots as "Time series of globally-averaged surface temperature….datasets." The inference one could reach from this is that the agreement between the curves is evidence of robustness of the trends plotted in the Figure. The reality is that the parent data from which the three groups obtain their data is essentially the same.

The Executive Summary even states "*Independently-performed adjustments to the land surface temperature record have been sufficiently successful that trends given by different data sets are very similar on large (e.g. continental) scales.*"

---

[20] Davey, C.A., and R.A. Pielke Sr., 2005: Microclimate exposures of surface-based weather stations - implications for the assessment of long-term temperature trends. Bull. Amer. Meteor. Soc., Vol. 86, No. 4, 497–504.

[21] Pielke Sr., R.A., C. Davey, and J. Morgan, 2004: Assessing "global warming" with surface heat content. Eos, 85, No. 21, 210-211.

The data used in the analyses from the different groups, however, are not different but have very large overlaps! This statement in the Executive Summary is incorrect and misleading.

The report needs to answer this question,

Question:What is the overlap in the raw data that utilized by the three groups?

The best estimate that I am aware of has a 90-95% overlap. The analyses from the three groups are hardly independent assessments, and this should not be hidden in the report.

The overlap is particularly important for the grid points analyzed in the analyses where only 1 or 2 observational data points exist. We have documented for the tropical land areas, for example (20N to 20S) about 70% of the grid points have had zero or less than one observation site![22]. Thus to compute an average surface temperature trend over land in the tropics, which is the area where the report narrowly focuses, almost all of the raw data used on the three analyses is from the same source. Thus to present a Figure to purportedly illustrate uncertainty in the surface temperature trends is misleading.

Response: It is true that there are substantial, though not complete, overlaps between the data sources used in the three global surface temperature analyses. But the unimportance of this problem is shown by the abovementioned observation that the trends show strong coherence between adjacent grid boxes, even in the tropics (Figure 3.6d). Thus if the three global surface temperature analyses were to be deliberately based on different, well-distributed sets of one third of the grid boxes, their global trends would still be in good agreement. Moreover it was shown by Jones et al. (1997) that on the annual global space scale there are only about 60 degrees of spatial freedom in surface temperature anomalies.

We note also that the three global surface temperature analyses are based on different methods, corroborating the validity of the analyses. The MSU groups use identical input data and yet yield estimates that differ by the same magnitude as they searched for signal. Why the surface record is being systematically identified as being a problem because of raw data overlap when this applies to all datasets is somewhat of a mystery. The analysis in this report implies that structural uncertainty is greater aloft than at the surface. It is not an altogether surprising result. The surface record is based upon instruments which remain in-situ, are generally calibrated and maintained on a regular basis, and observing practices are relatively constant. Monitoring of the upper-air is achieved either by "fire and forget" single-use radiosondes or by satellites which have at most a lifetime of several years. It is much easier to change practices and introduce significant non-climatic influences in these latter records which very likely explains the larger spread in these estimates than those at the surface.

A final overarching question is

---

[22] Davey. C. and R.A.Pielke Sr., 2005: Comparing station density and reported temperature trends for land surface sites, 1979-2004. (in preparation).

Question: What is the value-added of using annually-averaged surface temperatures to assess global climate system heat changes ("global warming") over the last several decades in lieu of assessing the regional, zonally-average and global trends in ocean, and other climate component heat storage in units of Joules?

Response: The report's focus is consistent with the topic addressed by this Synthesis and Assessment Product: "Temperature trends in the lower atmosphere – steps for understanding and reconciling differences". Global mean temperature is a common and simple index used regularly in discussions of global-average warming. As stated in the Preface, previous discrepancies between surface and tropospheric temperature observations challenged the correctness of climate model simulations and the reality of greenhouse gas-induced global warming. The report discusses the considerable progress that has been made in resolving many of the earlier discrepancies.

The reviewer raises an interesting question that was, in fact, discussed during the meetings of the Author Team. However, the alternate approach the reviewer suggests would entail research beyond the scope of this assessment and was not considered to be feasible at this time. For example, changes in the heat content of the climate system have not been systematically evaluated. Heat content is dependent on the moisture content of the climate system. Changes in this quantity have not been regularly calculated by the science community, probably because of a dearth of readily available reliable long-term data.

Pielke, Sr., GEN-4: With respect to the assessment of tropospheric temperature trends, the heat storage and fluxes into the atmosphere from the surface are a more robust procedure to explain observed trends over the last several decades.[23] The Report should have addressed the issue as to why the reconciliation of a global- and zonally averaged surface temperature trend with the tropospheric trends is even an important policy issue.

Response: The aim of the Report was to interpret temperature trends in the lower atmosphere vis-à-vis the surface. Heat storage in the ocean is important, and results so far strongly support the reality of anthropogenic effects on climate, but it is not directly related to the surface versus troposphere issue.

Pielke, Sr., GEN-5: Reanalyses. The use of current reanalyses to assess trends was minimized in the Report, and was a recommendation of the NRC Review.[24][25] However, not commented on by the Review was their use to assess trends in regions where the

[23] Pielke Sr., R.A., 2003: Heat storage within the Earth system. Bull. Amer. Meteor. Soc., 84, 331-335.

[24] However, on page 20 lines 405-407 of Chapter 2 of the CCSP Report, the only caution is not to use them when stratospheric information are considered. Thus, while the Executive Summary of the CCSP Report states otherwise, even the CCSP Report, by inference, indicates reanalyses are valuable for long term lower tropospheric temperature trend assessments.

[25] As a contradiction in the Report itself, Chapter 5 refers to a paper by Santer (2003a) where they used reananlyses in their study to assess long term tropopause trends.

magnitude of the trends has been large and for seasonal averages, such that accurate comparisons with satellite and radiosonde observations can be made. This approach has been shown to be robust[26] Chase et al (2000), with text included on this need in the final version of Chapter 6 that I was CLA (Appendix B). The treatment of the current reanalyses as inadequate for long-term temperature trends ignores the value-added by winds in particular in defining the tropospheric layer-averaged temperatures in the mid- and high-latitudes[27]. This is an added source of information with which to quantitatively compare with the other data sets.

Response: ERA-40 is better than NCEP but can only be used for climate analysis after 1979 and then with great caution. See Simmons et al. 2004. Regional analyses are not the subject of this report as now made clearer in the revised Preface which also gives the reasons. It may be true in the future that a climate-quality reanalysis could play a significant role in ironing out small-scale inhomogeneities in the surface temperature observing system which undoubtedly exist.

Pielke, Sr., GEN-6: The reanalyses can, therefore, provide critical information on regional temperature trends. Since weather is determined by the spatial pattern of tropospheric temperatures, rather than a global- or tropical zonally-averaged mean, the reanalyses are particularly well suited for this assessment. Indeed, the 2005 National Research Council report concluded that:

"regional variations in radiative forcing may have important regional and global climate implications that are not resolved by the concept of global mean radiative forcing."

And furthermore:

"Regional diabatic heating can cause atmospheric teleconnections that influence regional climate thousands of kilometers away from the point of forcing."

This regional diabatic heating produces temperature increases or decreases in the layer-averaged regional troposphere. This necessarily alters the regional pressure fields and thus the wind pattern. This pressure and wind pattern then affects the pressure and wind patterns at large distances from the region of the forcing which we refer to as teleconnections. This major issue, which should have been a major focus of the Report, as recommended in the 2004 Asheville Workshop, was inadequately covered in the Report. In Chapter 5, for example, of the seven figures shown, only one presented a spatial map of the trends, and even then no quantitative evaluation of the regional skill of the models in replicating the January 1979 to December 1999 trends is given. In the Executive Summary, only a reference to fingerprint studies is present (referring to Box 5.5.) with a selected summary of previous papers given.

[26] Chase, T.N., R.A. Pielke Sr., J.A. Knaff, T.G.F. Kittel, and J.L. Eastman, 2000: A comparison of regional trends in 1979-1997 depth-averaged tropospheric temperatures. Int. J. Climatology, 20, 503-518.
[27] Pielke, R.A. Sr., T.N. Chase, T.G.F. Kittel, J. Knaff, and J. Eastman, 2001: Analysis of 200 mbar zonal wind for the period 1958-1997. J. Geophys. Res., 106, D21, 27287-27290.

These comparisons should be also performed for seasonal averages and not just annual averages, which is another overlooked assessment in the Report.

To illustrate the value of using the relationship between winds and the temperature field, Figure 5.5 of the CCSP Report could have been used to compute the trends annually averaged east-west wind change that would be expected with the reported tropospheric temperature change. This would have provided an independent evaluation of the temperature trends. Using the thermal wind equation, an annual, zonally-averaged and tropospheric-layer averaged increase of 1 degree Celsius per 1000 km in mid-latitudes would produce a 4.3 meters per second increase of zonally averaged wind speed at 200 hPa. This text was also in Chapter 6, but was deleted in the ad hoc replacement Chapter.

Specific questions for the Committee for this subject area are the following:

Question: What is the magnitude in of the regional tropospheric layer-averaged temperature gradient annual- and season-averaged trends in the middle and higher latitudes as diagnosed from the horizontal winds using the thermal wind relation? How does this analysis compare with the layer-averaged temperature trends as computed with the available radiosonde and satellite data sets?

Question: What is the quantitative skill in degrees Celsius regionally of the temperature annual- and season-averaged trends between the models and the observed tropospheric temperatures from the satellite and radiosonde data, and from reanalyses over the recent decades?

Response: Wind data are of unknown accuracy and would only be useful if the geographical gradients of temperature trend were sufficiently large, and then only in the extra-tropics. Quite apart from the fact that over 1979-2004 the geographical gradients of temperature trends aloft are quite small, there is no published literature that this Assessment could review.

Owing to natural internal variability, models cannot be expected to reproduce regional patterns of trend over a period as short as 20 years from changes of radiative forcings alone.

Pielke, Sr., GEN-7a, Models: Although Chapter 5 contains a very informative summary of the latest global climate model simulations, the survey is incomplete. While the forcings listed in Table 5.2 of Chapter 5 are an improvement over past model studies, they remain a subset of the recognized climate forcings[28]. Moreover, the forcings included even from the Table varied among the modeling groups.

[28] National Research Council, 2005: Radiative forcing of climate change: Expanding the concept and addressing uncertainties. Committee on Radiative Forcing Effects on Climate Change, Climate Research Committee, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, The National Academies Press, Washington, D.C., http://www.nap.edu/openbook/0309095069/html/

One particular serious omission is the lack of description as to what indirect aerosol effects were actually used in the few models that were listed as having this forcing. The indirect aerosol forcings are diverse and significant and include the "first indirect aerosol effect", the "second indirect aerosol effect", the "semidirect effect", the "glaciation effect", the "thermodynamic effect", and "the surface energy budget effect".[29] Table 1 in the Executive Summary is titled "*Summary of the most important global-scale climate forcing factors*", but all of the most important climate forcings as identified by the 2005 National Research Council Report were not listed. This further illustrates the cherry picking of information for this Report.

Response: This report makes use of results from the so-called "20CEN" experiment recently performed in support of the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC FAR). The integrations analyzed in Chapter 5 were performed with 19 different models, and involve modeling groups in nine different countries. As discussed in Chapter 5, these 20CEN runs were performed with "…new model versions, and incorporate historical changes in many (but not all) of the natural and human forcings that are thought to have influenced atmospheric temperatures over the past 50 years" (page 104, column 2, para. 1).

The authors of this Report were in no position to influence the design of the 20CEN experiment. The 20CEN runs analyzed here had been completed, or were in the process of being performed, at the time work on this Report commenced.

Chapter 5 is fair and balanced in its discussion of these new model results. It explicitly notes that individual modeling groups used different sets of external forcings (see Tables 5.2 and 5.3), and that the "…selection and application of natural and anthropogenic forcings was not coordinated across modeling groups" (page 104, column 2, para. 2). It also points out that "In practice, experimental coordination is very difficult across a range of models of varying complexity and sophistication" (page 104, footnote 41).

The Reviewer notes that "While the forcings listed in Table 5.2 of Chapter 5 are an improvement over past model studies, they remain a subset of the recognized climate forcings[30]. Moreover, the forcings included even from the Table varied among the modeling groups."

The first part of this comment implies that we somehow have perfect knowledge of all "recognized" climate forcings, and how they have changed over space and time. This is not the case. As pointed out in some detail in Section 3 (pages 95-97), our level of scientific understanding is quite low for some of these forcings. It is noted that "…we

---

[29] http://www.nap.edu/books/0309095069/html/40.html from National Research Council, 2005: Radiative forcing of climate change: Expanding the concept and addressing uncertainties. Committee on Radiative Forcing Effects on Climate Change, Climate Research Committee, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, The National Academies Press, Washington, D.C.,

[30] National Research Council, 2005: Radiative forcing of climate change: Expanding the concept and addressing uncertainties. Committee on Radiative Forcing Effects on Climate Change, Climate Research Committee, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, The National Academies Press, Washington, D.C., http://www.nap.edu/openbook/0309095069/html/

will never have complete and reliable information on all forcings that are thought to have influenced climate over the late 20th century. A key question is whether those forcings most important for understanding the differential warming problem are reliably represented. This question is currently difficult to answer" (page 96, column 2, para. 3).

As mentioned above, the second part of the Reviewer's comment ("Moreover, the forcings included even from the Table varied among the modeling groups.") is discussed in some detail in our Chapter. The Reviewer's comment suggests that there are universally agreed upon "best" datasets for specifying "recognized climate forcings" such as the spatial and temporal changes in land surface properties over the 20th century, or the burdens of soot aerosols in the atmosphere. In practice, however, there are significant uncertainties in our knowledge of the space-time changes in these and many other external forcings. The fact that different modeling groups have used different datasets for specifying a given forcing is both a weakness and a strength – a weakness because the 20CEN runs convolve uncertainties in climate forcings with uncertainties in the model response to forcings (see Recommendation 1, page 91), and a strength because the 20CEN results span "…a wide range of uncertainty in current estimates of climate forcings" (page 104, column 2, para. 3).

We turn next to the Reviewer's comment that "One particular serious omission is the lack of description as to what indirect aerosol effects were actually used in the few models that were listed as having this forcing." Here, the Reviewer is requesting highly technical information. Our Report is not targeted for a technical audience. For the four U.S. models whose 20CEN results are featured in more detail (see Figures 5.5 and 5.7), we do provide complete information and references on the datasets used for specifying forcings. The technical information requested by the Reviewer is available in those references.

Finally, the Reviewer claims that "Table 1 in the Executive Summary is titled "*Summary of the most important global-scale climate forcing factors"*, but all of the most important climate forcings as identified by the 2005 National Research Council Report were not listed. This further illustrates the cherry picking of information for this Report."

This comment is puzzling. It alleges some there was some intent on our part to selectively filter information provided to the readers of this report. We strongly refute this allegation. The NRC Report mentioned by the Reviewer is cited in Chapter 5, and Tables 5.2 and 5.3 provide details of the natural and anthropogenic forcings that were varied in the 20CEN runs analyzed in here.

Pielke, Sr., GEN-7b, Models: The Preface of the CCSP Report (page 5, lines 102-106) provides clear evidence of the incompleteness of the Report;

"To help answer the questions posed, climate model simulations of temperature change based on time histories of the forcing factors thought to be important, have been compared with observed temperature changes. If the models replicate the observed temperature changes, this increases confidence in our understanding of the observed temperature record and reduces uncertainties about projected changes."

First, forcing factors "thought" to be important are left out of the studies as discussed earlier in this Section. The surface temperature data also has significant uncertainties (as overviewed in Section 3.1) which raises questions about the accuracy of comparing model data. Even more importantly, the statement is silent on the spatial scale of the model-observational comparisons. Thus,

Why should the models be assumed as skillful in hindcasts if important first-order climate forcings are ignored?

Response: The report is using the most up-to-date model versions available. Within Chapter 5, section 3, we explicitly state that we can never be sure to have included all external forcings relevant to the "differential warming" problem (see **Response to Pielke Sr., GEN-7a**). Quite frankly, the Reviewer's position on this issue borders on the ludicrous. If one follows his statements through to their logical conclusion, then we should never undertake assessments of how well models perform in hindcasting 20th century climate change, because we will never have perfect knowledge of historical changes in all forcings that the reviewer deems to be "first-order". We do not subscribe to this extreme position. Our job is to address important climate science questions – questions that are obviously of great relevance to policymakers – with state-of-the-art climate models, and with our current best estimates of historical changes in external climate forcings.
Finally, our Report is not "…is silent on the spatial scale of the model-observational comparisons." The Preface (page V) explicitly notes that much of the motivation for this Report arises from apparent discrepancies between observed surface and tropospheric temperature changes that were manifest at very large spatial scales (averages over the globe and over the tropics). Chapter 5 also addresses the "spatial scale" issue raised by the Reviewer:

"Our primary focus is on the tropics, since previous work by *Gaffen et al*. (2000) and *Hegerl and Wallace* (2002) suggests that this is where any differences between observations and models are most critical. We also discuss comparisons of global-mean changes in atmospheric temperatures and lapse rates. We do not discount the importance of comparing modeled and observed lapse-rate changes at much smaller spatial scales (particularly in view of the incorporation of regional-scale forcing changes in many of the runs analyzed here), but no comprehensive regional-scale comparisons were available for us to assess." (page 105, column 2, para. 1, and page 106, column 1, para. 1).

We note that the paragraph quoted immediately above was inserted in Chapter 5 in response to comments made by the Reviewer before the Reviewer's resignation from the Lead Author team of this Report.

Pielke, Sr., GEN-7c, Models: What are the magnitudes of the uncertainties identified in Section 3.1 of this Public Comment?

Response: The uncertainties in the model results are discussed in detail in Section 5 of Chapter 5, and are quantified in Tables 5.4A and 5.4B (for global-mean and tropical trends in stratospheric, tropospheric, and surface temperatures, and for trends in tropospheric lapse rates). As discussed at length in the **Response to Pielke Sr., GEN-7a**, model uncertainties arise from uncertainties in both the imposed forcings and the climate model responses to these forcings. Structural uncertainties in the observations are quantified in Chapter 3 (see Tables 3.3 and 3.4), and their derivation is discussed in Chapter 4.

Pielke, Sr., GEN-7d, Models: What is the quantitative skill of the model hindcasts on the regional scale for the period January 1979 to December 1999 both in terms of annual and seasonal averages?

Response: See response given above to **Pielke Sr., GEN-7b** comment. Detailed studies of regional hindcast skill were not available for all of the models discussed in Section 5 of Chapter 5, and so could not be provided. However, several of the models presented in Chapter 5 have been subjected to regional-scale assessments of model skill. Such work suggests that at least some current climate models do have skill in simulating observed surface temperature changes over the 20th century (see page 102, column 1, para. 3, and column 2, para. 1).

We note that the "signal-to-noise" (S/N) problem involved in regional-scale model-data comparisons is not mentioned by the Reviewer. This problem is non-trivial. It is discussed in Section 4.4 of Chapter 5 (page 102, column 1, para. 2). The implication of the S/N problem is that even with a hypothetical "perfect" model and complete knowledge of the space-time changes in all important climate forcings, regional-scale evaluation of model skill is still a difficult problem. This is essentially because of the chaotic nature of the climate system.

From our perspective, it is somewhat puzzling that the Reviewer is emphasizing regional-scale evaluation of model "hindcasts". The focus of this Report is on the apparent large-scale discrepancy between observed surface and tropospheric temperature changes (and between modeled and observed tropospheric temperature changes). It is not on model performance in the Amazon Basin, or in Outer Mongolia. It is at these regional scales that models are less skillful, signal-to-noise problems are more serious, and uncertainties in spatially-heterogeneous forcings are likely to be largest.

Bottom line: Although we agree with the Reviewer that regional-scale evaluation of climate models is an important exercise, it was not an exercise central to this Report. The question at the core of our Report relates to a problem manifest at very large spatial scales. The large-scale nature of the discrepancy between observed surface and tropospheric temperature changes (and between modeled and observed tropospheric temperature changes) was what initially attracted the attention of scientists and policymakers.

Pielke, Sr., GEN-7e, Models: This lack of a quantitative evaluation of the skill of the models in replicating the regional trends evident in the satellite, radiosonde, and reanalysis data since 1979 is a serious omission in the Report. The second finding in Chapter 5 that "results from many different fingerprint studies provides consistent evidence for a human influence on the three-dimensional structure of atmospheric temperature over the second half of the 20th century" is not documented by specific comparisons to the regional data from the satellites, radiosondes, and reanalyses. Indeed, this section was expanded from the August 2005 version apparently to give lip service to the need in the report to consider a regional perspective. It is very inadequate and selective in its summary of regional lower atmosphere temperature trends.

Response: See response given above to **Pielke Sr., GEN-7d** comment. We evaluated the limited number of rigorous assessments of "regional hindcast skill" that were available in the published literature (see, *e.g.*, page 102, column 1, para. 3, and column 2, para. 1). We were charged with assessing existing scientific research, and not with performing and publishing new research specifically for the purposes of this Report. We could not assess work that does not exist.

Our brief was to consider a scientific problem manifest at very large spatial scales. It was not to perform new assessments of model "hindcast skill" at regional scales. Such regional assessments are of limited usefulness owing to the large, chaotic variability of the climate system. Because of this variability, models cannot be expected to reproduce observed regional patterns of temperature trends over a period as short as 2-3 decades, even with hypothetical perfect models and complete knowledge of radiative forcing changes.

The Reviewer mentions Key Finding 2 ("Results from many different fingerprint studies provide consistent evidence for a human influence on the three-dimensional structure of atmospheric temperature over the second half of the 20th century"). The Reviewer states that this finding "is not documented by specific comparisons to the regional data from the satellites, radiosondes, and reanalyses". He fails to note that Key Finding 2 is documented by literally dozens of rigorous statistical studies. Details of these studies are provided in Section 4.4 of Chapter 5. The focus of these studies is on comparison of detailed patterns of modeled and observed temperature change, either in terms of global latitude-longitude maps, zonally-averaged profiles through the Earth's atmosphere, *etc*.

The important point here is that Key Finding 2 is supported by compelling scientific evidence. The Reviewer's comments obfuscate this evidence by again reverting to discussion of "regional data".

Bottom line: We do not "…give lip service to the need in the report to consider a regional perspective". We discuss existing and relevant published assessments of how well models perform in simulating regional aspects of observed temperature changes. Such assessments are currently limited in number and in scope. Our Report is not about regional climate change – it is about a very specific problem manifest at large spatial scales. The Reviewer's interpretation of our scientific charge is quite different from our own interpretation of that charge.

Pielke, Sr., GEN-7f, Models: The International Geosphere-Biosphere Programme (IGBP) report entitled "Vegetation, water, humans and the climate: A new perspective on an interactive system"[31] provides extensive documentation of significant and obvious fingerprints of a human climate forcing (in this case land use/land cover change and variability). The authors of Chapter 5 discuss fingerprint studies in Box 5.5, but fail to include the spectrum of papers on this subject that are outside their expertise, yet were made aware of during the course of the Report preparation.

Response: The Reviewer's definition and understanding of "fingerprinting" and "detection and attribution" is not the same as that discussed in Chapter 5, or in the literature in general. Our focus is on rigorous statistical comparisons of modeled and observed temperature changes. Such work explicitly considers whether the climate "signal" in response to an imposed forcing change (such as a change in land surface properties) is statistically identifiable relative to the "noise" of natural climate variability. We have included all formal detection and attribution studies that are germane to evaluating the causes of surface and free atmosphere temperature changes.

The studies referred to by the Reviewer are largely qualitative in nature. Typically, they do not involve any attempt to assess the formal statistical significance of results. Discussions with the Reviewer (prior to the Reviewer's resignation as a Lead Author of this Report) prompted us to include some discussion of this more qualitative work in Chapter 5 (see, *e.g.*, Boxes 5.3 and 5.4 on pages 96 and 97). From our perspective, however, rigorous fingerprint studies are much more useful for investigating the causes of recent temperature changes.

Pielke, Sr., GEN-7g, Models: The 8th Finding in Chapter 5 also is disingenuous. The statement that changes "in black aerosols and land use/land cover (LULC) may have had significant influences on regional temperature, but these influences have not been quantified in formal fingerprint studies" is incorrect. The role of these forcings is so categorical that fingerprint studies are not required.[32]

---

[31] Kabat, P., Claussen, M., Dirmeyer, P.A., J.H.C. Gash, L. Bravo de Guenni, M. Meybeck, R.A. Pielke Sr., C.J. Vorosmarty, R.W.A. Hutjes, and S. Lutkemeier, Editors, 2004: Vegetation, water, humans and the climate: A new perspective on an interactive system. Springer, Berlin, Global Change - The IGBP Series, 566 pp.

[32] See the summary of these forcings in National Research Council, 2005: Radiative forcing of climate change: Expanding the concept and addressing uncertainties. Committee on Radiative Forcing Effects on Climate Change, Climate Research Committee, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, The National Academies Press, Washington, D.C., and Kabat, P., Claussen, M., Dirmeyer, P.A., J.H.C. Gash, L. Bravo de Guenni, M. Meybeck, R.A. Pielke Sr., C.J. Vorosmarty, R.W.A. Hutjes, and S. Lutkemeier, Editors, 2004: Vegetation, water,

Response: The Reviewer is mistaken. Key Finding 8 is scientifically accurate. At the time this Report was written, the specific influences of carbonaceous aerosols and land use/land cover changes had not been "quantified in formal fingerprint studies". The Reviewer cannot simply assert that "The role of these forcings is so categorical that fingerprint studies are not required". Rigorous fingerprint studies are an essential part of investigating cause-and-effect relationships in the climate system. We cannot quantify the magnitude of LULC effects on global-scale lapse-rate changes simply by eyeballing the differences between modeled and observed temperature fields that are complex space-time vectors!

LULC forcing may indeed cause large temperature changes at local and regional scales (see Box 5.4). However, climate noise typically increases with decreasing spatial scale. Thus a large local or regional climate change does not necessarily translate to a statistically significant change. Again, this is why need rigorous statistical assessments of S/N properties.

If the effect of LULC changes is really as "categorical" as the Reviewer claims, and if this effect is evident at the largest spatial scales, then it should be an easy task for the Reviewer to use well-documented fingerprint methods to quantify the magnitude of LULC effects on the vertical structure of atmospheric temperature changes. The Reviewer has not done so. Nor have other investigators applied standard fingerprinting methods to the climate signals arising from changes in carbonaceous aerosols and LULC. We recommend that such investigations should be performed with newer climate model runs that now include these forcings (see Recommendations 2 and 4 in Chapter 5).

---

humans and the climate: A new perspective on an interactive system. Springer, Berlin, Global Change - The IGBP Series, 566 pp. The Reviewer's definition and understanding of "fingerprinting" and "detection and attribution" is not the same as that discussed in Chapter 5, or in the literature in general. Our focus is on rigorous statistical comparisons of modeled and observed temperature changes. Such work explicitly considers whether the climate "signal" in response to an imposed forcing change (such as a change in land surface properties) is statistically identifiable relative to the "noise" of natural climate variability. We have included all formal detection and attribution studies that are germane to evaluating the causes of surface and free atmosphere temperature changes.

The studies referred to by the Reviewer are largely qualitative in nature. Typically, they do not involve any attempt to assess the formal statistical significance of results. Discussions with the Reviewer (prior to the Reviewer's resignation as a Lead Author of this Report) prompted us to include some discussion of this more qualitative work in Chapter 5 (see, *e.g.*, Boxes 5.3 and 5.4 on pages 96 and 97). From our perspective, however, rigorous fingerprint studies are much more useful for investigating the causes of recent temperature changes.

Finally, we note that Key Finding 8 (the Finding cited by the Reviewer) relates to the question of whether recent forcing by carbonaceous aerosols and LULC changes has had a significant effect on lapse rates at the large space scales that are of primary interest to the report. The preliminary answer to this question is "no". This does not mean that these forcings will prove unimportant at smaller spatial scales, as is made abundantly clear in the text of Chapter 5, and in the two "bullets" of Key Finding 8 (page 91).

Pielke, Sr., GEN-7h, Models: In the Executive Summary regarding the models (page 5, lines 100-107), the authors make an astounding claim,

"On decadal and longer time scales, however, while almost all of the model simulations show greater warming aloft, most observations show greater warming at the surface. These results have at least two possible explanations, which are not mutually exclusive. Either the amplification effects on short and long time scales are controlled by different physical mechanisms, and models fail to capture such behavior; and/or remaining errors in some of the tropospheric data sets adversely affect their long-term temperature trends. The second explanation is judged more likely."

Thus despite the caution of the NRC review of the Report earlier this year

"It should also be clearly emphasized that data is being used to test models and not vice-versa"[33],

the authors ignore this caution by the NRC Committee. They accept the model results (which is a hypothesis) as truth and blame the data when it does not agree. And not any data, but just the data that does not conform to their prejudices (i.e., the surface temperature data in the tropics is assumed robust, which as overviewed in Section 3.1 of this Report still contains unquantified uncertainties).

Response: This Key Finding (which has now been slightly modified) is not an "astounding claim". It is merely a statement of the results of Chapter 5. The revised text (on page 90, Key Finding 6, bullet 5) now reads:

"These results could arise due to errors common to all models; to significant non-climatic influences remaining within some or all of the observational data sets leading to biased long-term trend estimates; or a combination of these factors. The new evidence in this Report (model-to-model consistency of amplification results, the large uncertainties in observed tropospheric temperature trends, and the independent physical evidence supporting substantial tropospheric warming) favors the second explanation".

Instead of "favors the second explanation", the public review version stated that the second explanation was "more likely". Use of the new phrase "favors the second expression" is a simple, factual description of the majority opinion of the Lead Authors

[33] http://www.nap.edu/books/030909674X/html/39.html

of this Report, and does not express any value judgment regarding the relative likelihood of the two posited explanations (see response to **Douglass CH5-1**).

Despite the Reviewer's strident claims to the contrary, we are cautious and circumspect in our interpretation of model-data comparisons. In Chapter 5, we explicitly state that:

"As pointed out by *Santer et al*. (2003b) and *Christy and Spencer* (2003), we cannot use such model-data comparisons alone to determine whether the UAH or RSS T2LT data set is closer to (an unknown) "reality". As the next section will show, however, models and basic theory can be used to identify aspects of observational behavior that require further investigation, and may help to constrain observational uncertainty" (page 112, column 2, para. 3).

We do not "accept model results (which is a hypothesis) (*sic*) as truth and blame the data when it does not agree". We point out that: 1) Models, data and basic theory all show consistent behavior in terms of how the month-to-month and year-to-year changes in tropical surface temperatures are amplified in the free troposphere; 2) This consistency breaks down – at least for some observational datasets – when one considers temperature changes on decade-to-decade timescales; 3) Chapters 3 and 4 have shown that the basic structural uncertainty in the observations is much larger than was hitherto believed and can easily span the model results (see responses to **Douglass CH5-1 and CH5-7**); and 4) There is other complementary evidence (such an increase in tropospheric water vapor) that provides independent physical support for recent tropospheric warming).

What the Reviewer fails to mention is that one of the datasets used in this Report (the UAH T_{2LT} data) initially showed cooling of the tropical lower troposphere over the satellite era. This cooling was in stark contradiction to all model results, to basic theory, and to our understanding of the physics of the tropical atmosphere. In the process of work on this Report, *Mears and Wentz* (2005) identified an error in the procedure used by the UAH group to adjust for drift in sampling the diurnal temperature cycle. Correcting the error changed the sign of the UAH tropical T_{2LT} trend. A cooling trend became a warming trend. Clearly, knowledge of theoretical and model-based "amplification factors" was helpful in trying to understand and interpret the anomalous UAH T_{2LT} result. Sometimes it is useful to confront observational data with basic theory and with model results, particularly when the structural uncertainties in observations are very large. The Reviewer's perspective – that models and theory are never useful for discriminating between wildly differing observational datasets – does not seem sensible to us.

Bottom line: The new UAH T_{2LT} results still show tropospheric damping of decadal-timescale surface temperature changes. This result implies that in the real world, different physical mechanisms govern "amplification behavior" on short and on long timescales. The Reviewer provides no indication of how or why the basic physics might vary with timescale. No have any other Reviewers of this Report.
Alternately, the RSS T_{2LT} result, and the latest analyses of radiosonde data by *Sherwood et al*. (2005) and *Randel and Wu* (2006) imply amplification behavior that is consistent with models and theory across a range of different timescales, and consistent with

independent physical evidence of recent tropospheric warming. Occam's Razor suggests that the simpler and internally-consistent explanation is preferable to the more complex "different (but unknown) physics" explanation.

Pielke, Sr., GEN-7i, Models: Specific questions to ask the Committee include:

What is the uncertainty in the estimates of the zonal and global averaged tropospheric temperature trends on annual and seasonal averages due to the neglect of all of the first-order climate forcings? Achieving correspondence with the observations when a subset of recognized first-order climate forcings are neglected is not a demonstration of skill.

Response: Many of the 20CEN runs analyzed in Chapter 5 incorporate a broad range of natural and anthropogenic forcings. The Reviewer's claim of "neglect of all of the first-order climate forcings" is demonstrably untrue, as is readily apparent from examination of Tables 5.2 and 5.3 in Chapter 5. See **Response to Pielke Sr., GEN7a**.

Without systematic experimentation – which is exactly what we advocate in Recommendation 1 of Chapter 5 – we have no way of separating forcing uncertainties from climate response uncertainties. The Reviewer's question simply cannot be answered at present, and is unlikely to be answerable in the foreseeable future. This does mean, however, that we should all go home and do no model experimentation until we somehow obtain perfect knowledge of all historical changes in "first-order" climate forcings. Useful science can be done with the existing 20CEN runs analyzed in Chapter 5. In fact, one of the interesting results emerging from Chapter 5 is that inter-model forcing differences seem to have surprisingly little impact on simulated amplification behavior, at least at very large spatial scales.

Pielke, Sr., GEN-7j, Models: What is the quantitative uncertainty in the model hindcasts of regional tropospheric temperatures in terms of annual and seasonal averages?

Response: See Response to Pielke Sr., GEN-7a,b,c,d,e. Regional analyses were not part of our direct mandate.

Pielke, Sr., GEN-3k, Models: What added information on regional surface and tropospheric temperature trends are provided from regional climate models?

Response: See Response to Pielke Sr., GEN-7a,b,c,d,e. Regional analyses were not part of our direct mandate.

Swanson GEN-1 Comment: In October 2003, a report which I wrote was published in the GRL (Swanson, 2003). As a result, I was invited to attend the RVTT workshop in Ashville, NC, which began the process that produced this Draft Report. The paper had just been published a few days before the Workshop, so I provided copies at the workshop. Most of my comments are derived from the findings presented in this report and an unpublished follow on paper.

I can find no reference in the Draft to the unexpected annual cycle I found in the UAH T_{2LT} data (Swanson, 2003), even though there was brief mention of this problem in the earlier Draft for Peer Review. Since writing my report, I have found that the UAH T data does not exhibit this anomalous annual cycle (See Figures 1 & 2 below). I suggest this fact lends support to my 2LT contention that the UAH T data is impacted by strong influence from the surface at high southern latitudes. Mears and Wentz (2005) in their latest results, point to high altitude effects as a reason to exclude Antarctic data for latitudes above 70S. In my report, I suggested that one explanation for the anomalous annual cycle was the impact of the sea-ice cycle, since at high latitudes the ground path of the scan swaths becomes mostly north-south and as a result, the scans include the large annual cycle in sea-ice around Antarctica.

Given the choice of Mears and Wentz to exclude data for these high southern latitudes, I strongly recommend that the UAH team also exclude these latitudes from their data sets. A similar exclusion of data for the Arctic should also be considered, as well as the possible extension of the exclusion to latitudes greater than 60 degrees. This recommendation presents an unfortunate situation, as the UAH satellite data provides the only wide area coverage of the Antarctic. I suggest, however, that it would better to remove this data until the difficulties I suggest are resolved. Otherwise, the data may be misused by others who are unaware of a possible problem.

Response: The effect of sea-ice anomalies on MSU-based temperature anomalies will be much smaller than the total effect of sea ice on actual MSU-based temperatures because sea-ice anomalies are much smaller than the annual cycle of sea ice. More importantly, the polar regions are not the primary focus of this report. Observations in this region from surface and upper-air instruments are poorest. Concentrating upon sea-ice effects does not help in our resolving the long-standing tropical discrepancy, as there is no sea-ice in the tropics.

Trenberth GEN-1 Comment: There is, in my view, too much emphasis on linear trends and nowhere a clear statement that linear trends are not a good fit to the data (the Appendix in fact claims otherwise but gives examples chosen to make this so). This is especially so in the stratosphere with the volcanic perturbations, in the tropics with ENSO, and it is also true especially for longer intervals such as 1958 to 2004 where the trends in troposphere and stratosphere are very different after 1976 from those before then. As a result, sampling issues and sensitivity to small differences at start and end of series is real. It makes a big difference whether the trends begin in 1976 or 1979. This becomes a major issue for comparisons with model results that do not have such a shift or ENSOs in the right sequence and magnitude. Error bars are missing in many places, including 2 figures in exec summary.

Response: Linear trends are used as a summary statistic. The justification for this and the possible shortcomings are discussed in Chapter 3, pages 29-30, lines 645-652 and footnote 12, as well as in the Appendix. We make note of any important nonlinear changes both in the chapter text as well as in the key findings. For example, the climate regime shift in the troposphere and the possible nonlinearities in the stratosphere due to volcanic eruptions are discussed in sections 3.3.1 and 3.3.2. For model comparisons the 1976 vs. 1979 start date is not an issue since all of our model comparisons involve the satellite era (i.e., a start date of 1979). Utilizing a relatively large ensemble of model simulations allows us to quantify the effect of variability due to ENSO (as well as other internally driven modes of variability) via the spread of results from models. While there is only a single realization for the observations, as we state in the Appendix regarding the use of alternate methods to estimate trends that are less sensitive to the choice of endpoints "… for the data used in this report tests using different trend estimators give results that are virtually the same as those based on standard least-squares …". Our philosophy for the display or non-display of error estimates is discussed at length in the Appendix. (NOTE: See also the response to Trenberth ES-1)

Trenberth GEN-2 Comment: The summary is also deficient on issues of land vs. ocean. This is related to max vs. min changes and how those would be seen in the troposphere vs. surface; i.e., expect max. to be seen from deeper mixing but not min. Surface changes are much larger over land than ocean and muted in troposphere (see chapter 1), but in troposphere changes are more zonally symmetric and larger over oceans than at surface. This relates to the issue of where and how the surface can increase more than troposphere. Chapter 1 makes the point that there are really not good reasons why these should be strongly linked, yet much of the report misses this point. In chapter 4, where huge differences occur over Africa in T2LT, it does not come to grips with this issue (note also that the diurnal cycle of surface temperature is order 30ºC over the Sahara).

Response: A figure (Fig 4.5) has been added in Chapter 4 that shows the difference between the two T_{2LT} datasets and the surface, and text that points out how the differences in diurnal adjustment method may impact these difference maps has also been included.

Trenberth GEN-3 Comment: There is little discussion of issues on urban heat island effects etc. It is briefly mentioned in chapter 4 but inadequate. It is a complex issue and the effects are real, so it while one can say that the global mean is OK because it is not contaminated by unrepresentative very local UHI effects, those changes are real. This is not dealt with in the report. There is now quite a bit of literature related to the "weekend effect" whereby statistics differ by weekday and presumably relate to aerosols and interactions with clouds.

Response: These effects are real locally but not important on the large scales being considered in this report, e.g. Parker (2004).

Trenberth GEN-4, This is supposed to be an assessment. It falls short especially in chapters 2 and 3, where it should refer ahead to chapter 4. In chapter 4 there is some useful assessment but it falls back on "all datasets are equal" in spite of strong evidence otherwise. This is a major limitation of the report.

Response: A Table has been inserted in the Preface to guide readers and reduce cross-referencing and duplication. The report is structured as such that data shortcomings are discussed in Chapter 4.

Trenberth GEN-5 Comment: The report pretends that the radiosondes are global, and insufficient accounting is made of the fact that they are not close to that. Zonal means are also biased by land distribution. Errors of 0.2ºC can occur in global means from the distribution of sondes (Hurrell et al 2000) although effects on trends seems to be modest (0.03ºC decade-1) this is not guaranteed.

Response This is discussed in Chapter 2. It should also be noted that the report is more concerned with long-term stability than the inter-monthly error.

Trenberth GEN-6 Comment: Very little account is taken of the works that show major shortcomings in the radiosondes (Sherwood et al 2005, Randel and Wu 2005) in chapters 2 and 3. They are discussed in chapter 4 and conclusions drawn that sondes are biased cold but then this is ignored elsewhere. There is no sound basis for believing the profiles in Fig 3.7, for instance.

Response: The report is structured as such that data shortcomings are discussed in Chapter 4. Chapter 4 contains a discussion of problems with the radiosonde datasets, including those presented by Sherwood et al. (2005) and Randel and Wu (2005). The Preface has been modified to make clearer the structure of the report with regards to the purpose of each chapter. The purpose of Chapter 3 is to present the observations taken at face value, so it is therefore appropriate to present all of the observed data (such as in Fig. 3.7).

Trenberth GEN-7 Comment: The UAH record has once again been revised but the new T2LT values are at odds with surface temperature trends. Chapter 4 falls short in not presenting maps of this difference. Accordingly, this dataset ought to also be discounted. Given the UAH algorithm that is designed to minimize trends, this dataset ought to be given lower weight, but no commentary appears on this issue.

Response: See response to Trenberth GEN-2 above. While the UAH diurnal adjustment method may cause regional problems, such as the one over Africa, their method should not cause problems when averaging over latitude bands, and this is not a basis for "discounting" this dataset for global or zonal averages. Note that the author team did not think, on the basis of published or "in press" research that is was possible to assign relative credibility levels to individual data sets.

Trenberth GEN-8 Comment: The reanalyses are not considered seriously for no good reason other than opinions that are baseless. For NCEP, these fears are well grounded and some references are given but for ERA-40, major efforts went into bias correction and a major advantage of ERA-40 is that all observations were assimilated at the exact time they were made, overcoming diurnal cycle issues, a major advantage relative to all the other datasets. The bias corrections to the sondes in ERA-40 likely makes them better than the sonde records themselves. Nevertheless the reanalyses are seriously flawed and have to be used with care (see Trenberth and Smith 2005; given below under chapter 1).

Response: ERA-40 is used for climate analysis, but it is recommended that its use should be limited to the period after 1979 and then with great caution. See Simmons et al. 2004.

Trenberth GEN-9, In places the document is unduly dumbed down to the point where the text is not factual. Why is it necessary to have an appendix that is dominated by basic statistical text book material?

Response: Appendix A was added to the report largely in response to Major NRC Review Comment 3b, which stated: "A more thorough discussion of the detailed statistical trend calculations for the various data sets is needed. This discussion might be appropriately placed within an appendix." It is included in recognition of the fact that the report is intended to be understandable to readers who have no formal training in the use of statistical techniques.

Trenberth GEN-10, What is the vintage of this report? It mostly does not include papers submitted or in press but there are exceptions? It would help to make clear the time frame and cut off for considering literature.

Response: The cut off time was when the report was submitted for Public Comments. The Author Team must have had a copy of the article or paper made available to them.

Trenberth GEN-11, The report is very long, not generally readable as a result, and contains a lot (far too much) basic tutorial material.

Response: The report is intended for a wide-range of experts and policy-makers with various backgrounds and varying degrees of previous knowledge on the topic. The Executive Summary and the Abstract, in particular, are written at a very general level and attempts to reach a broad range of individuals with little or no climate science background. The main body of the text is written to be accessible to trained scientists from all disciplines, so has more tutorial information than would be the case if the audience were merely other climate scientists.