

Response of the authors for S&A Product 1.1 to the NRC review of the draft report

The NRC review is available at:
<http://www.nap.edu/catalog/11285.html>

Responses are keyed to comment numbers used in the NRC Report with specific Review Panel wording in *red italics*, followed by the response.

PREFACE

Major Comments

1. *The Temperature Trends report should include an improved discussion of the motivation for this report, which will increase the report's effectiveness for a variety of audiences. The committee also suggests more explicit clarification of the context and intended audience for this report. This background should occur in the preface or in an introduction and address the specific scientific issues that motivated the work (which surprisingly go unmentioned in the draft), what has been done previously, and what are the key outstanding issues. This section should be accessible to general educated readers and also scientifically sound.*

Response: We have strengthened the motivation for this report by substantially revising the Preface. There is now a much longer discussion of background, new data sets, and analysis since the IPCC (2001) and the NRC (2000) report which addressed this topic previously. We have added a lexicon of terms to help our readers understand areas of scientific agreement and areas where uncertainty remains, even among our teams of experts.

2. *To help the report communicate effectively to an array of audiences, the committee suggests changes to the presentation style within each chapter. The key findings for each chapter, should be brought to the front of that chapter possibly in the form of bulleted highlights, with a one-sentence summary and brief discussion for each key point (similar to the format within Intergovernmental Panel on Climate Change and World Meteorological Organization Assessments). The key points should be based on the detailed discussions within each chapter. The chapters should in some cases include more scientific rigor (as detailed in the chapter reviews that follow) and be aimed at the broad climate science community. One possible strategy for including scientific rigor in the chapters is to use footnotes to describe technical details, as was done in Chapter 5.*

Response: We have incorporated the suggestion to improve clarity by providing Key Findings and Recommendations at the beginning of each of Chapters 2-5. Chapter 6 and the Executive Summary build on these findings and recommendations. We have also increased the use of footnotes for scientific details and we have added an Appendix to discuss the nuances of statistical significance testing and uncertainty error bars for the data presented in this report.

Executive Summary

Note that the Executive Summary has been completely rewritten in a style recommended by the Review Panel, ensuring much better traceability (see response to question #8 below). The text of the Executive Summary has also been substantially shortened, eliminating much of the text that was criticized by the Review Panel. Similarly, the number of Figures has been reduced from nine to four, and only one of the original Figures (original Figure 3, now Figure 1) remains. Criticisms of many of the original Figures are no longer relevant.

Major Comments

- 3. More explicit discussion of the statistical characterization of uncertainty in trends is needed This discussion might be appropriately placed within an Appendix. When comparing trend differences between two estimates of the same quantity (e.g., tropospheric temperatures from radiosondes and satellites), it is more appropriate to examine the trend of the difference time series*

Response: A detailed Statistical Appendix has been written explaining the various statistical methods used in the Report, explaining why statistical trend uncertainties are not always given, describing methods for comparing time series, and giving specific examples.

- 4. a more critical evaluation of the trend differences between the University of Alabama, Huntsville (UAH) and Remote Sensing Systems (RSS) satellite microwave data sets (is needed) as well as a discussion of the implication of the differences.*

Response: These issues are discussed in detail in the main text and in the new Statistical Appendix.

- 8. Changes in the Executive Summary are needed to make key results more accessible and ensure traceability to the results in Chapters 1-6. A possible strategy is to bring forward key bullet points from each chapter followed by brief explanatory text, key figures, and implications for understanding within each chapter The first page of the Executive Summary should concisely summarize the key results of the report in a short abstract.*

Response: The Executive Summary (ES) has been completely rewritten. In particular, the recommendations stated here have been followed.

Review of the Executive Summary

The Executive Summary (needs to be) a crisp summary of the key major conclusions, accomplishments, and future challenges (and) ensure traceability to the results developed in Chapters 1-6.

The revised version takes account of these points.

Major Comments

- 1. Repeats #8 above.*
- 2. The first two paragraphs*

Response: These paragraphs have been removed.

- 3. It appears that some subjectivity is necessary in making the optimistic-sounding statements in the first two paragraphs the data can neither reject nor confirm the hypothesis that the models are in some sense reliable.*

Response: Research published since the draft Report was produced has provided justification for “the optimistic-sounding statements in the first two paragraphs”. The data do show that “the models are in some sense reliable”. These points are now made in individual chapters, and the points are now repeated verbatim in the ES.

- 3. The choice to present the results without any statistical significance testing or confidence intervals is highly questionable and ordinarily not allowed in the scientific literature.*

Response: This is not of direct concern to the ES. However, the issue is now discussed at length in the new Statistical Appendix.

- 4. additional support for the conclusions in the first two paragraphs (is needed).*

Response: These paragraphs have been removed.

- 5. The traceability of the conclusions in the Executive Summary is not entirely clear A way to clarify might be to develop major conclusions in the chapters and then move these statements forward to the Executive Summary.*

Response: This recommendation has been followed.

Specific Comments

- 1. The Executive Summary should start with a statement about why the reader should care about the subject of the report.*

Response: Such a statement appears in the Preface.

- 2. The first two paragraphs appear to be the abstract of the document and contain the main takeaway messages Does consistency merely mean that the observations and the model results have overlapping uncertainties?*

Response: These paragraphs have been removed. A brief Abstract is now given (albeit not labeled specifically as an “Abstract”). The word “consistency” is still used elsewhere in the ES, with the meaning clearly stated.

- 3. In lines 71-73, the conclusion that the report increases confidence in our understanding of recent climate change seems optimistic and inconsistent with the supporting evidence.*

Response: Research published since the draft Report was produced has provided justification for the claim of increasing confidence. These reasons are now explained in individual chapters, and the points are repeated verbatim in the ES.

- 4. Figure 1 is not effective.*

Response: This Figure has been removed.

- 5. Figure 2 is the basis for the major new conclusions summarized in the first two paragraphs of the (ES); however, it is not developed in detail and appears only in the Executive Summary and not in the supporting chapters suggesting a lack of traceability Figure 2 deals only with the 20-year period from 1979-1999.*

Response: Figure 2 has been replaced by new Figures (3 and 4) in the revised ES. Information is now given not only for global means but also for averages over 20S to 20N. These Figures still only cover 1979-99. However, comparisons with model results from 1958 onwards have not been made, so the 1979 start point is necessary. Where the material comes from in earlier chapters is now stated specifically. Information for 1958-2004, and its source, is now given in new Figures 1 and 2.

6. *This comment concerned Figures 2-5, and 7. A specific comment is “A rather strange nomenclature is developed”.*

Response: For Figure 2, see above. Figure 3, slightly modified, is still included, with the source of the material stated. The other Figures have been removed. The nomenclature (which, of course, comes from other chapters), has been revised.

7. *An important statement in the (ES) is “the climate models simulate greater warming in the troposphere than at the surface which is not apparent in the observations” (lines 67-68). Whether this statement is true depends on whether the results by Fu et al. are correct the Fu et al. results should be discussed in the report chapters and then distilled in the Executive Summary. Failure to account for the stratospheric contribution in the comparisons between data and models may compromise the report’s conclusions. This should be addressed in Chapter 5 and also summarized in the Executive Summary.*

Response: The issue of the work by Fu et al. is a problem for earlier chapters. Revised versions of these chapters have discussed this work in more detail and in a more balanced way. Individual chapter conclusions that are based in part on this work have been transferred verbatim to the ES.

The claim that model results reported in Chapter 5, or in the ES, might have failed to take account of stratospheric cooling effects on, for example, MSU channel 2 temperatures is not an issue here because we compare model simulated temperature with identical layers based on observations.

8. *In Figure 4, why were the negative weights for TLow-Trop truncated off the plot at the left edge?*

Response: Figure 4 has been removed.

9. *Figure 5 should include error bars.*

Response: Figure 5 has been removed. In any event, there were good reasons for not including error bars in this Figure. These reasons are now explained in the Statistical Appendix.

10. *lines 123-153. The section on “Motivation for this Report” suggests that the main purpose of the report is to address the single issue of surface versus tropospheric temperature trends over the past 20 years, yet the six questions that were to be addressed and the main body of the Report seem to be a somewhat more general approach to the question of temperature trends. Is the Temperature Trends Report intended to be a summary and extension of the 2000 NRC Report or a more general statement of knowledge about temperature trends?*

Response: The motivation for the Report is now stated in the Preface. The questions raised above are addressed there.

11. *In lines 163-169, it would be helpful to specify exactly which are the new data sets that have lead to new interpretations What specifically happened since NRC (2000) and IPCC (2001) and how has this changed perceptions?*

Response: In response to this, observed data details are given in Chapter 3, Table 3.1. Model information is more difficult because most of the newer models used in Chapter 5 do not yet have full published documentation. The source for model information (PCMDI, LLNL) is flagged in Chapter 5. Regarding what is new since 2000/2001, the text has been clarified to identify new results.

12. *In lines 353-359, it seems misleading to use the phrase: "at any one level". The comparison has not been made for levels, but for very deep layers as specified by the MSU channels.*

Response: This text has been deleted, and related text clarified.

13. *In Section 3.2, "Radiosonde data", no specific information is given about uncertainties in the two radiosonde datasets.*

Response: This Section has been removed and replaced by verbatim information from the relevant chapter.

14. *Does the statement on lines 426-427 include an assessment of the sampling uncertainty and a statistical confidence level?*

Response: This statement has been removed.

15. *lines 557-579 (should) discuss succinctly and in common language the meaning of Figures 8 and 9. The discussion of fingerprinting is not really necessary.*

Response: Figure 8 and 9 have been removed. The discussion on fingerprinting has been removed and replaced with a reference to Box 5.5 in Chapter 5.

16. *In line 586 the statement that not including the indirect effect of aerosols is the most important deficiency of the global model simulations should be better justified.*

Response: The confusion here is between the simulations *per se* (where both model and forcing deficiencies are concatenated) and forcing deficiencies alone. The text was meant to apply only to the latter. In any event, this text has been deleted.

17. *Because Figure 6 is repeated in the top half of Figure 8, perhaps Figure 6 can be removed.*

Response: Both Figure 6 and Figure 8 have been removed.

18. *In lines 620-626 the statistical significance is very important, so it would be helpful to highlight the statistically significant areas in Figures 8 and 9, if they are known. If they are not known, then that is important too. Throughout the Executive Summary, the reader is invited to take the values presented literally, even though they may have very large statistical uncertainty in addition to residual structural uncertainty.*

Response: Both Figure 6 and Figure 8 have been removed. The issue of statistical significance is now discussed in a separate Statistical Appendix.

19. *In Figure 9, it is somewhat of a misnomer to call this "Mid-Tropospheric Temperature". It contains a significant contribution from stratospheric trends that increases in magnitude*

with latitude. Much of the apparent agreement in this Figure is simply the result of negative trends in the stratosphere and the increasing fraction of the stratosphere that the weighting function samples as one moves toward the poles.

Response: Figure 9 has been removed.

20. In lines 620-626, are the only trends that are statistically significant the ones above 100 mb, and are they believed to be spurious trends associated with the radiosonde instrument? If so, how are the conclusions about different trends in the surface and troposphere supported? It seems reasonable to suppose that the spurious negative trend in the radiosonde data extends below 100 mb, albeit with decreased magnitude.

Response: This comment refers to Figures 8 and 9. Both Figures have been removed.

21. In lines 639-640, given the uncertainties in both the direct aerosol forcing and the indirect aerosol effects, the report should provide better justification for the conclusion that the aerosol effects have almost certainly been underestimated.

Response: Although this statement can easily be justified, it has been removed.

22. Are the statements in lines 646-652 true for the radiosonde era, or just for the satellite era?

Response: This text has been removed. However, new text dealing with this issue makes it clear whether results apply to the radiosonde era (post 1958) or just to the satellite era (post 1979).

23. Section 6 "Improving our understanding" starts with rather general philosophical statements, whose connection to the recommendations that follow is unclear. The recommendations are broad and unspecific, despite the fact that the Report raises some very specific problems. Some of the recommendations are not argued clearly elsewhere in the Report, in particular Chapter 6.

Response: Part of the reason for the statements in this Section was my perception that Chapter 6 was deficient (in accord with the Review Panel's perception), and this was an attempt to remedy some of those deficiencies. Chapter 6 has been completely rewritten, and the corresponding text in the Executive Summary completely revised in a way that ensures traceability to Chapter 6.

24. It would be useful to have a summary table showing all datasets and models used in this report to avoid long figure captions.

Response: See response to question #11. There are now only four Figures. Three of these still have quite long captions, but this is judged necessary in order to keep the Figures (with captions) self-contained.

CHAPTER 1

Major Comments

1. *The explanation of the greenhouse effect should more clearly describe its effect on the atmospheric temperature structure. In particular, the chapter should explain how the addition of infrared absorbing gases causes the characteristic emission level to be at a higher altitude, where temperatures are colder and where the reestablishment of radiative balance with space calls for warming at this level and communication of this warming to the surface (Goody and Yung, 1989; Lindzen and Emanuel, 2002). Thus, for example, the absence of any warming within the troposphere might suggest that the greenhouse effect is not responsible for the surface warming. A related topic concerns the question of whether temperature changes originating at the surface necessarily lead to temperature changes within the troposphere.*

Response: The chapter now contains a more thorough discussion of the greenhouse effect on atmospheric temperature structure as suggested by the reviewers, including details regarding the relationship between surface temperature changes and possible temperature changes within the troposphere.

2. *Similarly, a discussion of the relation between cumulus convection and the moist adiabat provides an opportunity to use such differential trends to understand the coupling between the surface and the lifting condensation level. Indeed, in the tropics, the temperature structure consists of a surface mixed layer (up to about 500 m) and a trade wind boundary layer (up to about 2 km) above which is the free troposphere. Each of the boundary layers is topped by an inversion which tends to isolate the layer from the region above (Sarachik, 1985). Outside the tropics, the surface communicates with upper levels primarily by quasi-horizontal motions along isentropic surfaces (e.g., Hoskins, 2003). Consequently, the report and the scientific community should move beyond the naïve notion that the lapse rate is a rigid constraint operating from the surface to the tropopause. Instead the observations this report is concerned with should be exploited in order to answer important questions about climate. This objective provides meaningful motivation for ascertaining the accuracy of the temperature measurements and the resulting time series. That said, it should be emphasized that the temperature changes being considered are changes on the order of tenths of a degree (although local changes may be much greater), and current theories may prove inadequate for such small changes.*

Response: We have added a new figure depicting latitude-height depictions of DJF and JJA climatological temperatures, with an expanded discussion of the physical processes whereby the temperature structure in the vertical may vary, including more details regarding how regional processes can contribute to zonal or global averages that may be several tenths of a degree but physically consistent with the processes that regulate vertical temperature profiles.

3. *In general, spatial and temporal sampling is not adequately dealt with in the Temperature Trends CCSP report. Given the fact that horizontal temperature variability at the surface tends to get smoothed as one rises to the free troposphere, there may be serious issues of sampling. Horizontal smoothing over large scales occurs above the boundary layer, but that at the surface and within the boundary layer, there can be much more horizontal variation of temperature. Thus, much more data may be required at the surface to get characteristic temperatures.*

Response: We have added more text regarding the effects of spatial and temporal sampling on quantifying changes in vertical temperature structure.

4. *For Chapter 1, explanations of the processes involved in determining vertical profiles of temperature should represent the current state of understanding or lack thereof. The chapter should focus less on details of the vertical profile of temperature that are not resolved by the observations that are the focus of the report. For example, the satellite data are only reported in coarse vertical layers.*

Response: We have added text with discussion of vertical sampling and resolution issues with regards to how the observations may be able to represent the vertical temperature profile.

5. *For discussions that are felt to be too detailed for the body of the text, footnotes are a reasonable device.*

Response: We now employ footnotes for technical details that are inappropriate for the main body of the text.

Specific Comments

1. *The chapter should include more discussion of theories that provide physical constraints on the apparent differences between surface and tropospheric records.*

Response: Agreed. Done.

2. *The discussion on lines 69-80 should be replaced with a more accurate figure as well as a description of the differences between the tropics, the extratropics, and the polar regions. In the tropics, the temperature is hardly linear with height, given that the lapse rate associated with the moist adiabat goes from about 5 K/km near the surface to almost 9.8 K/km at the tropopause near 16 km. It should also be noted that the tropopause descends sharply to 12 km near 30 degrees latitude and to around 8 km near the poles. The existence of the near surface inversion layer at high latitudes should also be noted as well as its dependence on meteorological conditions.*

Response: Agreed. There is a new Fig. 1 which now illustrates the zonal-mean vertical profiles for Dec-Jan-Feb and June-July-August mean conditions, as obtained from NCEP reanalyses data. Accompanying details in the text include: description of the tropopause in the tropics, extratropics and polar regions; variation of lapse-rate with height in the tropics; sharp change of tropopause level at 30 degrees latitude; existence of surface inversion at high latitudes and its dependence on meteorological conditions.

3. *Relatedly, lines 100-131 should be replaced by a more complete discussion wherein it is noted that a radiative-convective balance is only likely to be of dominant relevance in the tropics, while in the extratropics, the lapse rate and the tropopause height are mostly determined by the same baroclinic instability that gives rise to weather systems (Schneider, 2004). Planetary-scale forced waves in winter and other circulation features, such as the Hadley and Walker cells, should be mentioned.*

Response: Agreed. Mention that radiative-convective balance likely dominant only in the tropics. Control of lapse rate and tropopause height in extratropics due to baroclinic instability. Mention of circulation features – Hadley and Walker circulations, planetary-scale wave forcing owing to land-sea contrast and flow over topography.

4. *The authors should provide further discussion of the role played by dynamics. The discussion of dynamics in lines 128-131 should introduce the concept of Rossby radius.*

This vital concept shows that dynamics tend to homogenize temperatures (above the boundary layer) over horizontal scales that vary from the planetary scale near the equator to a couple of thousand kilometers at midlatitudes and to a few hundred kilometers near the poles.

Response: Done.

- 5. The discussion in lines 133-137 should be strengthened, in particular so that it distinguishes between specific and relative humidity.*

Response: Corrected.

- 6. Remove “especially critical” from line 141.*

Response: Modified.

- 7. In lines 150-156, the question of internal variability needs to be improved and clarified. For one thing, there can be internal variability without external forcing, and even without air-sea interaction. Further, there are limitations associated with using numerical models to examine the importance of internal variability because such models poorly characterize such things as El Niño/Southern Oscillation (ENSO), the 1976 regime shift, and the quasi-biennial oscillation (QBO) at the levels of tenths of a degree. It should be emphasized that most rules of thumb used for atmospheric structure may not be appropriate at the level of the small temperature changes being considered in this report.*

Response: Phrase modified to “atmosphere-ocean-land-ice/snow climate system”. Limitations of models with regards to ENSO, QBO simulations mentioned.

- 8. It would be worth stressing that the temperature changes that are being discussed are only a few tenths of a degree. Much of our thinking is based on more substantial changes. There is an extensive literature arguing for and against the relevance of the moist adiabat in the tropics (e.g., Xu and Emanuel, 1989). However, even those arguing for its relevance would not argue that it should hold to better than a few tenths of a degree. Similarly, it might be argued that the role of motions should cancel when averaged over the earth. But the above is not strictly true. The existence of radiation leads to irreversibility, and when the strong changes in water vapor with latitude are taken into account, changes in circulation can lead to changes in global mean temperature that might be on the order of a few tenths of a degree.*

Response (applies to 7&8): Noted that caution has to be exercised as the report is focusing on climate trends of a few tenths of degrees.

- 9. In lines 194-202 and in lines 283-291, the report should be more cautious in arguing that local changes in radiative constituents can lead to local changes in temperature profile in light of such processes as the mean circulation in the tropics, which homogenizes temperature, and quasi-geostrophic dynamics in the extratropics.*

Response: Agreed. Clarification given.

- 10. In line 206, while the radiative impact of clouds is undoubtedly very important, further explanation is needed if one is to attribute to them a role as a “regulator”.*

Response: The word “regulator” is dropped, as it connotes something different from the focus intended.

- 11. In lines 222-224, it should be mentioned that greenhouse gas forcing in the tropics is not*

uniform owing to the current distribution of clouds and water vapor. Thus, greenhouse gas forcing from anthropogenic sources is greatest in dry regions.

Response: Agreed. This was an oversight. Corrected. Also, tropical feature has been mentioned.

12. *The claim of local radiative influence in lines 233-234 should either be explained or omitted.*

Response: Omitted.

13. *In lines 251-252, caution should be suggested in adding unknown forcings because these can easily become nothing more than adjustable parameters. Of course, care should also be taken to include all forcings that are quantitatively known.*

Response: Agreed, and wording has been changed accordingly.

14. *In lines 254-255, it should be noted that while the air-sea interaction can play a role in internal variability, such variability can also occur in the atmosphere alone.*

Response: Agreed, and clarification has been added.

15. *In lines 258-261, while water vapor and clouds are indeed critical to the high climate sensitivity of many models, the references cited (Stocker et al., 2001; NRC, 2003) carefully note that water vapor and especially clouds are areas of major uncertainty in models, and even in nature.*

Response: Agreed, and wording has been clarified.

16. *The discussion of volcanic influence on lines 309-312 should be reworked to include additional work that has been done on this subject. For example, there is more on the effects of volcanoes on European temperatures in Jones et al. (2003) and in Robock and Oppenheimer (2003). The most affected region is Northern Europe—not North America and certainly not Siberia. The two studies cited in lines 309-312 are also basically model studies, and evidence from observations is less convincing.*

Response: Agreed, suggested changes made and references added.

17. *The claim on lines 331-336 should note the substantial uncertainty of such factors as solar variability (Frohlich and Lean, 2004), historical volcanic forcing (Bradley, 1988), and aerosols (Charlson et al., 1992; Anderson et al., 2003).*

Response: Agreed, suggested changes made and references added.

18. *The report appropriately notes that the radiosondes show an abrupt increase in temperature in the troposphere around 1976 and the fact that this is missed in the satellite data which starts in 1979. It has been argued that the surface warming is simply the response to this jump with a delay due the heat capacity of the ocean (Lindzen and Giannitsis, 2002). This is distinctly relevant to the present report.*

Response: This is a valid point, and discussion noting the 1970s regime change coincident with the beginning of larger global warming has now been added, including the issue mentioned and the reference added, and we also note that models with changes in external forcing also simulate the 1970s increase in warming, and draw attention to the relationship between external forcing and internal decadal variability as a research problem that is currently under study.

Chapter 2

Major Comments

There are two main issues in the NRC review of Chapter 2 which have bearing on several of the Major and Specific comments. These will be discussed first.

- A. The quantitative discussion of precision and accuracy, and error knowledge of individual adjustments, though desired by the NRC panel is extremely difficult to affect. The original version of this chapter contained a table in which each potential error (i.e. diurnal drift, geographic sampling, etc.) was listed by columns with each observing system listed by rows. Unfortunately, very few of the entries could be confidently determined; indeed most entries were ad hoc and opinionative. The chapter authors elected instead to produce a table of “readiness” of the various observing systems as a “stop light” chart (green, yellow, red) in an attempt to describe in a weakly quantitative but defensible manner what we understood the strengths and weaknesses of the present situation of observing systems to be. In a number of places we have now inserted quantitative values for various errors of the datasets, but this is certainly not comprehensive as there is no source for such information.
 - B. There are several comments offered concerning statistical issues, i.e. trend calculation, (least squares regressions vs. median of pair-wise slopes vs. least anomaly regressions etc.), autocorrelation, and various types of uncertainty (Major Comments 1, 2, 7, 8, and 9). Following the NRC panel suggestion we have consolidated the discussion of statistical matters into an appendix.
1. *In Chapter 4, the following is asked: what is our understanding of the contribution made by observational or methodological uncertainties to the previously reported vertical differences in temperature trends? The nature of the discussion in Chapters 2 and 4 needs to be focused to reduce redundancy and avoid omissions in both Chapters 2 and 4. Chapter 2 appears to be focused on answering Chapter 4's question, rather than Chapter 2's questions. In general, some of the topical divisions between Chapters 2 and 4 are artificial, so some redundancy in material presented is inevitable. However, the committee suggests that Chapter 2 focus on the various observing systems and Chapter 4 focus on trends in the observations, as differentiated in the following:*

Chapter 2 should focus on:

- a. *explaining the measuring systems and instrumentation, their accuracy and precision, and spatial temporal variability for global measurements of temperature;*
- b. *addressing measurement issues both for surface temperature measurements and atmospheric temperature measurements;*
- c. *addressing spatial and temporal sampling errors; and*
- d. *discussing any particular geographic regions where measurement and retrieval errors are particularly large.*

Chapter 4 should focus on:

- a. *errors associated with trends; and*
- b. *assessing which of the bias errors in Chapter 2 could influence the trends, and why*

they do or do not do so.

The discussion related to trend estimation and uncertainties in Chapter 2 should be moved to Chapter 4. Text on reanalysis trends from lines 266-277 should be moved to Section 7 (“Reanalysis”) in Chapter 4. Also, Chapter 4 should add a section on “Methodological uncertainties” by including from Chapter 2 most of the text about linear trends in Section 2b (lines 385-460), discussions on structural uncertainty from lines 480-521, and the summary on “Errors or differences related to analysis or interpretation” from lines 584-601.

Alternatively, all material on trend estimation and uncertainties may be brought together in an appendix to the Temperature Trends report. In addition to the above material, discussion of statistical uncertainty in Chapter 3 (pages 39-40) could be included in the appendix.

Response: The NRC panel understands that the questions of chapters 2 and 4 have considerable overlap. We have edited and re-edited our drafts numerous times to find that balance of separating information, but including needed redundancy for flow. See (B).

- 2. Quantitative information is needed about the strengths and limitations of the observing systems. Specifically, quantitative discussion of the following sources of uncertainties should be included: accuracy and precision of the sensor, uncertainties in converting the fundamental measurement into temperature, and spatial and temporal sampling errors. There should be a summary of studies (with references) in which the different measurement types (e.g., radiosondes, active sensors, different satellite retrievals) have been intercompared and evaluated on a pixel level.*

Response: See (A)

- 3. Increased discussion is needed on surface temperature measurements and trends, to parallel the detailed discussion provided on atmospheric temperatures. From reading this document, the impression is given that global surface temperature measurement is a solved problem, but this is not the case. Description of skin and bulk sea surface temperature (SST) in Chapter 4.5.1 should be moved to Chapter 2 and should reference the recent work of Chelton (2005). Errors associated with sea surface temperature measurement are not adequately covered in either Chapters 2 or 4. Discussion of microwave SST and blended infrared/microwave products should be included. Note, the bulk SST is probably the suitable variable for trend estimation, but the skin SST values are needed to understand the variations in both bulk SST and atmospheric temperatures. Issues related to land-surface temperature measurement (skin versus screen) are not adequately addressed (see Jin and Dickinson, 2002).*

Response: See (B). Skin vs. Bulk is briefly mentioned. SST error discussion from Chapter 4 is now included. No land surface dataset from satellites is being considered, thus no discussion of land-skin is included other than in effects on microwave tropospheric temperatures. Due to space constraints the discussion is essentially limited to describing the datasets actually examined in this report.

- 4. Because we need to understand the processes contributing to the trends as well as measure the trends themselves, geographical regions having particularly large uncertainty should be addressed. For example, regional problems in surface temperature measurement should be discussed, including the Arctic Ocean and Southern Ocean, warm current regions, and the Indian Ocean.*

Response: Geographic regions of large uncertainty have now been mentioned, though it should be self-evident that regions with poor sampling will have greater uncertainty. An estimate of the impact on global trends is now included.

5. *Four and a half pages (pages 9-13) are devoted to "Reanalysis". Uncertainties in reanalysis trends are nicely summarized, and it is shown how the data are used by National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis and European Center for Medium-Range Weather Forecasts (ECMWF) Reanalysis (ERA-40). The conclusion is that there are considerable uncertainties in reanalysis trends, so reanalysis results are downplayed and not used in drawing conclusions in this report. The committee agrees with the authors' decision to deemphasize reanalysis data in the trend analyses in their report. However, this long discussion of reanalysis should be moved from Chapter 2 to page 19 in Chapter 3, where reanalysis temperature "data" are presented. Chapter 2 focuses on observing systems instead of particular datasets, and reanalysis products are not in fact "datasets". In addition, the four and a half pages of reanalysis discussions seem overly long in comparison to the approximately two pages for surface air temperatures and approximately four pages for upper air temperatures.*

Response: The discussion on Reanalyses remains in chapter 2. The reasoning, as stated in the text, is that the main conclusions of this report require a more mature set of reanalyses than are available to us at this time and the following chapters were tasked to analyze datasets selected for this report. Reanalyses also led to a finding and recommendation of chapter 2, thus warrant a bit more discussion here as they may hold significant promise for the future.

6. *Scientific justifications for future observing systems are listed in Chapter 6, such as why we need reference radiosondes, but this is not mentioned in Chapter 2. For example, after "no absolute standards" in line 512, one sentence can be inserted to state that reference instruments are needed for future networks, such as the global reference radiosonde network proposed by the Global Climate Observing System (GCOS).*

Response: Mention reference radiosondes

7. *The report states that two main methods are widely used for calculating trends: linear regression and a "nonparametric" method attributed to Gilbert (1987). In the statistics literature, a technique is said to be "robust" if it is insensitive to violations of the underlying assumptions (the presence of outliers is one example of how underlying assumptions could be violated). In this sense, linear (least squares) regression is not robust, though it is not clear that this is an issue in any of the climatic time series under discussion. Gilbert's method does not seem to be widely used, but there are other methods (e.g., methods based on minimizing the sum of absolute deviations instead of squared deviations as in least squares) that have a large literature and should be referenced. These include the use of R functions to perform robust regression (Venables and Ripley, 2002), semi-parametric regression methods (Ruppert et al., 2003) and additive models (e.g., Hastie and Tibshirani, 1990). The distinction between least-squares and robust methods is not likely the main source of uncertainty in analyzing climatic time series. Non-linear trends are discussed on pages 18-19 of Chapter 2, though without making a clear-cut recommendation. It is self-evident that the trend is non-linear over any respectably long time interval, but nevertheless, fitting a linear trend could be the best thing to do if one is simply interested in coming up with one number to represent a trend over a stated time period. In the view of the committee, it is not unreasonable to use linear trends in this kind of analysis, with two caveats: (i) it is important to remember that linear trends for different time periods will be different, and (ii) such linear trends should not be used for predicting future values. A further important issue is that when comparing observations and coupled model results, ENSO can appear*

in different sequences and magnitudes, making sampling a major issue. While linear removal of ENSO can ameliorate this problem, it is in fact impossible to remove all ENSO aspects even with multiple indices. As for ENSO (and similar) effects, the report discusses these on pages 18-19 of Chapter 2 but does not mention the most direct solution, that is, including ENSO (or other “natural variability” components) as additional covariates in the regression. There are arguments both for and against doing this, but comparing both analyses could be a useful reality check on the results.

Response: See (B)

8. *The report barely mentions the issues of autocorrelation, i.e., the fact that correlations in time series could severely affect the estimation of a trend, especially in the calculation of standard error. Chapter 2 discusses error bars extensively without mentioning this issue. Chapter 3 mentions it tangentially, with discussion of error bars on lines 876-886 and a passing reference to the first-order autocorrelation in the captions of Tables 6.1 and 6.2, but with no details about the method. Given the importance of correct treatment of autocorrelation in the assessment of linear trends, this seems to be a major omission. The report should acknowledge that autocorrelation is a problem, as it is generally done incorrectly, and recommend how to properly account for its influence. The standard errors of estimated trends, allowing correctly for autocorrelation and other effects, are likely comparable to the “uncertainties” due to instrument shifts and effects of that nature quoted at numerous places in the report. This could lead to a quite different perspective on the relative importance of “structural” as opposed to simple statistical errors.*

In fact, the method of Santer et al. (2000) seems to rely on the assumption that after subtracting trends, the time series is of AR1 form, which can indeed be characterized by the first-order autocorrelation. However, the AR1 assumption may not be correct and is certainly unnecessary as it is possible to fit a general ARMA (autoregressive, moving average) model with scarcely any more work. The “arima” function in the freely available R statistical package allows for fitting a linear regression component with ARMA errors, where the autoregressive and moving average components are of arbitrary order. The method is exact maximum likelihood, and standard errors are calculated for both the regression coefficients and the ARMA parameters. It should be noted that earlier versions of this method have been in use in the climatology literature for some time (Karl et al. 1996, 1998). Earlier discussions of time series approaches (e.g., including those based on fractional ARIMA models) have been given by Bloomfield (1992) and Bloomfield and Nychka (1992).

Another issue is whether to include an ENSO signal directly as a covariate in the analysis. In an analysis of annual hemispheric temperature averages, Smith et al. (2003) argued that inclusion of the Southern Oscillation Index as a covariate, though not having a great effect on the estimated trend, allows for specification of a lower-order AR model (AR1 rather than AR4) and in this sense simplifies the analysis. It would be of interest to see whether the same applies with the time series under discussion in this report.

Another method mentioned in Chapter 6 of the report is the adaption of methods from longitudinal data analysis (e.g. medical data in which individual subjects are followed for some period of time) a book by Diggle et al. (1996) is mentioned in this respect. While it is conceivable that these methods could be adapted to the estimation of trends in climatological time series, it also seems unnecessary, given that the AR/ARMA/ARIMA approach is quite well established. Therefore, discussion of this method should be omitted.

Response: See (B)

9. *Direct discussions with the authors of the report made it clear that they had given more*

consideration to statistical assessment of trends in time series than is apparent in the written report, but nevertheless, it was the strong view of the committee that the issues should be dealt with explicitly in the report. Based on the overall structure of the document, such discussion would logically belong with the “uncertainty” discussion in Chapter 4 rather than Chapter 2 but the authors might alternatively consider writing a separate appendix on the statistical issues associated with estimating trends in climatic time series.

Response: See (B)

- 10. There are insufficient bibliographic references to the technical aspects of temperature measurements and error determination and far too many references associated with climate variability and trends (these are more suitable to other chapters). Recent references (since NRC, 2000) should especially be included.*

Response: The number of references has been supplemented.

- 11. Cross evaluation and intercomparison of different technologies (including surface-based remote sensing) to measure temperature should be described.*

Response: Multiple systems, and reference systems are mentioned, but the main discussion is in Chapter 6 where recommendations are addressed in full.

Specific Comments

- 1. Observations not used in this report should be mentioned, such as why Television Infrared Observation Satellite Program (TIROS) Operational Vertical Sounder/Infrared (TOVS/IR) was never used for trend analysis. This probably should be mentioned after line 193. A discussion of TOVS temperature profiles is needed. Note, the tuned regression type analysis used by the National Oceanic and Atmospheric Administration (NOAA) is not the only temperature available from TOVS. The Pathfinder effort and French 3I effort represent research-quality retrievals. John Bates at NOAA is in the process of doing a careful calibration of TOVS so that trends can be determined.*

Response: The authors do not have access to a climate dataset based only on TOVS retrievals. NCEP perhaps comes closest; however, over land and island stations, NCEP accepts the radiosondes rather than TOVS. A footnote has been added to explain briefly the idea of a statistical, sonde-type retrieval.

- 2. Table 2.2 is used to answer the first question of this chapter but provides insufficient emphasis on the long-term temperature changes due to anthropogenic effects (i.e., temperature trends), which is the sole focus of this report. It would be useful to add one column to list the “Outstanding issues” regarding specific variation, which includes inconsistencies among different datasets (or observing systems) and what future data are needed for better characterizing and understanding this variation. The column “Effect on trend estimates” needs more quantitative information if available, such as how much the temperature trends change before and after removing ENSO signals in the time series.*

Response: Add to Table 2.2 “Outstanding issues”. Quantify “effect on trend estimate.”

- 3. It appears that Table 2.1 and the text on pages 22-24 were used to try to answer the second question for this chapter. The information given here is too general and too qualitative. More quantitative information and some references should be given on pages 22-24. For example, the authors can summarize how insufficient spatial sampling of the radiosonde network affects the temperature trend from Agudelo and Curry (2004) and others. The authors can give more information about radiosonde errors in the upper*

troposphere and lower stratosphere—such as radiation errors, their magnitudes and characteristics—errors in existing radiation corrections and how they affect the trends. Table 2.1 should include specific instruments and pixel size for satellite measurements. Humidity and wind measurements should be excluded from the table, although the authors may want to discuss how these measurements can be useful proxy diagnostics if measured carefully with climate-quality monitoring.

Response: See (A).

- 4. In lines 88-89, near-surface air temperatures over land are measured about 1.5-2 m above the ground level at official weather stations, rather than 1.5 m.*

Response: Height of surface sensor 1.5-2m is now in the text.

- 5. A reference is needed for this statement.*

Response: Uncertain what statement is being examined “A reference is needed for this statement”

- 6. The caption for Figure 2.2 should state that the pressure levels at the y-axis are radiosonde “mandatory reporting levels”.*

Response: Mandatory levels are now identified in Fig. 2.2

- 7. In lines 217-226, the reference for Global Positioning System-Radio Occultation (GPS-RO) is Kursinski et al. (1997). The comparison between GPS-RO and radiosonde data has shown that the GPS-RO soundings are of sufficiently high accuracy to differentiate performance among the various radiosonde types (Kuo et al., 2005). Also, the report should discuss the findings of Schroder et al (2003) on MSU versus GPS. In particular, Schroder et al. (2003) found that UAH T4 retrievals in the Arctic lower stratosphere in winter were biased relative to temperatures derived from GPS Radio Occultation measurements.*

Response: Reference for GPS-RO is Kursinski et al. 1997 is mentioned. We have introduced GPS as a future source of data to address the issue of vertical temperature trends. However, the authors have not seen a climate dataset suitable for analysis that is useful at this time. Hence the GPS datasets were mentioned, but not with elaboration.

- 8. The statement “the method of calibrating a radiosonde before launch may introduce time-varying biases” in lines 543-544 needs clarification.*

Response: Clarify “the method of calibrating a radiosonde before launch may introduce time-varying biases”. Added statement about ground calibration by typical temperature sensor vs. calibration against a traceable standard.

- 9. The references at the end of this review include several additional papers that should be considered for inclusion in Chapter 2 of the Temperature Trends report.*

Response: See references in this NRC report for addition to Chapter 2.

Chapter 3

Major Comments

1. A major issue is the drop in temperature associated with the introduction of the Vaisala sonde. It is stated that this affects the stratosphere, but it is unclear how deeply this systematic bias might extend into the troposphere. This is an important research problem that should be addressed

Response: We agree that the jumps in temperature associated with changes in radiosonde instruments, Vaisala or otherwise, are an important issue. The magnitude and vertical structure of resulting systematic biases remains an open question. More detailed discussion of such biases is deferred to Chapter 4. We include in our recommendations a need to better explore the nature of, and try to remedy the time-varying biases of the upper-air datasets.

2. *Mentioning the similarity of the basic data in the surface dataset while highlighting many of the potential problems (almost all of which have been adequately handled) sows doubts in the minds of readers. This gets picked up and emphasized in Chapter 6. The report should provide more explanation of how various problems in the data have been addressed and how this leads to some level of confidence in examining trends. For example, it can be easily shown by sub-sampling the surface data that the resulting hemispheric and global trends from the sub-samples would be almost exactly the same. References should be made to the frozen grid analyses work done in the Climate Research Unit (CRU) in the mid-1980s to the mid-1990s.*

Response: This is an excellent point, so the text has been modified to explain how the various problems have been dealt with, leading to more confidence in the data. Reference to the frozen grid analyses done at the CRU have been added as well.

3. *In general, the chapter would benefit from a more careful dissection of the global mean and a recognition that radiosondes are not near global. It does not address global mapping or the need for evaluations at co-located sites of sondes (see Hurrell et al., 2000; Agudelo and Curry, 2004; and Free and Seidel, 2005). There are no latitude-time series presented. This chapter does identify differences over high latitude land as being the main reason for surface being larger than troposphere trends in the extratropics (page 30), but this is not carried forward to the Executive Summary or Chapter 5. Weakening or removal of inversions over cold land or ice is a very good reason why the surface should warm more and a good example of why the global mean should be dissected.*

Response: The points mentioned, while of interest in their own right, are not critical issues to this report. Chapter 2 now includes more discussion of the issue of co-location. We do cover some of the issues related to sub-global spatial scales by way of presentation of trend maps and zonally averaged trends.

4. *The chapter also places too much emphasis on linear trends. Only linear trends as a function of latitude are presented, however this presentation can hide many things. The claimed agreement between radiosondes is not shown except for the linear trends (e.g., Table 3.6.1) (see Free and Seidel, 2005). There is nothing on root mean square differences, which are very revealing (Hurrell et al., 2000), or on monthly differences (smoothing the time series can be misleading).*

Response: The reviewers' statement "The chapter also places too much emphasis on linear trends" is contradicted by a statement regarding chapter 2, major comment #7, "In

the view of this committee, it is not unreasonable to use linear trends in this kind of analysis ...". Furthermore, it should be noted that the title of this synthesis report includes "Temperature Trends". Other measures, such as RMS, while of more general interest, are not particularly relevant to the focus of this report.

- 5. In a number of places, the assertion is made that the troposphere has warmed more than the surface. However, the differences in trends are often quite small, particularly for the 1958-2004 period. It is not clear that these differences are statistically significant. Although statistical significance is assessed for the trends themselves, no analysis of the significance of trend differences is presented. When comparing trend differences between two estimates of the same quantity (e.g., tropospheric temperatures from radiosondes and satellites), it is more appropriate to examine the trend of the difference time series, rather than trends for each time series individually (because the data contain similar overall variability). This is an omission that should be corrected, and the text should reflect the results of such an analysis. In particular, any statement about differences in warming should be weakened considerably if the differences are not statistically significant.*

Response: The suggestion to use difference series and their trends is a good one. We have added such analyses for assessing differences between datasets measuring the same parameter where it is most appropriate, in Chapter 4. Examination of differences between measures of tropospheric and surface temperature is now covered in Chapter 3 in the section on lapse rates (7.2). Assessment of statistical significance is detailed in the Appendix.

- 6. The trends calculated from reanalyses are downplayed because the input data sets are not homogenized. Although there is potential independent value of reanalysis products, it is not clear what trends from a reanalysis model mean in the context of temporally varying inputs. Therefore, the committee agrees with the decision of the authors to downplay this source of information.*

Response: This is not an action item since the reviewers are simply agreeing with our decision to downplay reanalysis data.

- 7. The issue of regional land-use and land-cover changes is brought up in a number of places, but the implications are not clearly addressed. For example, in lines 94-96 it is suggested that regional land-use change must be considered in the development of land-based data sets. However, if regional changes are large enough to have a measurable influence on global temperature, then these changes will be sampled and detected by the existing land-based networks. As such, why is this an issue when analyzing the differences among the data sets? There is an issue related to land-use and land-cover changes that could be addressed here or in other chapters. In the modeling discussions in Chapters 1, 5, and 6, land-use and land-cover is considered to be a forcing (with uncertain magnitude in the past) that is incorporated in some models and not in others. The committee believes this is correct and that land-use and land-cover should be considered as a forcing. Any land-use and land-cover effects in observational datasets should therefore be left in and not commented upon as a problem in Chapter 6. In other words, Chapter 6 cannot have it both ways the data are affected by land-use and land-cover change, so they are somehow wrong, yet this forcing is omitted from many models.*

Response: We agree with the reviewers and have made appropriate changes in section 2.1 to reflect the fact that land use changes are considered a forcing and that any related changes in temperature will be detected by the existing land-based station network.

- 8. The Fu et al. results have the potential to be centrally important to the issue of tropospheric temperature trends and should be discussed more thoroughly in lines 863-868. Attempts to separate tropospheric and stratospheric contributions to trends are reasonable. They*

should not be rejected with the value statement that they are “controversial”. The only published criticism of the Fu et al. approach is by Tett and Thorne (2004), with other criticisms in the grey literature. The Fu et al. method has since been followed up by several studies which show that it is robust, including further research by Fu and colleagues and Gillett et al. (2004). The potential clarification that the Fu et al. method can contribute to the central issues is very significant.

Response: As suggested by the reviewers we will no longer refer to the Fu et al. technique as controversial. Since the first draft of this report several additional manuscripts related to this technique have been written suggesting that the technique can add insights to the issues of concern. However, the need to rely on this technique is somewhat reduced by virtue of the fact that we now have lower tropospheric satellite temperature datasets produced by two separate groups (UAH and RSS), that have virtually no contribution from the stratosphere.

- 9. The difference between the Remote Sensing Systems (RSS) and University of Alabama, Huntsville (UAH) trends is left as an open issue, with no relative value given. It is important to resolve this discrepancy, if possible. The trend difference in the mid-troposphere is the same size as the signal: zero for UAH and +0.1 K/decade for RSS. If no distinction can be made, then no conclusion can be drawn. Statements in lines 355-359 and elsewhere about discrepancies between RSS and UAH as being mostly due to the NOAA 9 satellite are misleading as can be seen by looking solely at the post 1987 period. In fact, examining differences between the two datasets, which are not shown in the report, reveals major issues remaining on adjustments for other satellites and diurnal cycle issues (especially as a function of latitude and in the tropics).*

Response: The reviewers make a good point in that the differences between the RSS and UAH measures of tropospheric temperature are worthy of more discussion and we have added such in Chapter 4. We also agree that the RSS-UAH discrepancy is due to more than just the NOAA-9 transition and have modified the text to reflect this.

- 10. In lines 791-816, if the tropical tropospheric temperature profile behaves as a moist adiabat, which to an approximation it does, then the lapse rate is expected to decrease as temperature increases (i.e., as the surface warms, the troposphere is expected to warm more). This is the “global change theory” the authors refer to in Section 6.2.1. Therefore, it is no surprise that when the surface warms due to ENSO, the troposphere warms relative to the surface (line 798), or that when the atmosphere warmed in 1976-1977, the lapse rate dropped (line 802). These results are currently presented with no link to physical theory. The authors say that “the variation in tropical lapse rate can be characterized as highly complex, with rapid swings over a few years, superimposed on persistent periods of a decade or more”, but our guess is that much of this variation can be explained by changes in the mean temperature. Further, the authors say that the enhanced warming of the troposphere associated with surface warming gives “enhanced static stability” (lines 799 and 803). A reference should be provided for this statement. It should be noted that the troposphere did warm relative to the surface in the tropics during the 1997-98 El Niño event, which is a large signal. Also, the report should reference a study by Gettelman et al. (2002) on changes in stability. This study highlights the observed increases in Convective Available Potential Energy (CAPE) that are not replicated by models (which remove all CAPE), and so it is also relevant to Chapter 5 of the report.*

Response: The purpose of Chapter 3 is to discuss what the temperature observations indicate rather than physical theory. We note that further discussion of moist adiabatic theory is given in Chapter 5, which is a more appropriate location for such. The reviewers note that the lapse rate changes are described as complex. Perhaps some confusion regarding this has resulted from the lack of an appropriate graphic accompanying the

discussion. As such we have added a new figure. Also, we have removed the statements regarding “enhanced static stability” since this was an inference, rather than a finding. The suggested reference by Gettelman et al. has been added.

Specific Comments

1. *The numerical system for numbering the figures is overly complicated and inconsistent. It would be simpler to number the figures 3.1, 3.2, 3.3 etc., rather than 2.4, 3.3, 4.4, 6.2, 6.2.2, 6.2.3, 7.1. In all the figures, the notations used to label the curves in the diagrams are different from the descriptions in the captions. For example, Figure 2.4 has the labels N, G, and U, and these are not defined in the caption. The same is true in different ways for 3.3, 4.4, etc. Also, without a very good color print, the different colored lines can be difficult to distinguish.*

Response: The numbering system for the figures will be unified for the final draft, as will the captions. Corrections and simplifications to the numbering of the figures and tables have also been made in the interim.

2. *In lines 53-55, comparing results from more than one dataset also provides a better idea of the uncertainties or at least the range of results.*

Response: The suggested edit has been made.

3. *In lines 86-88, the statement that homogenization procedures are “quite successful” at addressing these issues should be more nuanced. While we are in agreement with the statement with regard to biases introduced by changes in time of observation, we are less confident that other issues (e.g., exposure changes) can so readily be addressed because there is often a lack of metadata.*

Response: The literature suggest that statistical homogeneity adjustments on surface data work quite well and this is true even in the face of the fact that metadata are rarely complete and sometimes ambiguous or wrong. A reference to WMO guidance on metadata and homogeneity has been added.

4. *In lines 107-111, the benefits of sea surface temperature (SST) over night marine air temperature (NMAT) are discussed without saying anything about what the relationship between SST and NMAT is likely to be (e.g., is SST a good proxy for NMAT?).*

Response: As recommended, more discussion has been added to describe the differences and similarities between NMAT and SST.

5. *There should be a reference in line 163 to Jones et al. (1997, 2001). These papers give details of the procedure for allowing for changing numbers of observations through time.*

Response: The suggested reference has been added.

6. *This text in lines 180-183 is a bit wordy and does not follow on well from the previous sentence. The paper by Vose et al. (2005) should show that the differing techniques with the same data produce almost the same results.*

Response: The text was changed as suggested and the reference was added.

7. *In line 205, the text “since neither choice is optimal” suggests that there is a single optimal approach. This should be rephrased to “since each approach has advantages and disadvantages.”*

Response: The change has been made as suggested.

8. *In lines 229-231, the Radiosonde Atmospheric Temperature Product for Assessing Climate (RATPAC) data set incorporates different homogeneity adjustments before and after 1997. Has anyone evaluated the extent to which this might introduce an inhomogeneity into this data set?*

Response: Some consideration of this issue was made in the development of the dataset as discussed in the cited manuscript that describes this dataset.

9. *Lines 294-296 state, "There is some ambiguity about whether the temperatures return to their earlier values or whether they experience step-like falls". Surely this is just a matter of how best to describe the curves. A more important question is whether the observations agree with particular models (global circulation models or theoretical models). Has anyone suggested a plausible mechanism that would give a step-like cooling after a volcano (e.g., Douglass and Knox, 2005)?*

Response: The description of the temperature change may have implications for the physical mechanisms driving the variations so it is worth noting possible nonlinear behavior. However, any such proposed mechanisms would be subject to debate and is outside the focus of this report.

10. *In lines 297-299, is the interannual variability really mainly due to the Quasi-Biennial Oscillation (QBO)? If so, a reference should be provided.*

Response: We don't state that the interannual variability is due "mainly" to the QBO. We describe the associated variations as "small amplitude". A reference for the QBO has been added.

11. *The text in lines 299-301 makes it sound like the stratospheric cooling trend has been completely explained as a combination of the responses to stratospheric ozone depletion and cooling due to carbon dioxide. This is in disagreement with the Executive Summary, which indicates that the cooling cannot be fully explained by these forcings.*

Response: We never say "completely explained", rather we say "to a large extent".

12. *In lines 301-305, there are various descriptions of the curve including "the aforementioned step-like drops represent a viable alternative to a linear decrease". What do the authors mean by "a viable alternative"? Presumably, they do not mean one based on a physically-plausible mechanism. Again, this seems to just be a discussion of how best to describe the curve, whereas the real issue is whether the observations agree with theoretical predictions.*

Response: The cited reference (Seidel and Lanzante 2004) explains that the alternate descriptions of the behavior of the curve are both plausible from a statistical standpoint. See also response to #9.

13. *In lines 320-323, the change to the Vaisala radiosonde in certain tropical areas is given as a possible reason for the differences in the two radiosonde data sets. What analysis has been done to suggest this possibility? Or is this statement made simply because the timing is coincidental?*

Response: It is just a matter of coincident timing. We have simply pointed out that such changes have taken place at this time.

14. *The nomenclature of TMid-Trop-R and TMid-Trop-A are introduced without definition in line 358. At least a reference to Chapter 2, Figure 2.2 and related discussion should be included for those who may start reading here. Are these just the Microwave Sounding Unit (MSU) channels and their radiosonde integral equivalents, or something else? Also, the nomenclature in the figures and captions is inconsistent and not sufficiently defined.*

Response: A reference to Chapter 2 has been added as suggested. Chapter 2 has been modified to give a more thorough discussion of this issue. The nomenclature in the figures and captions has been unified.

15. *TTrop-UW-A is introduced without definition in line 396.*

Response: A reference to Chapter 2 has been added for clarity.

16. *In line 447, the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalyses go back to 1948. It is probably best to ignore the period between 1948-57 as this study only goes back to 1958.*

Response: As a matter of completeness we indicate the full period of available data.

17. *A reference to Simmons et al. (2004) might be needed in line 475, or a reference back to Chapter 2.*

Response: A reference back to Chapter 2 has been added.

18. *The Pielke and Chase (2004) reference in line 488 is missing from the reference list.*

Response: The missing reference has been added to the bibliography.

19. *It is not completely clear what is meant in lines 502-505. Presumably this relates to the abrupt change in the late 1970s.*

Response: The interpretation is correct. The statement has been made more explicit in the text.

20. *What do the authors mean by "it has been shown that such constructs are plausible" in lines 505-508? What criteria are used to judge their plausibility? Presumably it is just how well they fit the data. In this case, you could make a perfect fit to the data by regressing it on itself. Again, the real issue is whether the observations fit with theoretical predictions.*

Response: See response to comment #12.

21. *It is unclear what is meant in lines 515-518.*

Response: As stated in the text, the two stratospheric temperature measures behave inconsistently in comparing the two radiosonde datasets.

22. *In lines 521-527, The reanalysis models tend to agree better with the climate model predictions than do the raw observations. Is this an alternate explanation of the differences between the reanalyses and the raw observations (i.e., if the reanalysis model has similar physics to the climate models, then its troposphere will warm more than its surface)?*

Response: This chapter is not an appropriate place to comment on agreement between observations and models. Such analyses are discussed in Chapter 5.

23. *It is not entirely clear what is meant in lines 542-544. Do the authors mean something like “trends in land air temperature in coastal regions are generally consistent with trends in SST over neighboring ocean areas”?*

Response: The ambiguity has been cleared up by using the reviewers' suggested wording.

24. *In lines 548-552, the authors do not mention the most obvious explanation for enhanced warming over land, namely the smaller effective heat capacity over land than ocean. Enhanced warming over land is seen in every climate change simulation and does not relate primarily to the phase of ENSO, though this could be a contributor. Better justification should be provided for a link between warmer temperature over Siberia and ENSO. Siberia encompasses a large area, so be more specific and provide a reference.*

Response: The edits have been made as suggested by the reviewers. The discussion of Siberia has been removed.

25. *In lines 561-563, SSTs and NMAT have different trends for short periods owing to ENSO and changes in surface fluxes, as shown in other works.*

Response: The text has been changed as suggested and a reference has been added.

26. *In lines 568-570, these differences might be related to an increase in mean ship height above the sea surface.*

Response: This is not true since the cited reference used data that incorporated corrections to account for changing ship height. This is now explicitly stated in the text.

27. *While the explanation in lines 589-591 sounds plausible, has it ever actually been shown? Are the free tropospheric temperatures more highly correlated with maximum surface temperatures than with minimum temperatures? If there is not a published reference on this, then should be removed.*

Response: This has not been shown in any published works, so the text in question has been removed.

28. *In lines 600-601, it is not clear whether this relative change in trend in the troposphere and surface is statistically significant in the recent era. Visually, it does not seem that impressive or obvious.*

Response: This is a statement that is qualitative in nature and as stated in the text is self evident from comparing trends for the radiosonde and satellite eras. The nature of the relevant change in the tropics is now clarified by the introduction of a new figure discussed in section 7.2 (time series of surface-tropospheric temperature difference).

29. *The comparison the authors make in lines 627-629 is equivalent to assuming that the Maryland stratospheric trend is the same as that in the other two datasets (since the Fu et al. approach is just to fit to a regression model).*

Response: That is correct and a phrase to this effect has been added to a footnote in the sentence in question.

30. *In lines 655-656, “TLow-Strat-A” and “TLow-Strat-B” need definitions.*

Response: The terminology that replaces these is defined in Chapter 2.

31. *In lines 656-657, why is the cooling at the South Pole not more dramatic, especially given known problems over sea ice (Swanson, 2003) and the high ice sheet of Antarctica that greatly impacts channel 2? In fact, it looks like the cooling is larger in the northern hemisphere midlatitudes.*

Response: It is not clear why this is so. The primary purpose of Chapter 3 is to illustrate observed changes in temperature, which will be explained later in the report.

32. *Replace "Soviet stations" in line 672 with "stations located in Russia and other countries of the former Soviet Union".*

Response: A change as suggested has been made.

33. *The word "granularity" should be replaced in line 693.*

Response: Granularity seems like an appropriate word here.

34. *In line 696, replace "noisy patterns that result" with "noise that results".*

Response: The suggested change has been made.

35. *The figure labeling (a, b, c and d) in line 704 is incorrect.*

Response: Reference to the various figure panels is correct.

36. *No mention is made of the Antarctic in line 713.*

Response: Antarctic has been added as suggested.

37. *In lines 722-724, the sharp contrasts only seem to be around the western coasts of the Americas.*

Response: The clarification has been made as suggested.

38. *The unit for a lapse rate trend looks wrong in line 823. Surely it should be $K km^{-1} decade^{-1}$ or something with the same dimensions.*

Response: The units are correct since the quantity is a vertical temperature difference rather than an actual lapse rate. The reason for this and its implications are discussed at the beginning of section 7.2.

39. *Are there missing crosses for the surface in Figure 6.2b, or do they all overlap?*

Response: They overlap. For further clarification, the trend values from this figure have been placed in a new table.

Chapter 4

Major Comments

- 1. There should be better discussion of the Fu et al. approach. Simply stating that this is controversial is a value judgment and not an adequate reason for dismissing the approach. The review panel sensed that some of the authors had more specific objections to the approach, but these are not adequately documented. For example, why should it be a problem that the approach uses negative weights for part of the signal? The ultimate goal here is to eliminate or reduce stratospheric contributions to middle troposphere temperature trends. As described in Line 184, about 10 percent of the weight of Channel 2 comes from the stratosphere, but the integrated weight for Fu et al. weighting function is near zero. As stated on Line 187, the stratospheric contamination on TMid-Trop trends is about 0.05 K/decade, while the trend uncertainty due to the uncertainty in derived coefficients in the Fu et al. method is only about 0.01 K/decade. The potential for incorrect stratospheric temperatures to corrupt the mid-tropospheric values should receive greater emphasis in Chapter 4 and Chapter 5. In conclusion, the Fu et al. method appears to reduce stratospheric contributions and may represent a valuable resource for this report. The report could, if appropriate, include references to more recent work of Fu et al. and possibly other authors. The new papers might give more insights on controversial issues of negative weights in the Fu et al. method and its impacts on trends. The Fu et al. (2004) reference on line 487 is missing in the chapter's references.*

Response: We have added material concerning the Fu et al methods, and moved the discussion of these methods closer to the section that deals with tropospheric temperatures. We have included references to the more recent Fu et al work, as well as model based work supporting Fu's approach. We have also included a discussion of Fu's arguments concerning the vertical trend consistency of the UAH and RSS datasets.

- 2. The report gives a very even-handed discussion of the reasons for different trend estimates by UAH and RSS. Is there any way to go further, for example by stating which approach is better or proposing ways to reconcile the two approaches? Would the authors recommend further statistical analyses? If so, what form should these take? It appears to be the case that, although issues with diurnal corrections and the calibration target are important, these are not the major reason why the two groups obtain different trend estimates. These differences appear to hinge on the different treatments of the NOAA-9 satellite. It may be possible to do better using Bayesian statistical methods. For example, one could treat the unknown shift of the time series (resulting from the change of satellites) as a parameter with a prior distribution, construct a posterior distribution by analyzing data from both satellites, and then integrate out that posterior distribution using Monte Carlo methods to derive a reconstructed time series that allows for uncertainty in the shift. This method could potentially work better than current methods when there is only a very limited amount of overlapping data. In lines 294-300, the authors use the lack of a diurnal correction in the University of Maryland (UMD) dataset as an excuse for not discussing it. Because of the differences between UAH and RSS and the small residual uncertainty from diurnal sampling, it could be informative to use the UMD dataset as an independent check to understand and possibly reconcile the differences between UAH and RSS. The suggestion that the correction for target temperature is a function of latitude (or orbit relative to the Sun), as done by the UMD group (Grody et al., 2004), but not by UAH and RSS, is an interesting one and builds in some diurnal cycle corrections. These issues ought to be discussed openly.*

Response: The response to this comment is long and complex, mostly due to the rapidly evolving situation with regard to the availability of various datasets. We will start with the middle troposphere data (now called $T_{\text{sfc-75}}$).

Currently, we do not have enough information to resolve the differences in trends between the UAH and RSS datasets. We are confused by the review panel's statement that the target factors are not important – and that the difference “appear to hinge on different treatments of the NOAA-9 satellite”. The main difference in the “treatment” is the NOAA-9 target factor. We have reviewed the differences between the two datasets in greater detail, and we now attribute a significant amount of the discrepancy in the globally averaged time series to differences in the NOAA-11 target factor. We have also added more discussion of the differences on a regional scale, where differences in the diurnal adjustment and different approaches to calculating intersatellite offsets (one constant global offset vs. zonally varying offsets) are also important. We have expanded the section that discusses the UMD results, and discussed both the innovations and possible shortcomings of their new techniques. We have added a figure of the difference time series between the RSS and UMD results and the UAH results. Both these difference time series show large jumps during the NOAA-09 lifetime, limiting the usefulness of the UMD time series to the problem of differentiating between the RSS and UAH series.

Now we will discuss the situation with $T_{\text{sfc-350}}$ (formerly known as TLT or $T_{\text{Low-Trop}}$). Since the last draft of this report, RSS has produced a new version of $T_{\text{sfc-350}}$, which is discussed in a paper submitted to Science. In the process of creating this dataset, RSS found that the diurnal adjustment approach used by UAH was flawed in several ways, at least one of which has been agreed to by the UAH scientists. UAH has subsequently implemented an entirely new diurnal correction method for $T_{\text{sfc-350}}$, and constructed a new version of their dataset which shows significantly larger trends in the tropics, and is more consistent with other UAH datasets, as can be seen by comparing trends in the $T_{\text{sfc-150}}$ layer with the $T_{\text{sfc-350}}$ layer.

- 3. Overall, the satellite uncertainty is summarized in detail, and in depth, while the radiosonde uncertainty is described in less detail and less quantitatively (see below for more detailed comments). There is no discussion of the strengths and weaknesses of homogenized methods used by different dataset groups. There is a lack of attention to developing physical-based correction schemes. For example, radiosonde radiation error is the main source of errors for upper troposphere and lower stratosphere temperatures. It appears that none of the groups has implemented radiation corrections to non-corrected historical data or adjusted applied corrections. It is true that the trend analysis relies more on long-term homogeneity than on the absolute accuracy. But accurate data throughout the period would minimize the temporal inhomogeneity and can be used for other studies. Also, the report has no discussion of missing data within a month for radiosonde data. In Hadley Center Radiosonde Temperature (HadAT) only 12 soundings are required to make a monthly mean and two monthly means to make a season; there is no allowance for this in the error bars. Missing months are especially an issue in the tropics, where records are woefully incomplete, as shown by Hurrell et al. (2000). Free and Seidel (2005), however, find missing monthly data to have a fairly minor effect on trends.*

Response: The issue of radiation correction is discussed in more detail. We now mention the physical-model based corrections of Luers and Eskridge, as evaluated by Durre et al. However, since this report is narrowly focused on long term trends and other changes in atmospheric temperature, it would be inappropriate to focus on methods that only improve the accuracy of mean values (i.e. radiation corrections in the absence of instrumentation/procedural changes), however valuable they might be to the community at large.

We have added additional material concerning temporal sampling issues.

- 4. There is no discussion of statistical uncertainties in methodologies for calculating trends, calculating monthly mean values and creating global time series (i.e., spatial averaging techniques for radiosonde data). Some of this discussion appears instead in Chapter 2. Somewhere (Chapter 2, Chapter 4, or a separate appendix) there should be a separate section on statistical methods for estimating trends in time series, including standard errors or other measures of statistical uncertainty.*

Response: A separate appendix has been added to the document that discusses these and other statistical issues.

- 5. The largest discrepancy between radiosondes and satellite estimates of trends is in the stratosphere. More detailed discussions on the stratosphere discrepancy are needed in Section 2. Section 2.1 briefly describes two uncertainties associated with undetected changes in instrumentation and early bursting of balloons in early radiosondes. There can be significant biases in the radiosonde temperature data in the stratosphere due to radiation errors. Both radiosonde datasets do not include physical models for radiation adjustments. Durre et al. (2002) show that Luers and Eskridge (1998) adjustments make radiosonde temperatures more homogeneous in the stratosphere, although it frequently amplifies the discontinuities in the troposphere. Regarding the statement “The discrepancy ... is likely to be mostly due to pervasive uncorrected biases in the radiosonde measurements” on lines 96-98, can the authors be more specific about what those uncorrected biases are? What about time lag errors of radiosonde data that could cause a cold bias in the stratosphere? There are minimal discussions on the largest disparities in the tropics between two radiosonde datasets and between radiosonde and satellite data in Figure 6.2.2. in Chapter 3. How does the difference in station distributions between these two radiosondes contribute to the largest discrepancy in the tropics? Is the enhanced cooling in the tropics relative to the midlatitudes in the stratosphere in radiosonde datasets due to a lack of sampling over open oceans, or is it due to larger adjustments associated with the switch to Vaisala radiosondes for most of tropical stations? It seems that the former has minor impacts because Figure 6.2.3 in Chapter 3 shows that the stratosphere trends in the tropics are zonally uniform.*

Response: We have added a discussion of radiation and “time lag errors” to Section 2. We mention new results from Sherwood et al and Randall and Wu that may help explain the enhanced cooling in the tropics relative to the satellite measurements, and indicate that there may be uncorrected solar-heating induced biases in the LKS dataset. As mentioned above, we now mention the physical-model based corrections of Luers and Eskridge, as evaluated by Durre et al. We have also included results from Free and Seidel that indicate that spatial sampling errors are unlikely to be large enough to explain the sonde-satellite discrepancies.

- 6. It seems that the difference in homogeneity adjustment methods is the main contributor to disagreements in trends among different radiosonde datasets presented in Sections 2.1 and 3.1. Do the adjustments reduce or increase the discrepancies in trends (by comparing the trends before and after the adjustments)?*

Response: Since the majority of the adjustments made to the sonde data remove artificial cooling biases, the adjustments serve to move the sonde trends closer to the satellite trends. The question of whether the sonde trends come closer together doesn't make sense, since the sondes used in the NOAA dataset are all used in the UK dataset.

Specific Comments

1. *In lines 176-193, does the bias in the radiosonde-derived TMid_Trop from stratospheric errors have the same magnitudes of about 0.05 K/decade for NOAA and UK-Met datasets? As shown in the middle panel of Figure 6.2.2 in Chapter 3, the difference between TMid-Trop-U and TMid-Trop-N at around 5° N is about 0.1 K/decade. Adding ~0.05 K/decade to both datasets still cannot explain the large disparity between two datasets at this latitude.*

Response: The material the generated this comment has been removed, as it is no longer necessary in light of the expanded treatment of the Fu et al approach.

2. *In lines 335-347, how can the uncertainty of the lower troposphere temperature record be consistent with the mid-troposphere uncertainty, especially given that the mid-tropospheric record is biased low by contaminating lower stratosphere influences?*

Response: As discussed above, the lower-tropospheric (Tsfc-350) situation has changed dramatically since the last draft. The UAH Tsfc-350 product was biased low due to an error in the diurnal adjustment, and new evidence has emerged suggesting that even the radiosonde measurements at the pressures that make up this bulk of this layer may have an artificial cooling bias due to radiation errors that evolve in time. As a result, this part of the document has been completely rewritten.

3. *Section 4.3 fails to examine root mean square (RMS) differences (e.g., Hurrell et al., 2000) and only deals with average trends.*

Response: We do not focus on other statistical measures of agreement because the focus of this report is on long-term trends.

4. *The surface record also has problems that are not discussed in Section 5.1 of Chapter 4. In particular, no error bars are assigned to the systematic corrections.*

Response: The error bars shown in Fig 4.4 include errors in the systematic corrections.

Chapter 5

Major Comments

1. *More model results included.*

Response: In the version of Chapter 5 that was reviewed by the NRC, the section entitled "New Comparisons of Modeled and Observed Temperature Changes" showed results from three models only (PCM, CCSM3.0, and GFDL CM2.0). At a presentation to the NRC committee on Feb. 23, 2005, the CLA of Chapter 5 reported on a more comprehensive (but unpublished) analysis of surface- and atmospheric temperature relying on over a dozen climate models. The NRC committee requested that Chapter 5 should include "as many results from additional models as time allows" (NRC report, page 32, point #6). This has now been done. The revised version of Chapter 5 examines results from historical forcing runs ("20CEN" experiments) that were completed in support of the IPCC Fourth Assessment Report (FAR). Chapter 5 considers a total of 49 20CEN realizations performed with 19 different coupled atmosphere-ocean GCMs. The more comprehensive model results are incorporated in Figures 5.2, 5.3, 5.4, and 5.6, and are discussed in Section 5. Figures 5.5 and 5.7 (which previously provided results from NCAR and GFDL models only) now include results from the GISS-EH model, so that all major U.S. modeling groups are represented in these two Figures.

2. *Inclusion of T^*_G and T^*_T comparisons.*

Response: The reviewed version of Chapter 5 discussed "contamination" of MSU channel 2 by the cooling stratosphere, but did not attempt to remove this contamination in its comparisons of modeled and observed tropospheric temperatures. The NRC committee regarded this as a deficiency (NRC report, page 32, point #3). Chapter 5 now uses the approaches outlined by *Fu et al. (2004a, 2005)*¹ to remove the stratospheric contribution to T_2 , thus yielding T^*_G (for global-mean tropospheric temperature changes) and T^*_T (for tropospheric temperature changes in the deep tropics). Stratospheric influences are removed from both model and observational T_2 data. The new T^*_G and T^*_T results are shown in Figures 5.3, 5.4, and 5.6, and are described in Section 5.

3. *Inclusion of T_{2LT} Comparisons.*

Response: In the version of Chapter 5 that was reviewed by the NRC committee, the section on "New Comparisons of Modeled and Observed Temperature Changes" (Section 5) focused on trend comparisons involving T_4 , T_2 , T_{SFC} , and T_{SFC} minus T_2 . Trend comparisons involving T_{2LT} were not given. This was largely because of the dearth of information on structural uncertainties in T_{2LT} . At the time the first version of Chapter 5 was completed (in January 2005), only one group (the UAH group) had provided satellite-based estimates of changes in T_{2LT} , while three groups (UAH, RSS, and UMD) had supplied MSU-based estimates of changes in T_2 . Publication of a new T_{2LT} retrieval by the RSS group (*Mears and Wentz, 2005*) helped to quantify structural uncertainties in satellite-based estimates of lower tropospheric temperature change. It is now more meaningful to calculate changes in synthetic T_{2LT} from the IPCC 20CEN runs, and compare these with T_{2LT} changes derived from RSS and UAH (and with synthetic T_{2LT} changes from radiosondes). Such T_{2LT} comparisons are now provided in Figures 5.2B, 5.3D, G, 5.4D, G, 5.5, and 5.6A, C). The inclusion of these comparisons directly addresses a persistent concern raised by the CLA of Chapter 2.

¹ All references cited in this response can be found at the end of the revised version of Chapter 5.

4. *New box on fingerprint methods.*

Response: The NRC committee requested a description of fingerprint methods “that would be appropriate for a climate scientist who does not work directly in this area of research”. The committee made a specific suggestion regarding such a description. The LAs of Chapter 5 felt that the suggested text was too lengthy (nearly 1 ¼ of single-spaced text), and more technical than the rest of Chapter 5. Accordingly, we have shortened the NRC committee’s suggested text and tried to make it more suitable for a non-specialist audience. To avoid interrupting the “flow” of Chapter 5, we have placed the discussion of fingerprint methods in a separate box (Box 5.5). The new text explicitly addresses the “strengths and limitations of detection and attribution analyses” (NRC report, page 32, point #7), and also outlines a number of key choices that are typically made in the practical implementation of fingerprint methods.

5. *Discussion of black carbon aerosols and LULC changes shifted to boxes.*

Response: The reviewed version of Chapter 5 included an extensive discussion of forcing by black carbon aerosols and land use/land cover changes (LULC) in the middle of Section 3. The LAs of Chapter 5 felt that these issues were more suited to treatment in self-contained boxes. The new Boxes 5.3 (black carbon) and 5.4 (LULC) now provide two specific examples of spatially- and temporally-heterogeneous forcings, and describe some of their likely effects on surface and atmospheric temperatures.

6. *Key Findings and Recommendations are now given up front.*

Response: The “Key Findings and Recommendations” are now given at the beginning of Chapter 5. Previously, the “Key Findings” were at the end of Chapter 5, and no “Recommendations” section was provided. Some of the “Key Findings” have been modified in the light of new analyses of model and observational datasets (see points #1-#3 above). New “Key Findings” regarding modeled and observed amplification behavior have been added. Most of the old “Summary” section of Chapter 5 (section 7) has been deleted. The new “Summary” section now consists of a single paragraph.

7. *New Tables added.*

Response: We have added several new Tables. Table 5.1 introduces the 19 atmosphere-ocean GCMs whose results are discussed in Section 5. Table 5.2 summarizes the natural and anthropogenic forcings that were applied in the 20CEN runs performed with these 19 models. Table 5.4 provides information on basic statistic properties (mean, median, standard deviation, etc.) of the temperature trends computed from the 49 20CEN realizations.

8. *Vertical profile Figures dropped.*

Response: We have removed the two Figures (Figures 5.7A,B in the NRC-reviewed version of Chapter 5) showing the vertical profiles of global-mean and tropical-average temperature trends. The information on amplification of surface warming that these two Figures contained is now conveyed in a more comprehensive way in the new Figure 5.6.

9. *Discussion of statistical significance testing.*

Response: The revised version of Chapter 5 now includes a paragraph (on page 27, para. 3) explaining why statistical error bars are not used in Figures 5.3 and 5.4 (see NRC report, page 31, point #1).

10. References updated.

Response: A number of new references have been incorporated in response to NRC review comments.

Major Comments Response

1. Major Comment #1

- a. "...conclusions reached are often based on estimates of trends, neglecting uncertainty levels, and many statements on comparison are inaccurate because of this..."
- b. "...report should be more explicit about the choices made regarding the treatment of trend confidence intervals in model-data comparisons".

Response: Completed. See "Major Comments", point #9.

- c. "The second and third conclusions regarding the influence of volcanoes and El Niño...could be removed and only briefly mentioned earlier".

The ENSO conclusion has been removed. The effect of volcanoes on surface and atmospheric temperatures is now described in "Key Finding" #2 (page 2, lines 24-28).

"There should also be discussion of volcanoes in the context of *Douglass and Knox (2005)* and *Lindzen and Giannitsis (1998, 2002)*.

Response: The papers mentioned in this comment attempt to provide empirical estimates of climate sensitivity from the surface and/or atmospheric temperature response to volcanic eruptions. They are three specific examples of an extensive body of literature. They are atypical of the literature as a whole, in that all three arrive at very small estimates of climate sensitivity. Our opinion is that the three cited articles (and related papers) are not central to the mandate of Chapter 5. In a new footnote (footnote #13 on page 14), we mention these and other empirical estimates of climate sensitivity, and note that such investigations "...are not directly relevant to elucidation of the causes of changes in the vertical structure of atmospheric temperature, which is the focus of our Chapter".

2. Major Comment #2

- a. "Error bars are essential on the plots, notably Figures 5.3 and 5.4, and all dots should be horizontal bars to allow for sampling uncertainty. This is important because ENSO is not in the same sequence in the coupled models".

Response: See "Major Comments", point #9. Since we are now looking at a much larger range of model results (49 20CEN realizations performed with 19 different models, as opposed to 12 20CEN realizations performed with 3 different models), we display the model results in Figures 5.3 and 5.4 in histogram form rather than as discrete points. The model results contain 49 different sequences of El Niños and La Niñas (see page 31, para. 1 of revised text). Thus the comparisons between modeled and observed trends in these two Figures do "allow for sampling uncertainty", as well as for uncertainties in model forcings and responses. We hope that this issue has been clarified by the above-mentioned changes to the text, and by our use of a much larger sample of model results.

- b. "...model simulations cannot be definitive given the exceptional nature of the 1997-98 event".

Response: The observed 1997-98 El Niño event is not that exceptional when compared with some of the simulated ENSO variability in Figures 5.2B-D.

3. Major Comment #3

"...chapter notes the importance of the stratospheric contribution to the channel 2 temperatures and refers to Fu et al...but then never allows for this in subsequent comparisons".

Response: Completed. See "Major Comments", point #2. This issue is discussed on page 27, para. 2 of the revised text.

4. Major Comment #4

- a. "...regional trends differ a lot from global values...the large increase in temperature over northern land and the smaller decrease in the troposphere...are not examined in the models and not picked up. The chapter comes closest with Figure 5.5, but that fails to account for the stratospheric contamination".

Response: We agree that it would be useful to examine, on a regional basis, the correspondence between observed surface and tropospheric temperature trends and trends simulated in 20CEN runs performed with the IPCC models. Unfortunately, such comparisons were not available for assessment in Chapter 5. Nor was it possible to perform and publish such work for the specific purpose of this Report. We note, however, that the new version of Figure 5.5 shows spatial distributions of surface-minus- T_{2LT} trends in models and observations (the previous version showed surface-minus- T_2 trends), and thus avoids the "stratospheric contamination" problem referred to by the NRC committee.

- b. "The fact that sondes are not global is also not dealt with. Subsampling of the modeling data (sic) at sonde locations is not done".

Response: Like the regional trend comparisons mentioned above, this is an issue where further research is warranted, but was unavailable at the time Chapter 5 was being finalized. We note that several practical problems arise in subsampling model output at the locations of radiosonde locations. First, coupled models with relatively coarse horizontal resolution do not resolve many of the small islands on which tropical radiosonde stations are located. Most models represent these islands as ocean rather than land. At these island locations, some of the key physical properties that may influence surface warming rates (specific heat capacity, albedo, roughness length, *etc.*) will be quite different in models and in the real world. This diminishes the usefulness of subsampling model output at radiosonde station locations, particularly for the specific purpose of examining differential warming of the surface and troposphere. Second, the two radiosonde datasets used in Chapter 5 have differences in both spatial coverage and in the number of vertical levels. Any model "subsampling exercise" would convolve differences in the horizontal and vertical coverage of the two radiosonde datasets, thus hampering interpretation of results. Third, planetary waves tend to smooth anomalies far more effectively above the boundary layer than at the surface. Reliable estimation of large-scale temperature variability therefore requires a less-dense sampling network for T_{2LT} than for T_S . This explains why time series of monthly-mean, tropically-averaged anomalies of HadAT2 (synthetic) T_{2LT} data are more highly correlated with "full" tropical averages of HadCRUT2v T_S data (based on all available tropical T_S data) than with tropical

averages based on T_s data subsampled at the locations of HadAT2 radiosondes (see *Santer et al.*, 2005). Thus subsampling may actually introduce noise that hampers reliable estimation of amplification behavior, providing further justification for our decision to rely on unsubsampled model T_{2LT} and T_s data.

5. *Major Comment #5*

“...should be more explicit discussion of the specific responses to individual forcings and how these combine together”.

Response: Experiments involving changes to a single forcing factor are very useful in estimating the climate “fingerprint” arising from a given forcing. Unfortunately, very few modeling groups perform such “single forcing” experiments. The exceptions are the PCM, GISS, and HadCM3 models, which have completed numerous single forcing experiments, often with multi-member ensembles for each forcing that is varied (see, e.g., *Tett et al.*, 2002; *Santer et al.*, 2003a; *Hansen et al.*, 2005a). Most of the IPCC 20CEN runs analyzed in Chapter 5 involve changes in a combination of natural and anthropogenic forcings (see Tables 5.2 and 5.3), and do not allow estimation of “specific responses to individual forcings”. The need for “single forcing” experiments is now articulated in the new Recommendation #2 (page 6, lines 6-12). We note that there is some discussion of the likely effects of different forcings on atmospheric temperatures. This discussion occurs in Chapter 1, in Section 3 (in describing the PCM “single forcing” results shown in Figure 5.1), and in Boxes 5.3 and 5.4.

6. *Major Comment #6*

“...committee liked the two model plots...and would hope that these can be included in a revised chapter. Also included should be as many results from additional models as time allows”.

Response: Completed. See “Major Comments”, point #1. We note that the “two model plots” referred to by the NRC committee have now been published (in modified form) in *Santer et al.* (2005). The plots are discussed in the new Section 4 (“Tropospheric Amplification of Surface Temperature Changes”).

7. *Major Comment #7*

“...discussion of the main principles behind the (fingerprint) methodology...There needs to be a better understanding of the strengths and limitations of detection and attribution analyses”.

Response: Completed. See “Major Comments”, point #4.

Specific Comments

1. *Use of word “lockstep”.*

Response: Completed. Changed to “evolve in unison” (page 7, line 6).

2. *Provide references for differences of opinion...*

Response: Completed. The sentence in question has been deleted (“There are differences of scientific opinion about the relative merits of AGCM and CGCM experiments for studying the differential warming problem”).

3. *Mention of the NAO.*

Response: Completed. We have now added the following sentence to footnote #1: "Note also that even with the specification of ocean boundary conditions, the time evolution of modes of variability that are not forced by the ocean (such as the North Atlantic Oscillation; see Chapter 1) will not be the same in the model and in the real world (except by chance)" (see page 9).

4. *Expand sentence in lines 112-113.*

Response: We do not feel that expansion is necessary, particularly in view of the rather length footnote #2.

5. *Reason for using ensemble forecasts...*

Response: No action necessary.

6. *Add regional aspects in footnote 11?*

Response: Completed. The final sentence of the footnote now reads: "Even forcings with "low" or "very low" LOSU may have had significant climatic impacts at regional and even global scales" (page 12).

7. *Evidence of the 0.3°C cooling over India should be provided...*

Response: Completed. The discussion of the possible effects of black carbon aerosols on surface and tropospheric temperatures has been shifted to Box 5.3 (see "Major Comments", point #5). Box 5.3 no longer includes the sentence on the estimated cooling of surface temperatures over the Indian subcontinent (page 43).

8. *Mention of urban heat island effect.*

Response: The discussion of the possible effects of LULC changes on surface and tropospheric temperatures has been shifted to Box 5.4 (see "Major Comments", point #5). We have now included a specific reference to Chapter 4, where urban heat island effects are discussed. The first sentence of Box 5.4 now reads: "Humans have transformed the surface of the planet through such activities as conversion of forest to cropland, urbanization, irrigation, and large water diversion projects (see Chapter 4)". We think that Chapter 4 is the more logical place for treatment of urban heat island effects.

9. *Reference to Matthews et al.*

Response: Completed. There are two Matthews *et al.* papers (2003 and 2004). We now cite both papers. We have also added the following sentence to the end of Box 5.4: "Larger regional trends of either sign are likely to be evident (*e.g.*, Hansen *et al.*, 2005a)" (page 44; lines 21-22). To address the "signal-to-noise" issue raised by the NRC comment, we added footnote #78: "Note that larger regional trends do not necessarily translate to enhanced detectability. Although the signals of LULC and other spatially-heterogeneous forcings are likely to be larger regionally than globally, the "noise" of natural climate variability is also larger at smaller spatial scales. It is not obvious *a priori*, therefore, how signal-to-noise relationships (and detectability of a given forcing's climate effects) behave as one moves from global to continental to regional scales.

10. *Reference to Jones (1994).*

Response: Completed. This reference has been added (page 15, line 5). Footnote #15 now mentions the value of the residual T_{2LT} trend estimated by Jones (1994).

11. *Are the very small error bars in footnote #18 still believed?*

Response: No. This should be clear from the statements made on page 15, lines 14-21.

12. *Slow changes in lapse rate variability.*

Response: Completed. The sentence has been changed to: "The implication is that volcanic effects probably contribute to slow changes in observed lapse rates" (page 17, lines 1-2).

13. *Footnote 21.*

Response: Completed. This footnote (now #18) reads as follows: "The latter results were obtained with the HadCRUT2v surface data (Jones *et al.*, 2001) and version d03 of the UAH T_{2LT} data.

14. *Section 4.3 would benefit from more synthesis and assessment.*

Response: Completed. In the reviewed version of Chapter 5, Section 4.3 had four paragraphs. Paragraphs 2 and 3 have now been concatenated and modified. This helps to bring out common aspects of the findings of the Gaffen *et al.* (2000), Hansen *et al.* (1995) and Santer *et al.* (2000) papers.

15. *Which climate forcings does line 357 refer to?*

Response: Completed. Gaffen *et al.* (2000) concluded that "...the richer three-dimensional structure of natural and anthropogenic climate forcings may be required for more realistic simulations (of the vertical structure of atmospheric temperature trends". They were not more specific about the external forcings. To avoid any confusion, our description of the Gaffen *et al.* results does not mention external forcings, and now simply states that: "Model-based estimates of natural climate variability could not explain the observed tropical lapse rate changes over 1979 to 1997" (page 18, lines 12-13).

16. *Change "various datasets" to "various models".*

Response: Completed. There have been major changes to the discussion of fingerprinting (see "Major Comments", point #4). The sentence referred to by the comment no longer exists.

17. *IDAG reference missing.*

Response: Completed. This has been added (page 52, lines 19-21).

18. *Are there also different variables than temperature?*

Response: Yes. These are discussed in Section 6. Formal fingerprint detection work with sea-level pressure changes is cited in Section 6 (e.g., Gillett *et al.*, 2003).

19. *Positive detection results obtained in absence of some forcing.*

Response: Completed. Most of the original paragraph has been deleted. The new paragraph is on page 21, lines 9-17.

20. No "apparent contradiction".

Response: Completed. See point #19 above.

21. Relevance of final sentence in lines 498-500?

Response: Completed. We added the following text in the paragraph immediately after this sentence: "This yielded periods of agreement and periods of disagreement between the (fixed) aerosol fingerprint and the time-varying effect of aerosols on atmospheric temperatures" (page 24, lines 15-17). The relevance of the final sentence in lines 498-500 should now be clearer.

22. Degeneracy between the sulfate aerosol and greenhouse gas patterns.

Response: Completed. We do not think that degeneracy between the GHG and sulfate aerosol fingerprints is the explanation for the difficulty in detecting the sulfate response. Sulfate aerosol forcing has temporal and regional structure quite different from that of GHG forcing. Climate responses are likely to reflect these forcing differences, particularly in terms of the latitude-height patterns of temperature change. The key point here is that "space-time" detection schemes, which explicitly account for the complex spatial and temporal structure of the sulfate aerosol fingerprint, have successfully identified sulfate aerosol effects in observational data (see page 24, lines 23-28). We have made some changes to the text (particularly on page 24, lines 12-21), which should help to clarify this issue. The degeneracy problem is now discussed in the new Box 5.5 (page 46, lines 15-16, and footnote #83).

23. None of the model runs have been written up yet.

Response: This is not the case. A number of the IPCC 20CEN models and experiments have been described in the peer-reviewed literature. Some of the relevant publications have already appeared (e.g., Hansen et al., 2005a; Washington et al., 2000; Tett et al., 2002). Others are in press (e.g., Collins et al., 2005; Meehl et al., 2005; Delworth et al., 2005). One published paper (Santer et al., 2005) provides a brief introduction to the IPCC 20CEN runs discussed in Section 5. All of these references are given in Chapter 5. Note also that the revised version of Chapter 5 attempts to summarize (in Tables 5.2 and 5.3) the natural and anthropogenic forcings used in the 20CEN integrations (see "Major Changes to Chapter 5", point #7). The caption for Table 5.2 directs readers to PCM DI's website (http://www-pcmdi.llnl.gov/ipcc/model_documentation), where further model and experimental details are available.

24. HadCM3 has also run all the experiments.

Response: Completed. Results from both HadCM3 and HadGEM1 are now included in Section 5 (see "Major Comments", point #1).

25. Use of different forcings in the IPCC models sometimes an advantage.

Response: Completed. The introductory paragraphs of Section 5 have been modified to reflect the inclusion and analysis of a larger number of 20CEN runs. The revised text includes the following sentence: "Note, however, that the lack of a coordinated experimental design is also an advantage, since the "ensemble of opportunity" spans a wide range of uncertainty in current estimates of climate forcings" (page 26, lines 11-14).

26. Insert "partly" before "due".

Response: Completed. We have made some small but significant changes to the text. We now state that "The most likely explanation for this discrepancy is a residual cooling trend in the radiosonde data (Chapter 4). The neglect of stratospheric water vapor increases in most of the 20CEN runs considered here (*Shine et al.*, 2003) may be another contributory factor" (page 28, lines 15-18). We have also added the following footnote (#47): "Recent work suggests that this residual trend is largest in the lower stratosphere and upper troposphere, and is related to temporal changes in the solar heating of the temperature sensors carried by radiosondes (and failure to properly correct for this effect; see *Sherwood et al.*, 2005; *Randel and Wu*, 2005)". Footnote #48 has been expanded, and now points out that "To our knowledge, CH₄-induced stratospheric water vapor increases were explicitly incorporated in only two of the 19 models considered here (GISS-EH and GISS-ER; *Hansen et al.*, 2005a)".

27. Insert "Body" before "temperature".

Response: Completed. See page 37, line 12.

28. Add reference to *Robock and Oppenheimer (2003)*.

Response: Completed. See page 37, line 21.

29. Reference should be to *Clausius and Clapeyron*.

Response: We think that *Hess (1958)* is a reasonable primary reference. We have used this instead of *Clausius and Clapeyron* (see page 38, lines 12-13).

30. *Ocean temperature data has ambiguous implications.*

Response: We respectfully disagree. Recent papers by *Barnett et al. (2005)* and *Pierce et al. (2005)*, which are now cited in the revised version of Chapter 5 (page 39, line 24) show that two of the models discussed in Sections 5 and 6 (PCM and HadCM3), when forced by anthropogenic effects, successfully capture quite complex features of the "observed" vertical structure of ocean heat uptake (as portrayed in *Levitus et al.*, 2005) in a number of different ocean basins. These studies properly account for the sparse coverage of ocean temperature measurements in the *Levitus et al.* dataset, and subsample ocean model output at the locations where observations are actually available. Other studies relying on "infilled", spatially-complete *Levitus et al.* data (in which temperatures over large volumes of the world ocean have been infilled statistically, or with climatological means) have suggested that there is a significant mismatch between modeled and observed estimates of the variability of ocean heat content. This discrepancy is markedly reduced by subsampling model output at the location of actual temperature measurements (*Gregory et al.*, 2004; *AchutaRao et al.*, 2005). In our view, many of the "ambiguous implications" of modeled and observed ocean heat content comparisons have been clarified by this more recent work, which is now cited in the revised Section 6. We think we have been suitably cautious in describing the ocean heat content research (see, e.g., page 39, lines 10-17).

31. *More discussion about widespread and accelerating glacial retreat.*

Response: Completed. We have added a new footnote (#73). This reads as follows: "*Folland et al. (2001)* note that "Long-term monitoring of glacier extent provides abundant evidence that tropical glaciers are receding at an increasing rate in all tropical mountain areas". Accelerated retreat of high-elevation tropical glaciers is occurring within the

tropical lower tropospheric layer that is a primary focus of this report, and provides circumstantial support for warming of this layer over the satellite era .

32. Discuss response of the model to each individual forcing.

Response: See "Major Comments", point #5.

33. Volcanoes and ENSO do not make much difference to the trend.

Response: Completed. The section that this comment pertains to has been moved to the "Key Findings and Recommendations" ("KF&R"; see "Major Comments", point #6). The KF&R section no longer discusses the impact of ENSO and volcanoes on surface and tropospheric temperature trends. There is no "Key Finding" related to ENSO effects. There are, however, two "Key Findings" relevant to volcanic effects on atmospheric temperature:

- 1) "Large volcanic eruptions cause cooling of the surface and troposphere (over 3 to 5 years) and warming of the stratosphere (over 1 to 2 years)"
- 2) "The longer-term (multi-decadal) climatic effects of volcanic eruptions and solar irradiance changes are identifiable in some fingerprint studies, but results are sensitive to analysis details .

These are located on page 2, lines 24-28, of the revised text.

34. More explicit statements about which forcings contribute to agreement.

Response: See "Major Comments", point #5. We believe that the KF&R (see page 2, lines 14-19) are now explicit regarding the forcings whose climate fingerprints have been robustly identified in observational temperature data. The new Table 5.2 provides information on the forcings applied in the 20CEN runs analyzed in Section 5, and thus satisfies the NRC request to "be more explicit about exactly which forcings are included in the "all" integration .

Chapter 6

Chapter 6 asks: what measures can be taken to improve the understanding of observed changes? This chapter purports to respond to issues and shortcomings raised in Chapters 1-5 and develops a list of seven comprehensive recommendations. These recommendations address: (1) the need for improved observing standards that are rigorously implemented, (2) better use of existing data, (3) expanded use of regional and global climate models for assessing the impacts of forcings and feedbacks on temperature trends, (4) continued assessment of tropospheric trends using a full range of statistical techniques and modeling tools, (5) enhanced development of reanalyses, (6) improved metadata, and (7) development of scientific talent.

The committee finds that the recommendations in Chapter 6 are insufficiently specific and not clearly prioritized. Furthermore, the seven recommendations seem largely disconnected from the findings in Chapters 1-5, and even from the text in Chapter 6. This chapter needs a substantial rewrite, including re-organization of the text and reformulation of the recommendations.

Major Comments

1. *Chapter 6 should be reorganized into two parts:*

- a. *The first part should take findings from Chapters 1-5 to recommend specific opportunities to improve understanding of vertical temperature trends. These should focus on addressing remaining uncertainties in existing satellite and radiosonde data sets.*

Response: Done. Sections 1-5 refer. Text and recommendations are linked back to the supporting Chapters and focused on the key questions needing to be resolved

- b. *The second part should focus on future measurement opportunities in the context of the specific goals of the report for reconciling observations and understanding of temperature trends.*

Response: Done. Section 6 refers. Two key recommendations are in this final section which is on future climate monitoring. Section 6 thus differs in emphasis from Sections 1-5 but is not presented as a separate “part” and also links back to previous Chapters.

2. *Also in reorganizing Chapter 6, the committee recommends starting with the Global Climate Observing System (GCOS) implementation plan and reinforcing, adding to, or modifying that plan, rather than starting from scratch. It is important that the community speak with a unified voice as much as possible. The authors should also discuss the current efforts to improve the relevant temperature measurements in addition to GCOS, including Global Ocean Data Assimilation Experiment (GODAE), SEAFLUX, and Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel efforts to develop technologies for reference radiosondes, and discuss international efforts, not just U.S. efforts.*

Response: The GCOS Implementation Plan is referenced where it is the basis of the Recommendations and listed with other key documents in the opening Background section. GODAE is cited in Section 6 but SEAFLUX was deemed irrelevant. The work of the GEWEX Radiation Panel on radiosondes has been superseded by new WMO comparisons (and by a recent NOAA/GCOS Workshop); the former has been referenced, but only in a footnote as Chapter 4 contains lengthier discussion of the quality of radiosonde data.

3. *The organization of the chapter is centered around data types, such as “surface”, “tropospheric”, and “reanalyses”. A variety of new issues that do not directly map to the seven recommendations are brought up in these sections (much of which is not relevant or belongs in previous chapters). An alternate organization would be to have seven sections with section headings that are the first sentence of each recommendation. Then the text of each section would tie directly back to a need documented in the earlier chapters, would include discussion of the adequacy of current national and international plans to address this need, and make further specific recommendations for implementation of this recommendation.*

Response: The reorganization of Chapter 6 has been done in a similar way as recommended, but not exactly so. The Chapter is now organized around actions, not data types. The text is significantly shorter with frequent links to the previous Chapters. See response to (1a, b) above.

4. *A substantial amount of new information is introduced for the first time in Chapter 6, including material that should have been introduced in earlier chapters if it is deemed relevant and material that does not directly map to the seven recommendations. The following is specific information that is redundant or should be moved to previous chapters:*

- *The material in lines 54-71 should be mentioned in the context of Chapters 1 and 5.*

Response: This text is gone. Elements are in the new action-orientated Chapter 6 text but regional temperature trends are somewhat less emphasized as the global or tropical mean trends remain a key issue. However regional trends are important and implicit in the new Section 1 and Section 3.

- *Text on snow and sea ice and sampling inadequacies in lines 179-187 should be moved to Chapter 2.*

Response: Text moved

- *For lines 138-177, lines 240-254, and lines 300-317, text on combining surface temperature and dew point temperature is far too wordy, and the main point is lost. This concept should be included in a general recommendation on the need to evaluate and interpret the temperature data in the context of other data sets (e.g., humidity, winds, ocean heat content, etc.) and to understand issues such as the impact of changing land use on temperature trends, as stated in lines 347-355.*

Response: A much more concise form of these ideas is in the new Section 3, Multivariate Analyses

- *The text in lines 447-494 about recommending specific improved climate model parameterizations is not directly relevant to the present study, although it is appropriate to state in earlier chapters that inadequate parameterizations in numerical weather prediction models contribute to potential problems in using the reanalyses to determine temperature trends.*

Response: Text removed from Chapter 6.

- *The text in lines 98-105 should be moved to Chapters 2 and 4 as these points were not adequately made in those chapters.*

Response: Text moved.

5. *As far as the current recommendations in Chapter 6 still appear after the chapter is revised, here are comments on each of the current recommendations. The seven recommendations in Chapter 6 have been said numerous times before in other reports. Also, given the relative lack of traceability of these recommendations to the previous five chapters, it may be that a significant recommendation was omitted.*

a. *The first recommendation concerns reference measurements. The recommendation should be formulated to account for the adequacy or inadequacy of current national and international plans to address this need. If inadequate recommendations are made in previous documents (e.g., GCOS), then very specific recommendations should be made to address the sensor design, sampling, or other needs.*

Response: The old Recommendation 1 mixes an important aspect of the GCOS Climate Monitoring Principles contained in the GCOS Implementation Plan that is relevant to all measurements with recommendations on Reference Measurements. The reference measurements recommendation is the second Recommendation in the new Section 6, and wider GCOS monitoring principles aspects are in the first Recommendation of Section 6.

b. *The second recommendation concerns making better use of existing data. See comments for the first recommendation. This section should focus specifically on reprocessing of radiosonde data, resolving the differences between the different MSU analyses, and use of the TIROS Operational Vertical Sounder (TOVS) data, including some very specific recommendations to address the key issues. It should also discuss detailed intercomparison (at the pixel level) of the different data sets and cross checking with other variables. Better scientific uncertainty analysis of the data sets should be part of this recommendation. Specific recommendations here would add considerable value to this document.*

Response: This Recommendation has been split between the Recommendations of new Sections 1, 2 and 3 though TOVS is not explicitly mentioned. Thus multivariate cross checking aspects are in the Recommendation of new Section 3. Uncertainty analyses and detailed intercomparisons are included in the new Recommendations.

c. *The third recommendation concerns the use of climate models to interpret the cause of temperature trends. This recommendation needs to be reformulated or perhaps eliminated because it is too broad and inappropriate for the present study. What is recommended here should follow directly from Chapter 5 and any uncertainties or inconsistencies in the analyses that were identified. An alternative recommendation would be to "Improve the scientific understanding of the variations of the vertical temperature structure of the atmosphere". It should also be clearly emphasized that data is being used to test models and not vice-versa.*

Response: Recommendation reformulated substantially to reflect the emphasis of the revised Chapter 5. This is now the Recommendation of the new Section 5.

d. *The fourth recommendation concerns statistical trend analysis. A clear case has not been made in the previous chapters (or in Chapter 6) that there is a need for new research in the statistical analysis of trends. Rather, the committee would prefer that the report give explicit discussion to existing methods for dealing with such issues as autoregressive behavior and nonlinearities in trends, as already discussed in review comments on Chapter 2.*

Response: The new statistical Appendix covers the issues of calculating trends well. So we agree with the reviewers that the topic does not need sufficient further work for it to merit a recommendation in Chapter 6.

- e. *The fifth recommendation concerns climate quality reanalyses. Just as for the third recommendation, this one needs to be reformulated or perhaps eliminated. It is not useful to state such a broad recommendation that has already been made in other contexts. If there are any specific recommendations that would help address the temperature trend problem, then they should be formulated. Possibilities would include careful documentation about what assimilation data is actually assimilated into the model as a function of space and time, data assimilation experiments, etc.*

Response: We have reformulated this in Section 4 with specific guidance on data inputs. We feel it is an important topic for the Report.

- f. *The sixth recommendation concerns metadata. It seems that this issue is (or easily could be) covered in the first recommendation. It is not clear that accessibility of the data is a major issue.*

Response: Metadata are now explicitly included in the Recommendations of Section 2 on better using existing data. Availability of metadata and data is, we feel, still a sufficient issue to be included in the Recommendation. They are also implicitly included in the first Recommendation of Section 6 for future data through the full implementation of the GCOS Climate Monitoring Principles

- g. *The seventh recommendation concerns education. This recommendation is very diffuse and is not motivated by the previous chapters. It is hard to disagree with the statement that education in our field should include a stronger emphasis on the proper use of statistics and error analysis. However, this point could easily be incorporated into the second recommendation.*

Response: This recommendation has been removed.

- h. *An outstanding omission in terms of recommendations is the need for better methods to sense temperature or related variables from satellites, such as using instruments that are self calibrating, sounders with more channels for better vertical resolution, and the use of proxy measures such as refractive index and spectral TOA radiance.*

Response: The section on future climate monitoring attempts to remedy this omission. We don't make an explicit recommendation for specific new sensors here as work needs to be done to evaluate those listed in the context of the Report.

Specific Comments

1. *In lines 79-80, it is the committee's understanding that the U.S. Climate Reference Network been shelved or at least stalled.*

Response: This topic has been removed. (N.B. The network is running, but on a reduced budget which limits new deployments (T Peterson, pers. Comm.).)

2. *Most countries do not know about the GCOS Monitoring Principles, mentioned in lines 87-88.*

Response: This particular text has gone but this Report is mainly aimed at people in the USA who do know about the Principles. The GCOS Climate Monitoring Principles have been ratified by WMO Congress, so at least the heads of meteorological services worldwide know about them and have approved them and they are now part of GEOSS. The Principles are better referenced throughout the latest drafts of the Report and in Chapter 6.

3. *Many of the recommendations in lines 150-160 may be difficult to achieve based on cost considerations. The GCOS aim is to get the data first, then work on metadata. Getting pictures of sites will only be useful if they are taken at regular intervals.*

Response: We do not fully agree. Most countries abide by WMO guidelines in having regular station-inspection reports. It would nowadays cost little for digital photographs to be taken during routine inspections to develop a regular sequence.

4. *In lines 198-199, there is a GCOS working group of the Ocean Observing Panel for Climate (OOPC) and the Atmospheric Observing Panel for Climate (AOPC) looking at Sea Surface Temperature biases.*

Response: Two of the lead authors are members of these Panels and the convening lead author advises the AOPC, but we do not feel the need to mention the Panels explicitly.

5. *In line 287, locating the reference sonde stations for comparison with satellite overpasses requires observations at different times at each station. Thus, the CCSP authors may want to reconsider this recommendation.*

Response: Proper comparisons between radiosondes and satellites does require simultaneity. Otherwise uncertainty will persist, even given the use of an operational or reanalysis model to try to compensate for time-differences. So we maintain our original stance in the second Recommendation of Section 6.

6. *Better use of statistics is needed in lines 366-369.*

Response: This text has gone and is adequately covered in the new Appendix.

7. *In lines 402-407, there are at least two comments on Kalnay and Cai (2003) and there should also be a reference to Simmons et al. (2004). Several criticisms of the Kalnay and Cai approach have been identified.*

Response: This text has been removed as part of the shortening and refocusing of the Chapter. Only a brief discussion of the shortcomings of current reanalyses for climate-trend applications remains in Section 4, where the Simmons et al (2004) reference is listed.

8. *The recommendations for “tightly constraining” the dataset for reanalyses in line 425 is not possible or wise owing to continual changes in all observations, including sondes.*

Response: These words have also been removed. Section 4 and its Recommendation presents a strategy for climate-quality reanalyses. This is based on a paper the Convening Lead author is co-authoring, led by Lennart Bengtsson, which we hope to reference after the Public Review.