S.Gurney

https://www.edockets.state.mn.us/EFiling/edockets/searchDocuments.do?method=showPoup&documentId={3D3CEE85-9523-469E-B470-D8426DA1C6C5}&documentTitle=20159-113909-03

But see also the page of corrections/clarifications in
https://www.edockets.state.mn.us/EFiling/edockets/searchDocuments.do?method=showPoup&documentId={1DF404F6-A69C-437E-B279-35C30B463A19}&documentTitle=20159-114165-03

SURREBUTTAL TO: Happer, Lindzen, Bezdek

pp.3-15 PEABODY WITNESSES' USE OF FLAWED ARGUMENTATION

p.4 "To audiences outside the climate science community, the arguments presented by skeptics of ACC appear legitimate. Hence, in addition to presenting the technical evidence for ACC and refuting the individual technical challenges, it is critical to explain how these argument patterns, in and of themselves, are logically flawed. This is important because it does not require an audience to navigate the back-and-forth of the technical merits, which can often be extremely difficult due to both the volume of information and its technical nature."

The ALJ quoted extensively from these pages in her *Findings*, so they are not repeated, but people are urged to read the entire commentary, as it well-characterizes the Peabody testimonies.

p.16 "Furthermore, Dr. Happer has **limited expertise in the subject of climate science or economics**, as he has published no peer-reviewed papers in climate science or economics 6 (see http://physics.princeton.edu/atomic/happer/Publications.html) and has performed **no research related to climate modeling, the carbon cycle, or temperature measurements - all topics on which he has provided testimony.** He appears to lack the qualifications to opine on the expertise of persons engaged in areas of study of which he himself has limited knowledge."

.

pp.22-36 IPCC WGIAR4 pp.185-201

STATUTE 216B.2422, SUBDIVISION 3

SURREBUTTAL TESTIMONY AND ATTACHMENT OF KEVIN GURNEY

ON BEHALF OF

THE DIVISION OF ENERGY RESOURCES OF THE

MINNESOTA DEPARTMENT OF COMMERCE

AND

MINNESOTA POLLUTION CONTROL AGENCY

SEPTEMBER 10, 2015

SURREBUTTAL TESTIMONY AND ATTACHMENT OF KEVIN GURNEY
IN THE MATTER OF THE FURTHER INVESTIGATION INTO ENVIRONMENTAL AND SOCIOECONOMIC COSTS UNDER MINNESOTA STATUTE 216B.2422, SUBDIVISION 3

DOCKET NO. E999/CI-14-643
OAH DOCKET NO. 80-2500-31888

TABLE OF CONTENTS

I. PRELIMINARIES

Q. Please state your name and affiliation.

A. My name is Kevin Gurney. I am a consultant providing testimony at the request of the Minnesota Department of Commerce (Department or DOC), and Minnesota Pollution Control Agency (MPCA) (together, the Agencies.)

Q. Have you previously filed testimony in this case?

A. Yes. I filed Rebuttal Testimony on August 12, 2015.

Q. What is the purpose of your Surrebuttal Testimony?

A. I respond to the Rebuttal Testimonies of several witnesses retained by Peabody Energy Corp. ("Peabody")

First, however, I will explain a problem that permeated the testimony of certain Peabody witnesses, which was the persistent use of patterns of argumentation and reasoning that were misleading, biased or otherwise flawed.

II. PEABODY WITNESSES' USE OF FLAWED ARGUMENTATION

Q. What do you mean, when you say that certain Peabody witnesses' testimony persistently used patterns of argument and reasoning that were misleading, biased or otherwise flawed?

A. These Peabody witnesses used a series of argument patterns throughout their testimony that I have seen repeatedly over the last 30 years from the community I would call "skeptics" of anthropogenic climate change (ACC). These argument patterns reflect biased or flawed reasoning that I describe in my surrebuttal testimony below.

Q. Why is it important to understand these argument patterns?

A. To audiences outside the climate science community, the arguments presented by skeptics of ACC appear legitimate. Hence, in addition to presenting the technical evidence for ACC and refuting the individual technical challenges, it is critical to explain how these argument patterns, in and of themselves, are logically flawed. This is important because it does not require an audience to navigate the back-and-forth of the technical merits, which can often be extremely difficult due to both the volume of information and its technical nature.

Q. What are the argument patterns that you have seen throughout Peabody witnesses' testimony, and how do they result in incomplete or misleading assessment of the evidence for ACC?

A. The argument patterns can be classified into four categories. The first pattern involves the witnesses' use of selective citation.

A. SELECTIVE CITATION

Q. What is selective citation?

A. Selective citation is commonly referred to as "error by omission" or "cherry-picking" the information available to support a predisposed conclusion. This has been relied upon to a great extent in discussion of ACC. The selective citation pattern has two variations, both of which are used by the Peabody witnesses in their direct and rebuttal testimony. I will present these two variations and identify the instances when I found each of the two variations within the direct and rebuttal testimony of the Peabody witnesses.

Q. What is the first variation of selective citation?

A. The first variation is called "non-peer-review." This refers to the reliance upon non-peer-reviewed literature when presenting evidence that apparently contradicts the theory and evidence for ACC. As explained in DOC Ex.___ at 24-25 (Gurney Rebuttal), scientific assessment of any topic must rely on peer-reviewed academic literature. Arguments that rely on non-peer-reviewed literature are considered unreliable and potentially biased. In my opinion, heavy reliance on non-peer-reviewed literature is typically met with suspicion by the scientific community and often considered a deliberate attempt to obfuscate mainstream scientific thought.

Q. Do Peabody witnesses rely on non-peer-reviewed literature?

A. Yes. Large portions of the testimony of Dr. Bezdek relied almost completely on non-peer-reviewed literature. For example, in Peabody Ex. ___ at RHB-1, lines 97-136 (Bezdek Rebuttal), where Dr. Bezdek rebutted Dr. Polasky, Dr. Bezdek provided misleading testimony, offering to the Administrative Law Judge (ALJ) and Minnesota Public Utilities Commission (Commission) what he claimed was refutation of ACC in "peer-reviewed international scientific journals…" He then supplied purportedly supportive examples; however, nine of the thirteen examples he identified are not peer-reviewed papers but are a mixture of opinion pieces, institute reports, and online blog content. Of the three peer-reviewed papers, one was in the "Forum" portion of the peer-reviewed journal (*Bull. Am. Meteorological Soc.*), a section intended for opinion pieces. The remaining two papers were authored or co-authored by fellow Peabody witnesses (Drs. Tol and Lindzen). *None* of the thirteen papers listed were from either *Science* or *Nature* (two of the highest regarded journals in science) even though Dr. Bezdek claimed in his sworn testimony that refutation of ACC had appeared in these two important journals ((Peabody Ex. ___ at RHB-1, line 92 (Bezdek Rebuttal)).

When taking up the topic of scientific consensus in his Rebuttal Testimony (Peabody Ex.___ at RHB-1, lines 213-258 (Bezdek Rebuttal)), Dr. Bezdek cited 11 papers. Only 2 of these are peer-reviewed papers, and those happen to be the papers that he was countering in this section of his testimony. *None* of the 9 papers used to support his rebuttal position are peer-reviewed.

Dr. Bezdek purported to provide "empirical evidence" that counters ACC (Peabody Ex. ___ at RHB-1, line 168 (Bezdek Rebuttal)). That so-called empirical evidence consisted of an unreferenced Figure (Peabody Ex. ___ at RHB-1, lines 178-179 (Bezdek Rebuttal)) and congressional testimony (Peabody Ex. ___ at RHB-1, line 182 (Bezdek Rebuttal)) rather than peer-reviewed research.

Similarly, when Dr. Bezdek asserted that there is a "divergence between observations and climate model projections.." (Peabody Ex. ___ at RHB-1, line 190 (Bezdek Rebuttal)), the support for this assertion relies on a single instance of congressional testimony rather than peer-reviewed literature.

Finally, as noted in DOC Ex.___ at 25 (Gurney Rebuttal), out of the 54 endnotes to the entirety of the Direct Testimony of Dr. Bezdek, 52 were non-peer-reviewed literature citations.

Q. Are there other examples of the Peabody witnesses relying on non-peer-reviewed literature?

A. Yes. In the rebuttal testimony of Dr. Lindzen (Peabody Ex.___ at RSL-1, lines 51-59 (Lindzen Rebuttal)) he listed elements of his critique of the recent paper by Karl et al. (2015). Dr. Lindzen cited a non-peer-reviewed Cato Institute report and "numerous others," the latter of which are not referenced.

Q. What is the second variation of the "selective citation" type of argumentation?

A. The second variation of selective citation is "narrow citation." This is a type of argumentation where the witness cites peer-reviewed literature to support arguments but only a very narrow slice of the work on a topic is used. Examples of the narrow citation approach are where the witness relies only on those papers that support the witness's predisposed position, or cites papers that were later refuted without including the refutation literature, or cites papers without the context that would demonstrate their limited utility. Examples of narrow citation often show a preponderance of self-authored papers or a form of "circular" citation where a small group of ACC skeptics refer to a small set of papers authored by these same ACC skeptics giving the impression that there is a large body of literature when, in reality, the number of papers is small relative to a more comprehensive treatment of the literature on a subject.

Q. Why does narrow citation diminish the reliability of scientific claims and testimony?

A. As explained in my Rebuttal Testimony (DOC Ex.___ at 26-27 (Gurney Rebuttal)), in order to accurately assess the fact of a scientific topic and produce a reliable analysis, *all* the peer-reviewed literature on the scientific topic must be included, assessed and synthesized. Because the role of peer-reviewed publication is aimed at extending the boundaries of what is known, there is often a spectrum of evidence on any given topic. Hence, comprehensive assessment is an absolute necessity in order to arrive at a reasonable understanding of a topic at hand. This is one of the goals of the Intergovernmental Panel on Climate Change (IPCC) in forming and generating the series of multivolume assessment reports since the 1990s. The workforce to produce the assessments is entirely voluntary and comprised of scientists with specific expertise in the many sub-topics covered in the assessment reports.

Q. Did the Peabody witnesses use narrow citation?

A. Yes. The Peabody witnesses relied heavily on the narrow citation approach, as is seen in the testimony of Drs. Lindzen, Bezdek, Spencer and Happer.

Q. Can you provide examples?

A. Yes. Narrow citation was employed by Drs. Lindzen and Spencer when they discussed the topic of climate sensitivity. Peabody Ex.___ at RSL-2, pages 11-12 (Lindzen Rebuttal), Peabody Ex.___ at RSL-2, lines 447-475 (Lindzen Direct), Peabody Ex.___ at RWS-2, pages 5-6 (Spencer Direct), and Peabody Ex.___ at RWS-1, pages 22-23 (Spencer Rebuttal). CEO witness Dr. Abraham accurately noted that missing from Drs. Lindzen's and Spencer's presentations of an assessment of model climate sensitivity, were a series of peer-reviewed papers that directly refuted those cited by Drs. Lindzen and Spencer. CEO Ex.___ at 25-26 (Abraham Rebuttal).

Q. What do you conclude from Drs. Lindzen's and Spencer's reliance on the narrow citation approach?

A. Drs. Lindzen's and Spencer's testimony on the topic of climate sensitivity is not reliable because an objective, reliable assessment cannot be gleaned from testimony that narrowly cites one's own peer-reviewed work without citing or discussing peer-reviewed papers that directly refute that same work.

Q. Are there other examples of the Peabody witnesses relying on the narrow citation approach?

A Yes. Peabody Ex.____ at RWS-2, pages 21-22 (Spencer Rebuttal) provided another example of narrow citation. There Dr. Spencer testified that current surface temperature measurements have long-term biases due to urbanization, and he offered four references to support his testimony.

- His first reference was from a 1973 book that identified the problem associated with urbanization and other biases in surface temperature measurements. This citation was not relevant, however, to the question of the reliability of *current* temperature records because, during recent years, extensive effort has gone into corrections for urbanization effects, corrections that have been documented.
- Dr. Spencer's second reference was a paper on temperature records in New Zealand, in which the authors correct for "shelter-contaminated trends" and find a New Zealand warming trend of +0.28 ºC/century versus an uncorrected New Zealand trend of 0.91 ºC/century. New Zealand represents less than 0.2% of the land surface of the planet. There are numerous studies that have made

de Freitas et al(2014)
http://hot-topic.co.nz/nz-cranks-finally-publish-an-nz-temperature-series-but-their-papers-stuffed-with-errors
hot-topic.co.nz/danger-dedekind-heartbreak-ahead-still-wrong-still-digging-nz-still-warming-fast/

- adjustments for those stations potentially influenced by urbanization, but these are not cited by Dr. Spencer.
- The third paper Dr. Spencer referenced had no content relating to urbanization and temperature trends.
- The fourth paper identified the impact of urbanization on temperature trends in the urbanized portion of China. There are numerous studies that have made adjustments for those stations potentially influenced by urbanization, but these are not cited by Dr. Spencer.

This very narrow collection of papers demonstrates that there is a concern regarding the influence of urbanization on long-term surface temperature measurements. However, Dr. Spencer failed to account for, or disclose to the ALJ and Commission, the numerous papers and review efforts that have both developed techniques to correct for the effects of urbanization, or the results of those corrections. Most importantly, the results of these urbanization corrections have had little impact on the large-scale warming trends reviewed in the IPCC assessment reports. This missed literature and its results were well presented in CEO Ex.___ at 13-14 (Dessler Rebuttal) and CEO Ex.___ at 23-24 (Abraham Rebuttal).

This passage from the Spencer Rebuttal is also an example of "straw man argumentation" (argument pattern number 3 presented below). In this instance, the relevant question is not whether urbanization effects exist (the community that collects and analyzes long-term surface temperature records are certainly aware of this), but whether they have been adequately accounted for and the impact of those corrections on the analysis.

B. MISUNDERSTANDING OF SCIENCE OR CITED LITERATURE

Q. What was the second pattern of argument and reasoning in the testimony of the Peabody witnesses that was misleading, biased or otherwise flawed?

A. The second pattern was a misunderstanding of the science or cited literature. Though less common in proceedings of a serious nature, a misunderstanding of science or cited literature can, and does, occur.

Q. Dr. Gurney, can you provide examples of this error in the testimony of Peabody witnesses?

A. Yes. In Peabody Ex.___ at RHB-1, page 7, citation 15 (Bezdek Rebuttal), Dr. Bezdek cited a study authored by Strengers et al., 2015 in an effort to support his questioning of the consensus on ACC. Dr. Bezdek testified:

> The most recent study finds that less than half (43 percent) of climate scientists who research the topic and for the most part publish in the peer-reviewed literature agree with the IPCC's main conclusion that CO_2 is the dominant driver of climate change.

In reading through this Strengers et al. study, I can find no such statement or numerical result consistent with Dr. Bezdek's claim. It appears that Dr. Bezdek has combined the results of two separate questions, multiplying the percentage results of the two separate questions to arrive at the 43 percent value. This is incorrect. The only way to achieve an accurate assessment of the survey response is to ask the complete question to those being surveyed. Combining the results, as Dr. Bezdek has, represents flawed reasoning and would violate survey protocol. Indeed, the conclusions of the Strengers report, and more importantly, the peer-reviewed paper published based on the survey in this report (not cited or otherwise disclosed to the

ALJ and Commission by Dr. Bezdek: Verheggen et al., *Env. Sci. & Tech.*, 48, pp 8963-8971, 2014) came to the *opposite conclusion*. To quote the results (presented in the abstract) of the peer-reviewed paper:

> Consistent with other research, we found that, as the level of expertise in climate science grew, so too did the level of agreement on anthropogenic causation. 90% of respondents with more than 10 climate-related peer-reviewed publications (about half of all respondents), explicitly agreed with anthropogenic greenhouse gases (GHGs) being the dominant driver of recent global warming.

Q. Are there other examples of the second pattern - of a misunderstanding of the science or cited literature - by the Peabody witnesses?

A. Yes. In citation number 19 in Peabody Ex.___ at RHB-1, page 7 (Bezdek Rebuttal), Dr. Bezdek testified:

> A survey by the American Meteorological Society (AMS) found that only 25 percent of respondents agreed with the UN IPCC claims that humans are primarily responsible for recent warming.

I can find no such statement in the AMS survey report. The closest result to this testimony is the following (page 5 of the cited report):

> Respondents who indicated that global warming is happening were asked their views about its primary causes; a large majority indicted that human activity (59%), or human activity and natural causes in more or less equal amounts (11%), were the primary causes.

Again, the report concludes in *direct opposition* to the sworn testimony of Dr. Bezdek. As with his previous claim on the Strengers et al report, Dr. Bezdek does not disclose to the ALJ or Commission the peer-reviewed paper that resulted from this work: Stenhouse et al., *Bull. Am. Met. Soc.*, 2014, 95, pp. 1029-1040.

Q. Are there other examples of the second pattern - of a misunderstanding of the science or cited literature - by the Peabody witnesses?

A. Yes. As noted in my Rebuttal Testimony DOC Ex.___ at 21-23 (Gurney Rebuttal), the Direct Testimony of Dr. Happer (Peabody Ex.___ at WH-2, page 11 (Happer Direct)) contained references to papers that contain neither Dr. Happer's assertions nor the figures that Dr. Happer testified, are contained therein.

C. STRAW MAN ARGUMENTATION

Q. What was the third misleading, biased or otherwise flawed pattern of argument and reasoning used by Peabody witnesses?

A. As I noted above, the testimony of Peabody witnesses employed what is known as straw man argumentation. In this type of flawed argumentation, an argument is refuted, but it is not an argument advanced by an opposing witness. This type of argumentation results in the impression of successful refutation, but has no relevance to the proceeding.

Q. Dr. Gurney, can you provide examples?

A. Yes. This was best exemplified by Drs. Bezdek's and Happer's testimony regarding CO_2 fertilization. Peabody Ex.___ at RHB-1, pages 13-19 (Bezdek Rebuttal) and Peabody Ex.___ at WH-1, pages 2-4, and at WH-2, pages 16-17 (Happer Rebuttal).

Q. Please explain why this testimony is straw-man argumentation.

A. As I discussed in DOC Ex.___ at 5-7 (Gurney Rebuttal), the climate science community has not argued that there is no CO_2 fertilization effect or that CO_2

fertilization has a negative impact. The relevant question (within which the CO_2 fertilization effect resides) is whether or not climate change has a positive or negative impact on vegetation, particularly food crops, and whether or not this has been taken into account in scientific assessments and modeling efforts. The research suggests that the net effect of climate change on food crops is negative and the complete suite of effects have been included, to the extent of scientific knowledge on the subject.

D. ATTACKING THE MESSENGER

Q. What was the fourth misleading, biased or otherwise flawed pattern of argument and reasoning used by Peabody witnesses?

A. The testimony of Peabody witnesses employed an argumentation device known as attacking the messenger. This common form of argumentation has been used by skeptics of ACC, particularly when responding to content within the IPCC assessments.

Q. Dr. Gurney, can you provide examples of this pattern of argument and reasoning in the testimony of the witnesses of Peabody?

A. Yes. Testimony of Peabody witnesses mischaracterized the content of the IPCC reports, and used phrases such as "the IPCC claims" or "IPCC models find" and similar phrasing. As I explained in DOC Ex.___ at 25-28 (Gurney Rebuttal), however, the IPCC reports did nothing more than review the existing peer-reviewed literature and synthesized the information into an assessment of the scientific knowledge on the topic of climate change. Extensive effort goes in to how to express the results of the synthesis to best communicate the breadth of results. There is no such thing as "IPCC models." The authors of the IPCC reports are working scientists who volunteer their time to review the science. They often work in teams on particular chapters or report sections. The IPCC Secretariat itself is composed of a very small staff with no modeling or research capability whatsoever. The Secretariat staff serve a predominantly clerical function. The mischaracterizations by Peabody witnesses, and use of such misleading phrasing creates the impression the IPCC is a research entity imposing results with a predisposed agenda rather than a voluntary network of working scientists that engage in a review of *all* the peer-reviewed literature (including, it is worth noting, that published by self-proclaimed ACC skeptics) and write reports that assess and synthesize that peer-reviewed literature.

III. HAPPER

Q. Do you have additional concerns regarding the Rebuttal Testimony of Dr. Happer?

A. Yes. First, Dr. Happer commented negatively in his Rebuttal Testimony on the qualifications of Drs. Hanemann, Martin and Polasky (Peabody Ex. ___ WH-1, page 2, lines 10-12 (Happer Rebuttal)).

Q. What are your thoughts on those comments regarding qualifications?

A. Dr. Happer suggested that the witnesses "lack adequate formal training in the physical sciences, physics, meteorology, oceanography, biology or other areas needed to assess the scientific performance of climate models." (Peabody Ex. ___ WH-1, page 2, lines 10-12 (Happer Rebuttal)). In reading through the testimony of Drs. Hanemann, Martin and Polasky, it is clear to me that they are commenting on the economic aspects of the modeling endeavor, for which they are qualified. Expertise in academic settings is determined by degree topics and one's publication record.

Furthermore, Dr. Happer has limited expertise in the subject of climate science or economics, as he has published no peer-reviewed papers in climate science or economics (see http://physics.princeton.edu/atomic/happer/Publications.html) and has performed no research related to climate modeling, the carbon cycle, or temperature measurements - all topics on which he has provided testimony. He appears to lack the qualifications to opine on the expertise of persons engaged in areas of study of which he himself has limited knowledge.

Q. What other comments do you have regarding the Rebuttal Testimony of Dr. Happer?

A. Dr. Happer's Rebuttal Testimony repeated the assertions made in his Direct Testimony, except for one new assertion regarding "measurement errors."

Q. What are the repeated assertions?

A. The assertions repeated from his Direct Testimony regarding climate models have been adequately addressed in DOC Ex.___ at 13-16 (Gurney Rebuttal), in CEO Ex.___ at 12-18 (Abraham Rebuttal), and in CEO Ex.___ at 23-26 (Dessler Rebuttal). Further, the claims repeated from his Direct Testimony regarding the benefits of CO_2 to agriculture were adequately addressed in DOC Ex.___ at 19-23 (Gurney Rebuttal).

Q. Do you disagree with the one new assertion of Dr. Happer regarding "measurement errors"?

A. Yes. I disagree with the new assertion, which concerns what Dr. Happer referred to as measurement error; this assertion relates to a claim of Dr. Happer that surface measurement records are biased due to urbanization and the loss of measurement stations (Peabody Ex. ___ at WH-1, page 5, lines 82-97 (Happer Rebuttal)) and (Peabody Ex. ___ at WH-2, pages 19-20 (Happer Rebuttal)).

First, the Happer Rebuttal cited nine papers to support his claim, of which three are from the peer-reviewed literature. Of those three, one (a paper also cited by Peabody Witness Dr. Spencer) is a paper on corrections applied to surface temperature measurements in New Zealand (less than 0.2% of the land surface of the Earth) and one reports on the influence of urbanization in the temperature measurements in eastern China. The final peer-reviewed paper (Wang et al.) has no content related to either urbanization or measurement station loss.

Second, the issue of urbanization and other conditions impacting surface measurement locations have been extensively researched and thorough corrections applied. CEO Ex.___ at 13-14 (Dessler Rebuttal) and CEO Ex.___ at 23-24 (Abraham Rebuttal) offered a series of peer-reviewed publications that cover the extensive effort that goes into correcting for urbanization and other effects. In short, the scientific community has accounted for these effects in the temperature records used to support the observational evidence for ACC.

Q. What is your opinion regarding the issue of urbanization and other conditions impacting surface measurement locations?

A. The synthesis supplied by the IPCC is the best comprehensive review of the temperature records. Chapter 2 of the IPCC 5th Assessment Report (section 2.4, particularly section 2.4.1.3, a copy of which is included as Attachment 1) provides an extensive review of all the temperature records and discusses the siting issues, urbanization effects, and a long list of peer-reviewed papers that provide the methodological details and analysis. The ALJ should reject the assertion of Dr. Happer regarding what he referred to as measurement error.

IV. LINDZEN

Q. What comments do you have regarding the Rebuttal Testimony of Dr. Lindzen?

A. The Rebuttal Testimony of Dr. Lindzen repeated the central themes presented in his Direct Testimony and he provided only one item of new material.

Q. What was the one new topic?

A. The new topic purported to be a critique of a recent peer-reviewed paper that revises (very slightly) NOAA's temperature trend analysis. Peabody Ex.___ at RSL-1, lines 33-69 (Lindzen Rebuttal). Dr. Lindzen provided no peer-reviewed support for his critique of the analysis. Rather, his critique appeared to be a general statement about the motivation of researchers involved in analyzing the temperature records. A general statement without supporting evidence is immaterial to the matters at issue of this proceeding (whether the IWG's SCC is the best measure for determining the cost of carbon) and, were it pertinent, is a topic best analyzed by psychologists, an area of expertise Dr. Lindzen does not claim to possess.

V. BEZDEK

Q. What comments do you have regarding the Rebuttal Testimony of Dr. Bezdek?

A. Dr. Bezdek reviewed a series of statements made in the Direct Testimony of Dr. Polasky, and provided his rebuttal to each. I will comment on the content related to the physical and biological science aspects and leave the economic content to witnesses with expertise in economics.

First, as I noted in my above discussion of argument patterns, Dr. Bezdek consistently used common argument patterns with underlying logical flaws in his Rebuttal Testimony. For example, as support for his response to Dr. Polasky's testimony, that "The overwhelming majority of peer-reviewed articles on climate change and of scientists agree that emissions of CO_2 and other greenhouse gases have a warming effect on the planet and that the evidence is sufficiently strong to justify policy action," (Peabody Ex. ___ at RHB-1, pages 3-4 (Bezdek Rebuttal)) Dr. Bezdek relied almost entirely on non-peer-reviewed content of dubious quality. He claimed (Peabody Ex. ___ at RHB-1, page 3, line 92 (Bezdek Rebuttal)) to have evidence from the journals *Science* and *Nature* but failed to identify any such evidence.

Second, Dr. Bezdek purported to provide "empirical evidence" that counters ACC (Peabody Ex. ___ at RHB-1, line 168 (Bezdek Rebuttal). That so-called empirical evidence consisted of an unreferenced Figure (Peabody Ex. ___ at RHB-1, lines 178-

179 (Bezdek Rebuttal)) and congressional testimony (Peabody Ex. ___ at RHB-1, line 182 (Bezdek Rebuttal)) rather than peer-reviewed research.

Fourth, in this same line of argument, Dr. Bezdek purported to quote a study by Steinkamp and Hickler, and claimed that the study is "further evidence that 'global warming has ceased.'" Peabody Ex.___ RHB-1, lines 208-211 (Bezdek Rebuttal). Examination of this paper, however, shows no such statement nor does the paper imply such a conclusion. Indeed, the synthesis statement in the abstract of the paper states:

> *Synthesis*. Our results indeed suggest that dry forests have been experiencing increasing drought-induced mortality. However, this does not apply to forests in general and the spatial variability has been large. The poor correspondence between the simulated and reported mortality events indicates that models like LPJ-GUESS driven by standard climatologies, and soil input data do not represent drought-induced mortality well. But the poor detection of the reported drought events in our climate indices also suggests that drought stress might not be the main driver of all the reported drought-mortality events.

I reviewed and found nothing in this paper that supports Dr. Bezdek's claim. This statement from the study by Steinkamp and Hickler indicates that dry forests are experiencing increased mortality from drought and that the biological models employed ("LPJ-GUESS") do not do a good job at representing this type of mortality. It is unclear what Dr. Bezdek might have been quoting, but it was not the study by Steinkamp and Hickler.

Fifth, Dr. Bezdek mischaracterized the consensus around ACC as a "manufactured myth" ((Peabody Ex. ___ at RHB-1, line 213 (Bezdek Rebuttal)). As support for his characterization, he relied on mostly non-peer-reviewed research.

Further, the peer-reviewed studies that he relied on as purported support for his characterization are both misquoted and misunderstood as I described above, in my discussion of the argument patterns (under the heading, “Misunderstanding of Science or Cited Literature”).

Sixth, the next rebuttal argument provided by Dr. Bezdek reiterated his assertion of a pause in warming and the biases present in temperature records. Peabody Ex.___ at RHB-1, lines 276-336 (Bezdek Rebuttal). As support for his assertion, Dr. Bezdek provided no evidence based on peer-reviewed research, and chose instead to present figures presumably derived from newspaper stories and magazines such as *Forbes*. Peabody Ex.___ RHB-1, lines 306-309 (Bezdek Rebuttal).

Finally, the remainder of his Rebuttal Testimony that related to issues in physical or biological science reiterated the assertion made in his Direct Testimony that agriculture will benefit from CO_2 and warming (Peabody Ex.___ RHB-1, lines 393-540 (Bezdek Rebuttal)). As I discussed above, (under the heading “Straw Man Argumentation”) this is straw-man argumentation and I previously responded to this in DOC Ex.___ at 3-7 (Gurney Rebuttal).

Q. What are your conclusions regarding the Bezdek Rebuttal Testimony?

A. The ALJ should not adopt the Bezdek Rebuttal Testimony as to any of the seven topics that relate to issues in physical or biological science.

Q. Does this conclude your Testimony?

A. Yes.

2.4 Changes in Temperature

2.4.1 Land Surface Air Temperature

2.4.1.1 Large-Scale Records and Their Uncertainties

AR4 concluded global land-surface air temperature (LSAT) had increased over the instrumental period of record, with the warming rate approximately double that reported over the oceans since 1979. Since AR4, substantial developments have occurred including the production of revised data sets, more digital data records, and new data set efforts. These innovations have improved understanding of data issues and uncertainties, allowing better quantification of regional changes. This reinforces confidence in the reported globally averaged LSAT time series behaviour.

Global Historical Climatology Network Version 3 (GHCNv3) incorporates many improvements (Lawrimore et al., 2011) but was found to be virtually indistinguishable at the global mean from version 2 (used in AR4). Goddard Institute of Space Studies (GISS) continues to provide an estimate based upon primarily GHCN, accounting for urban impacts through nightlights adjustments (Hansen et al., 2010). CRUTEM4 (Jones et al., 2012) incorporates additional station series and also newly homogenized versions of many individual station records. A new data product from a group based predominantly at Berkeley (Rohde et al., 2013a) uses a method that is substantially distinct from earlier efforts (further details on all the data sets and data availability are given in Supplementary Material 2.SM.4). Despite the range of approaches, the long-term variations and trends broadly agree among these various LSAT estimates, particularly after 1900. Global LSAT has increased (Figure 2.14, Table 2.4).

Since AR4, various theoretical challenges have been raised over the verity of global LSAT records (Pielke et al., 2007). Globally, sampling and methodological independence has been assessed through sub-sampling (Parker et al., 2009; Jones et al., 2012), creation of an entirely new and structurally distinct product (Rohde et al., 2013b) and a complete reprocessing of GHCN (Lawrimore et al., 2011). None of these yielded more than minor perturbations to the global LSAT records since 1900. Willett et al. (2008) and Peterson et al. (2011) explicitly showed that changes in specific and relative humidity (Section 2.5.5) were physically consistent with reported temperature trends, a result replicated in the ERA reanalyses (Simmons et al., 2010). Various investigators (Onogi et al., 2007; Simmons et al., 2010; Parker, 2011; Vose et al., 2012a) showed that LSAT estimates from modern reanalyses were in quantitative agreement with observed products.

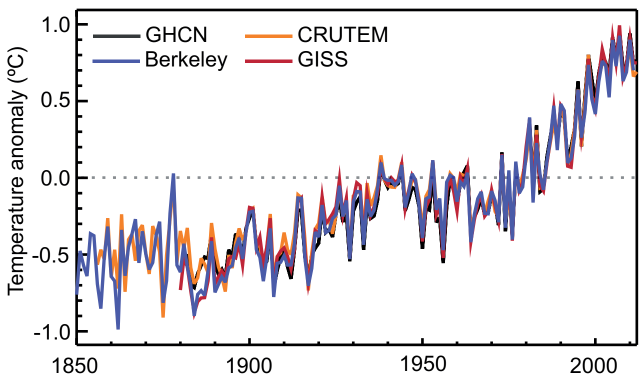


Figure 2.14 | Global annual average land-surface air temperature (LSAT) anomalies relative to a 1961–1990 climatology from the latest versions of four different data sets (Berkeley, CRUTEM, GHCN and GISS).

Particular controversy since AR4 has surrounded the LSAT record over the United States, focussed on siting quality of stations in the US Historical Climatology Network (USHCN) and implications for long-term trends. Most sites exhibit poor current siting as assessed against official WMO siting guidance, and may be expected to suffer potentially large siting-induced absolute biases (Fall et al., 2011). However, overall biases for the network since the 1980s are *likely* dominated by instrument type (owing to replacement of Stevenson screens with maximum minimum temperature systems (MMTS) in the 1980s at the majority of sites), rather than siting biases (Menne et al., 2010; Williams et al., 2012). A new automated homogeneity assessment approach (also used in GHCNv3, Menne and Williams, 2009) was developed that has been shown to perform as well or better than other contemporary approaches (Venema et al., 2012). This homogenization procedure *likely* removes much of the bias related to the network-wide changes in the 1980s (Menne et al., 2010; Fall et al., 2011; Williams et al., 2012). Williams et al. (2012) produced an ensemble of data set realizations using perturbed settings of this procedure and concluded through assessment against plausible test cases that there existed a propensity to under-estimate adjustments. This propensity is critically dependent upon the (unknown) nature of the inhomogeneities in the raw data records. Their homogenization increases both minimum temperature and maximum temperature centennial-time-scale USA average LSAT trends. Since 1979 these adjusted data agree with a range of reanalysis products whereas the raw records do not (Fall et al., 2010; Vose et al., 2012a).

Regional analyses of LSAT have not been limited to the United States. Various national and regional studies have undertaken assessments for Europe (Winkler, 2009; Bohm et al., 2010; Tietavainen et al., 2010; van

Table 2.4: | Trend estimates and 90% confidence intervals (Box 2.2) for LSAT global average values over five common periods.

Data Set	Trends in °C per decade				
	1880–2012	1901–2012	1901–1950	1951–2012	1979–2012
CRUTEM4.1.1.0 (Jones et al., 2012)	0.086 ± 0.015	0.095 ± 0.020	0.097 ± 0.029	0.175 ± 0.037	0.254 ± 0.050
GHCNv3.2.0 (Lawrimore et al., 2011)	0.094 ± 0.016	0.107 ± 0.020	0.100 ± 0.033	0.197 ± 0.031	0.273 ± 0.047
GISS (Hansen et al., 2010)	0.095 ± 0.015	0.099 ± 0.020	0.098 ± 0.032	0.188 ± 0.032	0.267 ± 0.054
Berkeley (Rohde et al., 2013)	0.094 ± 0.013	0.101 ± 0.017	0.111 ± 0.034	0.175 ± 0.029	0.254 ± 0.049

der Schrier et al., 2011), China (Li et al., 2009; Zhen and Zhong-Wei, 2009; Li et al., 2010a; Tang et al., 2010), India (Jain and Kumar, 2012), Australia (Trewin, 2012), Canada (Vincent et al., 2012), South America, (Falvey and Garreaud, 2009) and East Africa (Christy et al., 2009). These analyses have used a range of methodologies and, in many cases, more data and metadata than available to the global analyses. Despite the range of analysis techniques they are generally in broad agreement with the global products in characterizing the long-term changes in mean temperatures. This includes some regions, such as the Pacific coast of South America, that have exhibited recent cooling (Falvey and Garreaud, 2009). Of specific importance for the early global records, large (>1°C) summer time warm bias adjustments for many European 19th century and early 20th century records were revisited and broadly confirmed by a range of approaches (Bohm et al., 2010; Brunet et al., 2011).

Since AR4 efforts have also been made to interpolate Antarctic records from the sparse, predominantly coastal ground-based network (Chapman and Walsh, 2007; Monaghan et al., 2008; Steig et al., 2009; O'Donnell et al., 2011). Although these agree that Antarctica as a whole has warmed since the late 1950s, substantial multi-annual to multi-decadal variability and uncertainties in reconstructed magnitude and spatial trend structure yield only *low confidence* in the details of pan-Antarctic regional LSAT changes.

In summary, it is certain that globally averaged LSAT has risen since the late 19th century and that this warming has been particularly marked since the 1970s. Several independently analyzed global and regional LSAT data products support this conclusion. There is *low confidence* in changes prior to 1880 owing to the reduced number of estimates, non-standardized measurement techniques, the greater spread among the estimates and particularly the greatly reduced observational sampling. *Confidence* is also *low* in the spatial detail and magnitude of LSAT trends in sparsely sampled regions such as Antarctica. Since AR4 significant efforts have been undertaken to identify and adjust for data issues and new estimates have been produced. These innovations have further strengthened overall understanding of the global LSAT records.

2.4.1.2 Diurnal Temperature Range

In AR4 diurnal temperature range (DTR) was found, globally, to have narrowed since 1950, with minimum daily temperatures increasing faster than maximum daily temperatures. However, significant multi-decadal variability was highlighted including a recent period from 1997 to 2004 of no change, as both maximum and minimum temperatures rose at similar rates. The Technical Summary of AR4 highlighted changes in DTR and their causes as a key uncertainty. Since AR4, uncertainties in DTR and its physical interpretation have become even more apparent.

No dedicated global analysis of DTR has been undertaken subsequent to Vose et al. (2005a), although global behaviour has been discussed in two broader ranging analyses. Rohde et al. (2012) and Wild et al. (2007) note an apparent reversal since the mid-1980s; with DTR subsequently increasing. This decline and subsequent increase in DTR over global land surfaces is qualitatively consistent with the dimming and subsequent brightening noted in Section 2.3.3.1. Donat et al. (2013c) using HadEX2 (Section 2.6) find significant decreasing DTR trends in more than half of the land areas assessed but less than 10% of land with significant increases since 1951. Available trend estimates (–0.04 ± 0.01°C per decade over 1950–2011 (Rohde et al., 2013b) and –0.066°C per decade over 1950–2004 (Vose et al., 2005a)) are much smaller than global mean LSAT average temperature trends over 1951–2012 (Table 2.4). It therefore logically follows that globally averaged maximum and minimum temperatures over land have both increased by in excess of 0.1°C per decade since 1950.

Regionally, Makowski et al. (2008) found that DTR behaviour in Europe over 1950 to 2005 changed from a decrease to an increase in the 1970s in Western Europe and in the 1980s in Eastern Europe. Sen Roy and Balling (2005) found significant increases in both maximum and minimum temperatures for India, but little change in DTR over 1931–2002. Christy et al. (2009) reported that for East Africa there has been no pause in the narrowing of DTR in recent decades. Zhou and Ren (2011) reported a significant decrease in DTR over mainland China of –0.15°C per decade during 1961–2008.

Various investigators (e.g., Christy et al. (2009), Pielke and Matsui (2005), Zhou and Ren (2011)) have raised doubts about the physical interpretation of minimum temperature trends, hypothesizing that microclimate and local atmospheric composition impacts are more apparent because the dynamical mixing at night is much reduced. Parker (2006) investigated this issue arguing that if data were affected in this way, then a trend difference would be expected between calm and windy nights. However, he found no such minimum temperature differences on a global average basis. Using more complex boundary layer modelling techniques, Steeneveld et al. (2011) and McNider et al. (2012) showed much lower sensitivity to windspeed variations than posited by Pielke and Matsui but both concluded that boundary layer understanding was key to understanding the minimum temperature changes. Data analysis and long-term side-by-side instrumentation field studies show that real non-climatic data artefacts certainly affect maximum and minimum differently in the raw records for both recent (Fall et al., 2011; Williams et al., 2012) and older (Bohm et al., 2010; Brunet et al., 2011) records. Hence there could be issues over interpretation of apparent DTR trends and variability in many regions (Christy et al., 2006, 2009; Fall et al., 2011; Zhou and Ren, 2011; Williams et al., 2012), particularly when accompanied by regional-scale land-use/land-cover (LULC) changes (Christy et al., 2006).

In summary, *confidence* is *medium* in reported decreases in observed global DTR, noted as a key uncertainty in AR4. Several recent analyses of the raw data on which many previous analyses were based point to the potential for biases that differently affect maximum and minimum average temperatures. However, apparent changes in DTR are much smaller than reported changes in average temperatures and therefore it is *virtually certain* that maximum and minimum temperatures have increased since 1950.

2.4.1.3 Land Use Change and Urban Heat Island Effects

In AR4 Urban Heat Island (UHI) effects were concluded to be real local phenomena with negligible impact on large-scale trends. UHI and land-use land-cover change (LULC) effects arise mainly because the modified surface affects the storage and transfer of heat, water and airflow. For single discrete locations these impacts may dominate all other factors.

Regionally, most attention has focused on China. A variety of investigations have used methods as diverse as SST comparisons (e.g., Jones et al., 2008), urban minus rural (e.g., Ren et al., 2008; Yang et al., 2011), satellite observations (Ren and Ren, 2011) and observations minus reanalysis (e.g., Hu et al., 2010; Yang et al., 2011). Interpretation is complicated because often studies have used distinct versions of station series. For example, the effect in Beijing is estimated at 80% (Ren et al., 2007) or 40% (Yan et al., 2010) of the observed trend depending on data corrections applied. A representative sample of these studies suggest the effect of UHI and LULC is approximately 20% of the trend in Eastern China as a whole and of the order 0.1°C per decade nationally (Table 1 in Yang et al., 2011) over the last 30 years, but with very substantial uncertainties. These effects have *likely* been partially or completely accounted for in many homogenized series (e.g., Li et al., 2010b; Yan et al., 2010). Fujibe (2009) ascribes about 25% of Japanese warming trends in 1979–2006 to UHI effects. Das et al. (2011) confirmed that many Japanese sites have experienced UHI warming but that rural stations show unaffected behaviour when compared to nearby SSTs.

There is an important distinction to be made between UHI trend effects in regions underseeing rapid development and those that have been developed for a long time. Jones and Lister (2009) and Wilby et al. (2011) using data from London (UK) concluded that some sites that have always been urban and where the UHI has not grown in magnitude will exhibit regionally indicative trends that agree with nearby rural locations and that in such cases the time series may exhibit multi-decadal trends driven primarily by synoptic variations. A lack of obvious time-varying UHI influences was also noted for Sydney, Melbourne and Hobart in Australia by Trewin (2012). The impacts of urbanization also will be dependent on the natural LULC characteristics that they replace. Zhang et al. (2010) found no evidence for urban influences in the desert North West region of China despite rapid urbanization.

Global adjusted data sets *likely* account for much of the UHI effect present in the raw data. For the US network, Hausfather et al. (2013) showed that the adjustments method used in GHCNv3 removed much of an apparent systematic difference between urban and rural locations, concluding that this arose from adjustment of biased urban location data. Globally, Hansen et al. (2010) used satellite-based nightlight radiances to estimate the worldwide influence on LSAT of local urban development. Adjustments reduced the global 1900–2009 temperature change (averaged over land and ocean) only from 0.71°C to 0.70°C. Wickham et al. (2013) also used satellite data and found that urban locations in the Berkeley data set exhibited even less warming than rural stations, although not statistically significantly so, over 1950 to 2010.

Studies of the broader effects of LULC since AR4 have tended to focus on the effects of irrigation on temperatures, with a large number of studies in the Californian central belt (Christy et al., 2006; Kueppers et al., 2007; Bonfils et al., 2008; Lo and Famiglietti, 2013). They find cooler average temperatures and a marked reduction in DTR in areas of active irrigation and ascribe this to increased humidity; effectively a repartitioning of moist and dry energy terms. Reanalyses have also been used to estimate the LULC signature in LSAT trends. Fall et al. (2010) found that the North American Regional Reanalysis generated overall surface air temperature trends for 1979–2003 similar to observed records. Observations-minus-reanalysis trends were most positive for barren and urban areas, in accord with the results of Lim et al. (2008) using the NCEP/NCAR and ERA-40 reanalyses, and negative in agricultural areas.

McKitrick and Michaels (2004) and de Laat and Maurellis (2006) assessed regression of trends with national socioeconomic and geographical indicators, concluding that UHI and related LULC have caused much of the observed LSAT warming. AR4 concluded that this correlation ceases to be statistically significant if one takes into account the fact that the locations of greatest socioeconomic development are also those that have been most warmed by atmospheric circulation changes but provided no explicit evidence for this overall assessment result. Subsequently McKitrick and Michaels (2007) concluded that about half the reported warming trend in global-average land surface air temperature in 1980–2002 resulted from local land surface changes and faults in the observations. Schmidt (2009) undertook a quantitative analysis that supported AR4 conclusions that much of the reported correlation largely arose due to naturally occurring climate variability and model over-fitting and was not robust. Taking these factors into account, modified analyses by McKitrick (2010) and McKitrick and Nierenberg (2010) still yielded significant evidence for such contamination of the record.

In marked contrast to regression based studies, several studies have shown the methodologically diverse set of modern reanalysis products and the various LSAT records at global and regional levels to be similar since at least the mid-20th century (Simmons et al., 2010; Parker, 2011; Ferguson and Villarini, 2012; Jones et al., 2012; Vose et al., 2012a). These reanalyses do not directly assimilate the LSAT measurements but rather infer LSAT estimates from an observational constraint provided by much of the rest of the global observing system, thus representing an independent estimate. A hypothesized residual significant warming artefact argued for by regression-based analyses is therefore physically inconsistent with many other components of the global observing system according to a broad range of state-of-the-art data assimilation models (Box 2.3). Further, Efthymiadis and Jones (2010) estimated an absolute upper limit on urban influence globally of 0.02°C per decade, or about 15% of the total LSAT trends, in 1951–2009 from trends of coastal land and SST.

In summary, it is indisputable that UHI and LULC are real influences on raw temperature measurements. At question is the extent to which they remain in the global products (as residual biases in broader regionally representative change estimates). Based primarily on the range of urban minus rural adjusted data set comparisons and the degree of agreement of these products with a broad range of reanalysis products, it is *unlikely* that any uncorrected urban heat-island effects and LULC change effects have raised the estimated centennial globally averaged LSAT trends by more than 10% of the reported trend (*high confidence,* based on robust evidence and high agreement). This is an average value; in some regions with rapid development, UHI and LULC change impacts on regional trends may be substantially larger.

2.4.2 Sea Surface Temperature and Marine Air Temperature

AR4 concluded that 'recent' warming (since the 1950s) is strongly evident at all latitudes in SST over each ocean. Prominent spatio-temporal structures including the ENSO and decadal variability patterns in the Pacific Ocean (Box 2.5) and a hemispheric asymmetry in the Atlantic Ocean were highlighted as contributors to the regional differences in surface warming rates, which in turn affect atmospheric circulation. Since AR4 the availability of metadata has increased, data completeness has improved and a number of new SST products have been produced. Intercomparisons of data obtained by different measurement methods, including satellite data, have resulted in better understanding of errors and biases in the record.

2.4.2.1 Advances in Assembling Data Sets and in Understanding Data Errors

2.4.2.1.1 *In situ* data records

Historically, most SST observations were obtained from moving ships. Buoy measurements comprise a significant and increasing fraction of *in situ* SST measurements from the 1980s onward (Figure 2.15). Improvements in the understanding of uncertainty have been expedited by the use of metadata (Kent et al., 2007) and the recovery of observer instructions and other related documents. Early data were systematically cold biased because they were made using canvas or wooden buckets that, on average, lost heat to the air before the measurements were taken. This effect has long been recognized (Brooks, 1926), and prior to AR4 represented the only artefact adjusted in gridded SST products, such as HadSST2 (Rayner et al., 2006) and ERSST (Smith et al., 2005, 2008), which were based on 'bucket correction' methods by Folland and Parker (1995) and Smith and Reynolds (2002), respectively. The adjustments, made using ship observations of Night Marine Air Temperature (NMAT) and other sources, had a striking effect on the SST global mean estimates: note the difference in 1850–1941 between HadSST2 and International Comprehensive Ocean-Atmosphere Data Set (ICOADS) curves in Figure 2.16 (a brief description of SST and NMAT data sets and their methods is given in Supplementary Material 2.SM.4.3).

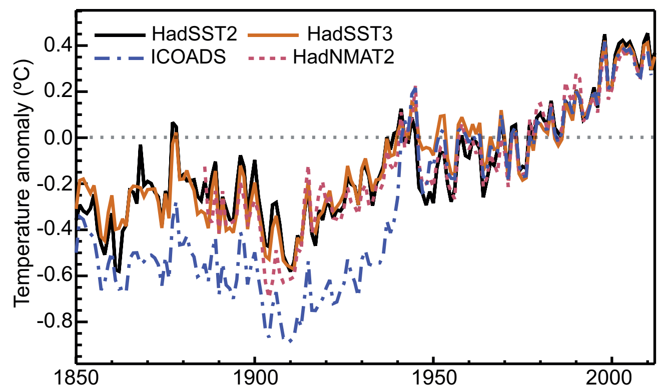


Figure 2.16 | Global annual average sea surface temperature (SST) and Night Marine Air Temperature (NMAT) relative to a 1961–1990 climatology from gridded data sets of SST observations (HadSST2 and its successor HadSST3), the raw SST measurement archive (ICOADS, v2.5) and night marine air temperatures data set HadNMAT2 (Kent et al., 2013). HadSST2 and HadSST3 both are based on SST observations from versions of the ICOADS data set, but differ in degree of measurement bias correction.

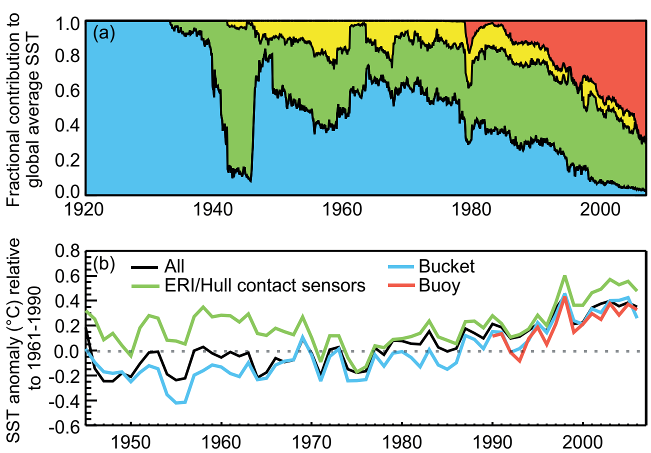


Figure 2.15 | Temporal changes in the prevalence of different measurement methods in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). (a) Fractional contributions of observations made by different measurement methods: bucket observations (blue), engine room intake (ERI) and hull contact sensor observations (green), moored and drifting buoys (red), and unknown (yellow). (b) Global annual average sea surface temperature (SST) anomalies based on different kinds of data: ERI and hull contact sensor (green), bucket (blue), buoy (red), and all (black). Averages are computed over all 5° × 5° grid boxes where both ERI/hull and bucket measurements, but not necessarily buoy data, were available. (Adapted from Kennedy et al., 2011a.)

Buckets of improved design and measurement methods with smaller, on average, biases came into use after 1941 (Figure 2.15, top); average biases were reduced further in recent decades, but not eliminated (Figure 2.15, bottom). Increasing density of SST observations made possible the identification (Reynolds et al., 2002, 2010; Kennedy et al., 2012) and partial correction of more recent period biases (Kennedy et al., 2011a). In particular, it is hypothesized that the proximity of the hot engine often biases engine room intake (ERI) measurements warm (Kent et al., 2010). Because of the prevalence of the ERI measurements among SST data from ships, the ship SSTs are biased warm by 0.12°C to 0.18°C on average compared to the buoy data (Reynolds et al., 2010; Kennedy et al., 2011a, 2012). An assessment of the potential impact of modern biases can be ascertained by considering the difference

Table 2.5 | Trend estimates and 90% confidence intervals (Box 2.2) for two subsequent versions of the HadSST data set over five common periods. HadSST2 has been used in AR4; HadSST3 is used in this chapter.

Data Set	Trends in °C per decade				
	1880–2012	1901–2012	1901–1950	1951–2012	1979–2012
HadSST3 (Kennedy et al., 2011a)	0.054 ± 0.012	0.067 ± 0.013	0.117 ± 0.028	0.074 ± 0.027	0.124 ± 0.030
HadSST2 (Rayner et al., 2006)	0.051 ± 0.015	0.069 ± 0.012	0.084 ± 0.055	0.098 ± 0.017	0.121 ± 0.033

between HadSST3 (bias corrections applied throughout) and HadSST2 (bucket corrections only) global means (Figure 2.16): it is particularly prominent in 1945–1970 period, when rapid changes in prevalence of ERI and bucket measurements during and after the World War II affect HadSST2 owing to the uncorrected measurement biases (Thompson et al., 2008), while these are corrected in HadSST3. Nevertheless, for periods longer than a century the effect of HadSST3-HadSST2 differences on linear trend slopes is small relative to the trend uncertainty (Table 2.5). Some degree of independent check on the validity of HadSST3 adjustments comes from a comparison to sub-surface temperature data (Gouretski et al., 2012) (see Section 3.2).

The traditional approach to modeling random error of *in situ* SST data assumed the independence of individual measurements. Kent and Berry (2008) identified the need to account for error correlation for measurements from the same "platform" (i.e., an individual ship or buoy), while measurement errors from different platforms remain independent.. Kennedy et al. (2011b) achieved that by introducing platform-dependent biases, which are constant within the same platform, but change randomly from one platform to another. Accounting for such correlated errors in HadSST3 resulted in estimated error for global and hemispheric monthly means that are more than twice the estimates given by HadSST2. The uncertainty in many, but not all, components of the HadSST3 product is represented by the ensemble of its realizations (Figure 2.17).

Data sets of marine air temperatures (MATs) have traditionally been restricted to nighttime series only (NMAT data sets) due to the direct solar heating effect on the daytime measurements, although corrected daytime MAT records for 1973–present are already available (Berry and Kent, 2009). Other major biases, affecting both nighttime and daytime MAT are due to increasing deck height with the general increase in the size of ships over time and non-standard measurement practices. Recently these biases were re-examined and explicit uncertainty calculation undertaken for NMAT by Kent et al. (2013), resulting in the HadNMAT2 data set.

2.4.2.1.2 Satellite SST data records

Satellite SST data sets are based on measuring electromagnetic radiation that left the ocean surface and got transmitted through the atmosphere. Because of the complexity of processes involved, the majority of such data has to be calibrated on the basis of *in situ* observations. The resulting data sets, however, provide a description of global SST fields with a level of spatial detail unachievable by *in situ* data only. The principal IR sensor is the Advanced Very High Resolution Radiometer (AVHRR). Since AR4, the AVHRR time series has been reprocessed consistently back to March 1981 (Casey et al., 2010) to create the AVHRR Pathfinder v5.2 data set. Passive microwave data sets of SST are available since 1997 equatorward of 40° and near-globally since 2002 (Wentz et al., 2000; Gentemann et al., 2004). They are generally less accurate than IR-based SST data sets, but their superior coverage in areas of persistent cloudiness provides SST estimates where the IR record has none (Reynolds et al., 2010).

The (Advanced) Along Track Scanning Radiometer (A)ATSR) series of three sensors was designed for climate monitoring of SST; their com-

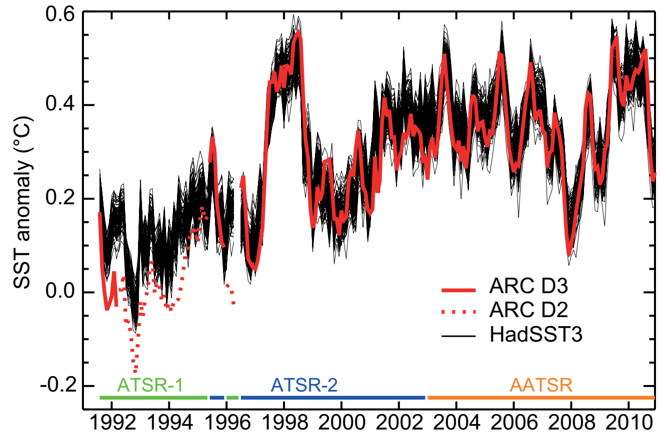


Figure 2.17 | Global monthly mean sea surface temperature (SST) anomalies relative to a 1961–1990 climatology from satellites (ATSRs) and *in situ* records (HadSST3). Black lines: the 100-member HadSST3 ensemble. Red lines: ATSR-based nighttime subsurface temperature at 0.2 m depth ($SST_{0.2m}$) estimates from the ATSR Reprocessing for Climate (ARC) project. Retrievals based on three spectral channels (D3, solid line) are more accurate than retrievals based on only two (D2, dotted line). Contributions of the three different ATSR missions to the curve shown are indicated at the bottom. The *in situ* and satellite records were co-located within 5° × 5° monthly grid boxes: only those where both data sets had data for the same month were used in the comparison. (Adapted from Merchant et al. 2012.)

bined record starts in August 1991 and exceeds two decades (it stopped with the demise of the ENVISAT platform in 2012). The (A) ATSRs are 'dual-view' IR radiometers intended to allow atmospheric effects removal without the use of *in situ* observations. Since AR4, (A)ATSR observations have been reprocessed with new estimation techniques (Embury and Merchant, 2011). The resulting SST products seem to be more accurate than many *in situ* observations (Embury et al., 2011). In terms of monthly global means, the agreement is illustrated in Figure 2.17. By analyzing (A)ATSR and *in situ* data together, Kennedy at al. (2012) verified and extended existing models for biases and random errors of *in situ* data.

2.4.2.2 Interpolated SST Products and Trends

SST data sets form a major part of global surface temperature analyses considered in this assessment report. To use an SST data set as a boundary condition for atmospheric reanalyses products (Box 2.3) or in atmosphere-only climate simulations (considered in Chapter 9 onwards), gridded data sets with complete coverage over the global ocean are typically needed. These are usually based on a special form of kriging (optimal interpolation) procedure that retains large-scale correlation structures and can accommodate very sparse data coverage. For the pre-satellite era (generally, before October 1981) only *in situ* data are used; for the latter period some products also use AVHRR data. Figure 2.18 compares interpolated SST data sets that extend back to the 19th century with the uninterpolated HadSST3 and HadNMAT2 products. Linear trend estimates for global mean SSTs from those products updated through 2012 are presented in Table 2.6. Differences between the trends from different data sets are larger when the calculation period is shorter (1979–2012) or has lower quality data (1901–1950); these are due mainly to different data coverage of underlying observational data sets and bias correction methods used in these products.

Table 2.6 | Trend estimates and 90% confidence intervals (Box 2.2) for interpolated SST data sets (uninterpolated state-of-the-art HadSST3 data set is included for comparison). Dash indicates not enough data available for trend calculation.

Data Set	Trends in °C per decade				
	1880–2012	1901–2012	1901–1950	1951–2012	1979–2012
HadISST (Rayner et al., 2003)	0.042 ± 0.007	0.052 ± 0.007	0.067 ± 0.024	0.064 ± 0.015	0.072 ± 0.024
COBE-SST (Ishii et al., 2005)	–	0.058 ± 0.007	0.066 ± 0.032	0.071 ± 0.014	0.073 ± 0.020
ERSSTv3b (Smith et al., 2008)	0.054 ± 0.015	0.071 ± 0.011	0.097 ± 0.050	0.088 ± 0.017	0.105 ± 0.031
HadSST3 (Kennedy et al., 2011a)	0.054 ± 0.012	0.067 ± 0.013	0.117 ± 0.028	0.074 ± 0.027	0.124 ± 0.030

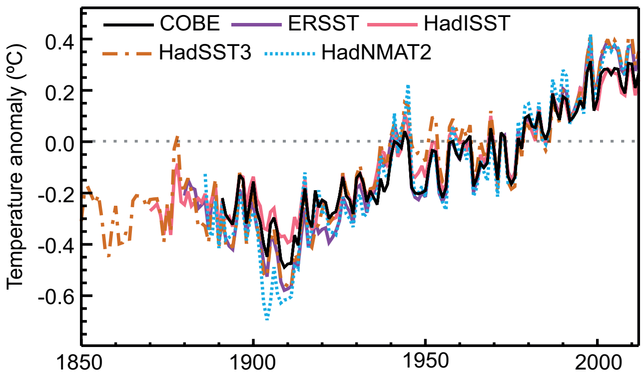


Figure 2.18 | Global annual average sea surface temperature (SST) and Night Marine Air Temperature (NMAT) relative to a 1961–1990 climatology from state of the art data sets. Spatially interpolated products are shown by solid lines; non-interpolated products by dashed lines.

Starting in the 1980s each decade has been significantly warmer at the Earth's surface than any preceding decade since the 1850s in HadCRUT4, a data set that explicitly quantifies a large number of sources of uncertainty (Figure 2.19). Each of the last three decades is also the warmest in the other two GMST data sets, but these have substantially less mature and complete uncertainty estimates, precluding such an assessment of significance of their decadal differences. The GISS and MLOST data sets fall outside the 90% CI of HadCRUT4 for several decades in the 20th century (Figure 2.19). These decadal differences could reflect residual biases in one or more data set, an incomplete treatment of uncertainties in HadCRUT4.1 or a combination of these effects (Box 2.1). The data sets utilize different LSAT (Section 2.4.1) and SST (Section 2.4.2) component records (Supplementary Material 2.SM.4.3.4) that in the case of SST differ somewhat in their multi-decadal trend behaviour (Table 2.6 compare HadSST3 and ERSSTv3b).

In summary, it is certain that global average sea surface temperatures (SSTs) have increased since the beginning of the 20th century. Since AR4, major improvements in availability of metadata and data completeness have been made, and a number of new global SST records have been produced. Intercomparisons of new SST data records obtained by different measurement methods, including satellite data, have resulted in better understanding of uncertainties and biases in the records. Although these innovations have helped highlight and quantify uncertainties and affect our understanding of the character of changes since the mid-20th century, they do not alter the conclusion that global SSTs have increased both since the 1950s and since the late 19th century.

2.4.3 Global Combined Land and Sea Surface Temperature

AR4 concluded that the GMST had increased, with the last 50 years increasing at almost double the rate of the last 100 years. Subsequent developments in LSAT and SST have led to better understanding of the data and their uncertainties as discussed in preceding sections. This improved understanding has led to revised global products.

Changes have been made to all three GMST data sets that were used in AR4 (Hansen et al., 2010; Morice et al., 2012; Vose et al., 2012b). These are now in somewhat better agreement with each other over recent years, in large part because HadCRUT4 now better samples the NH high latitude land regions (Jones et al., 2012; Morice et al., 2012) which comparisons to reanalyses had shown led to a propensity for HadCRUT3 to underestimate recent warming (Simmons et al., 2010).

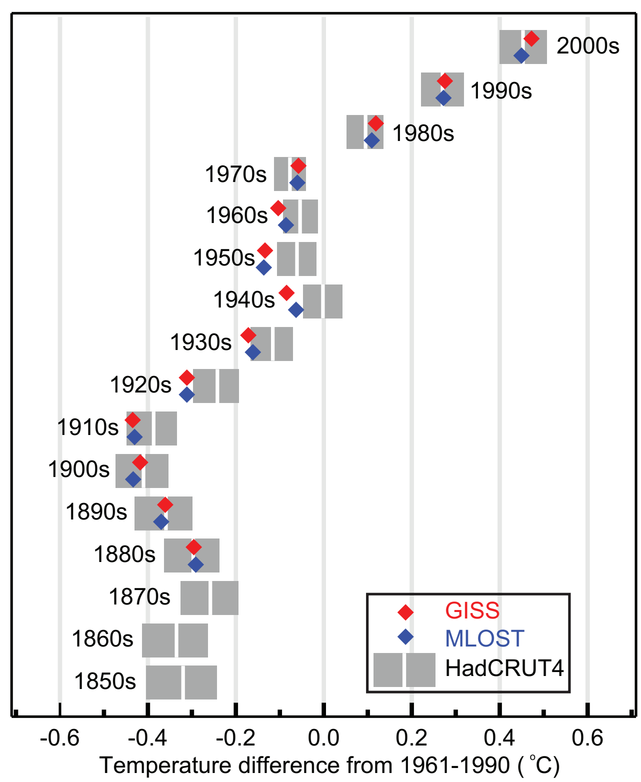


Figure 2.19 | Decadal global mean surface temperature (GMST) anomalies (white vertical lines in grey blocks) and their uncertainties (90% confidence intervals as grey blocks) based upon the land-surface air temperature (LSAT) and sea surface temperature (SST) combined HadCRUT4 (v4.1.1.0) ensemble (Morice et al., 2012). Anomalies are relative to a 1961–1990 climatology. 1850s indicates the period 1850-1859, and so on. NCDC MLOST and GISS data set best-estimates are also shown.

All ten of the warmest years have occurred since 1997, with 2010 and 2005 effectively tied for the warmest year on record in all three products. However, uncertainties on individual annual values are sufficiently large that the ten warmest years are statistically indistinguishable from one another. The global-mean trends are significant for all data sets and multi-decadal periods considered in Table 2.7. Using HadCRUT4 and its uncertainty estimates, the warming from 1850–1900 to 1986–2005 (reference period for the modelling chapters and Annex I) is 0.61 [0.55 to 0.67] °C (90% confidence interval), and the warming from 1850–1900 to 2003–2012 (the most recent decade) is 0.78 [0.72 to 0.85] °C (Supplementary Material 2.SM.4.3.3).

Differences between data sets are much smaller than both interannual variability and the long-term trend (Figure 2.20). Since 1901 almost the whole globe has experienced surface warming (Figure 2.21). Warming has not been linear; most warming occurred in two periods: around 1900 to around 1940 and around 1970 onwards (Figure 2.22. Shorter periods are noisier and so proportionately less of the sampled globe exhibits statistically significant trends at the grid box level (Figure 2.22). The two periods of global mean warming exhibit very distinct spatial signatures. The early 20th century warming was largely a NH mid- to high-latitude phenomenon, whereas the more recent warming is more global in nature. These distinctions may yield important information as to causes (Chapter 10). Differences between data sets are larger in earlier periods (Figures 2.19, 2.20), particularly prior to the 1950s when observational sampling is much more geographically incomplete (and many of the well sampled areas may have been globally unrepresentative (Brönnimann, 2009)), data errors and subsequent methodological impacts are larger (Thompson et al., 2008), and different ways of accounting for data void regions are more important (Vose et al., 2005b).

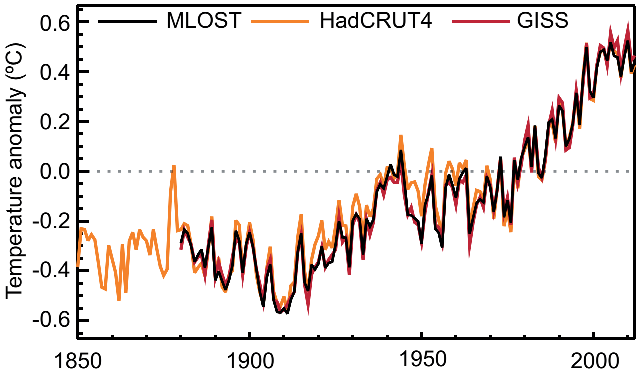


Figure 2.20 | Annual global mean surface temperature (GMST) anomalies relative to a 1961–1990 climatology from the latest version of the three combined land-surface air temperature (LSAT) and sea surface temperature (SST) data sets (HadCRUT4, GISS and NCDC MLOST). Published data set uncertainties are not included for reasons discussed in Box 2.1.

Much interest has focussed on the period since 1998 and an observed reduction in warming trend, most marked in NH winter (Cohen et al., 2012). Various investigators have pointed out the limitations of such short-term trend analysis in the presence of auto-correlated series variability and that several other similar length phases of no warming exist in all the observational records and in climate model simulations

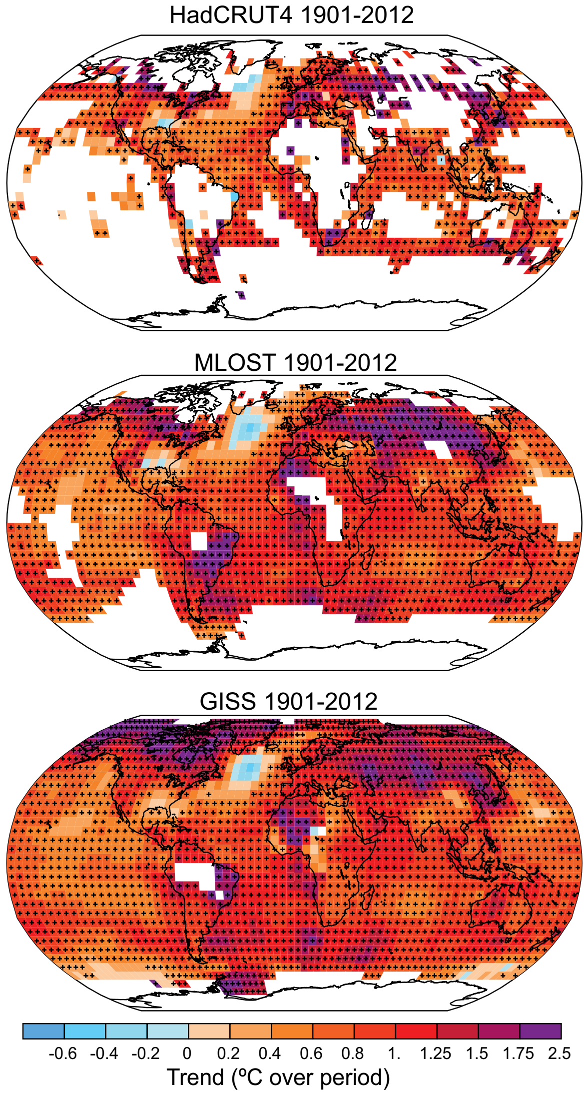


Figure 2.21 | Trends in surface temperature from the three data sets of Figure 2.20 for 1901–2012. White areas indicate incomplete or missing data. Trends have been calculated only for those grid boxes with greater than 70% complete records and more than 20% data availability in first and last decile of the period. Black plus signs (+) indicate grid boxes where trends are significant (i.e., a trend of zero lies outside the 90% confidence interval). Differences in coverage primarily reflect the degree of interpolation to account for data void regions undertaken by the data set providers ranging from none beyond grid box averaging (HadCRUT4) to substantial (GISS).

Table 2.7 | Same as Table 2.4, but for global mean surface temperature (GMST) over five common periods.

Data Set	Trends in °C per decade				
	1880–2012	1901–2012	1901–1950	1951–2012	1979–2012
HadCRUT4 (Morice et al., 2012)	0.062 ± 0.012	0.075 ± 0.013	0.107 ± 0.026	0.106 ± 0.027	0.155 ± 0.033
NCDC MLOST (Vose et al., 2012b)	0.064 ± 0.015	0.081 ± 0.013	0.097 ± 0.040	0.118 ± 0.021	0.151 ± 0.037
GISS (Hansen et al., 2010)	0.065 ± 0.015	0.083 ± 0.013	0.090 ± 0.034	0.124 ± 0.020	0.161 ± 0.033

Docket No. E999/CI-14-643
DOC Ex. ___ KG-S-1
Page 8 of 15

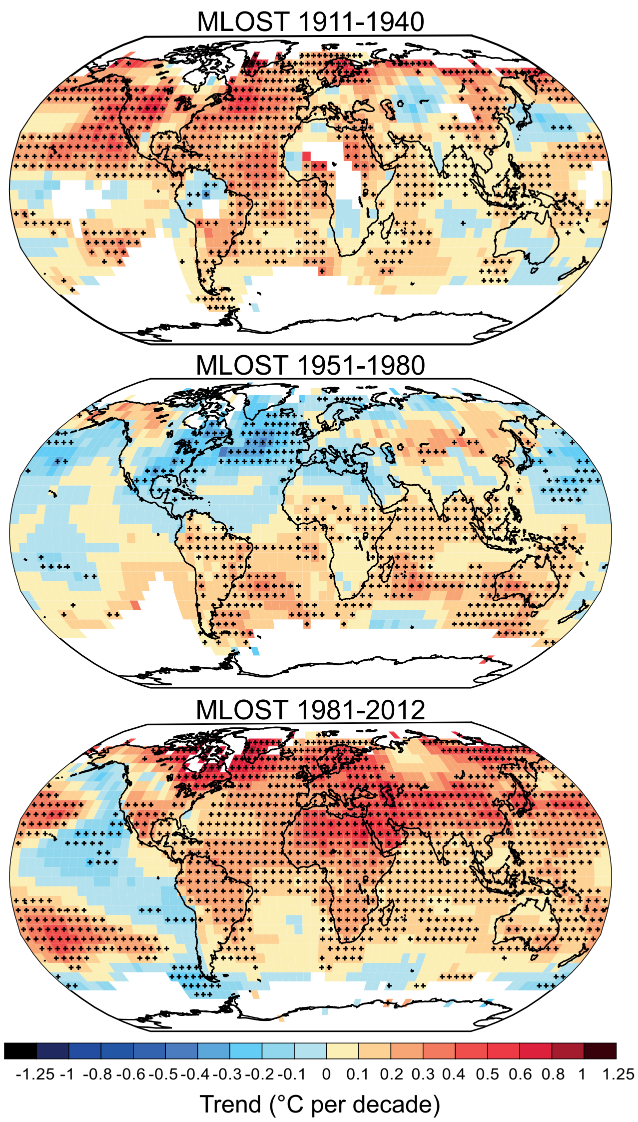


Figure 2.22 | Trends in surface temperature from NCDC MLOST for three non-consecutive shorter periods (1911–1940; 1951–1980; 1981–2012). White areas indicate incomplete or missing data. Trends and significance have been calculated as in Figure 2.21.

(Easterling and Wehner, 2009; Peterson et al., 2009; Liebmann et al., 2010; Foster and Rahmstorf, 2011; Santer et al., 2011). This issue is discussed in the context of model behaviour, forcings and natural variability in Box 9.2 and Section 10.3.1. Regardless, all global combined LSAT and SST data sets exhibit a statistically non-significant warming trend over 1998–2012 (0.042°C ± 0.093°C per decade (HadCRUT4); 0.037°C ± 0.085°C per decade (NCDC MLOST); 0.069°C ± 0.082°C per decade (GISS)). An average of the trends from these three data sets yields an estimated change for the 1998–2012 period of 0.05 [–0.05 to +0.15] °C per decade. Trends of this short length are very sensitive to the precise period selection with trends calculated in the same manner for the 15-year periods starting in 1995, 1996, and 1997 being 0.13 [0.02 to 0.24], 0.14 [0.03 to 0.24] and 0.07 [–0.02 to 0.18] (all °C per decade), respectively.

In summary, it is certain that globally averaged near surface temperatures have increased since the late 19th century. Each of the past three decades has been warmer than all the previous decades in the instrumental record, and the decade of the 2000s has been the warmest. The globally averaged combined land and ocean surface temperature data as calculated by a linear trend, show a warming of 0.85 [0.65 to 1.06] °C, over the period 1880–2012, when multiple independently produced datasets exist, about 0.89°C [0.69 to 1.08] °C over the period 1901–2012, and about 0.72 [0.49° to 0.89] °C over the period 1951–2012. The total increase between the average of the 1850–1900 period and the 2003–2012 period is 0.78 [0.72 to 0.85] °C and the total increase between the average of the 1850–1900 period and the reference period for projections 1986–2005 is 0.61 [0.55 to 0.67] °C, based on the single longest dataset available. For the longest period when calculation of regional trends is sufficiently complete (1901–2012), almost the entire globe has experienced surface warming. In addition to robust multi-decadal warming, global mean surface temperature exhibits substantial decadal and interannual variability. Owing to natural variability, trends based on short records are very sensitive to the beginning and end dates and do not in general reflect long-term climate trends. As one example, the rate of warming over the past 15 years (1998–2012; 0.05 [–0.05 to +0.15] °C per decade), which begins with a strong El Niño, is smaller than the rate calculated since 1951 (1951–2012; 0.12 [0.08 to 0.14] °C per decade)Trends for 15-year periods starting in 1995, 1996, and 1997 are 0.13 [0.02 to 0.24], 0.14 [0.03 to 0.24] and 0.07 [–0.02 to 0.18], respectively..

2.4.4 Upper Air Temperature

AR4 summarized that globally the troposphere had warmed at a rate greater than the GMST over the radiosonde record, while over the shorter satellite era the GMST and tropospheric warming rates were indistinguishable. Trends in the tropics were more uncertain than global trends although even this region was concluded to be warming. Globally, the stratosphere was reported to be cooling over the satellite era starting in 1979. New advances since AR4 have highlighted the substantial degree of uncertainty in both satellite and balloon-borne radiosonde records and led to some revisions and improvements in existing products and the creation of a number of new data products.

2.4.4.1 Advances in Multi-Decadal Observational Records

The major global radiosonde records extend back to 1958, with temperatures, measured as the balloon ascends, reported at mandatory pressure levels. Satellites have monitored tropospheric and lower stratospheric temperature trends since late 1978 through the Microwave Sounding Unit (MSU) and its follow-on Advanced Microwave Sounding Unit (AMSU) since 1998. These measures of upwelling radiation represent bulk (volume averaged) atmospheric temperature (Figure 2.23). The 'Mid-Tropospheric' (MT) MSU channel that most directly corresponds to the troposphere has 10 to 15% of its signal from both the skin temperature of the Earth's surface and the stratosphere. Two alternative approaches have been suggested for removing the stratospheric component based on differencing of view angles (LT) and statistical recombination (*G) with the 'Lower Stratosphere' (LS) channel (Spencer and Christy, 1992; Fu et al., 2004). The MSU satellite series also included a Stratospheric Sounding Unit (SSU) that measured at higher altitudes (Seidel et al., 2011).

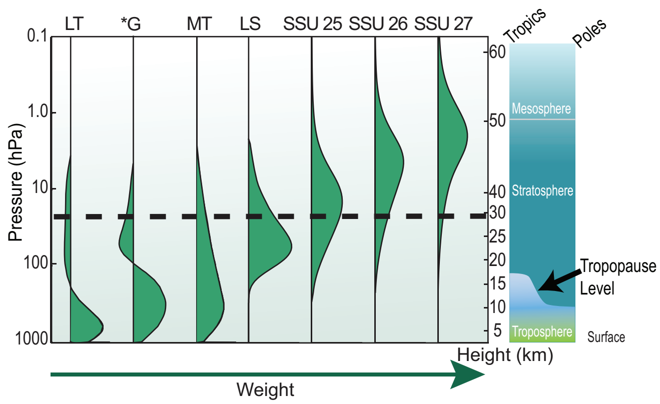


Figure 2.23 | Vertical weighting functions for those satellite temperature retrievals discussed in this chapter (modified from Seidel et al. (2011)). The dashed line indicates the typical maximum altitude achieved in the historical radiosonde record. The three SSU channels are denoted by the designated names 25, 26 and 27. LS (Lower Stratosphere) and MT (Mid Troposphere) are two direct MSU measures and LT (Lower Troposphere) and *G (Global Troposphere) are derived quantities from one or more of these that attempt to remove the stratospheric component from MT.

At the time of AR4 there were only two 'global' radiosonde data sets that included treatment of homogeneity issues: RATPAC (Free et al., 2005) and HadAT (Thorne et al., 2005). Three additional estimates have appeared since AR4 based on novel and distinct approaches. A group at the University of Vienna have produced RAOBCORE and RICH (Haimberger, 2007; Haimberger et al., 2008, 2012) using ERA reanalysis products (Box 2.3). Sherwood and colleagues developed an iterative universal kriging approach for radiosonde data to create IUK (Sherwood et al., 2008) and concluded that non-climatic data issues leading to spurious cooling remained in the deep tropics even after homogenization. The HadAT group created an automated version, undertook systematic experimentation and concluded that the parametric uncertainty (Box 2.1) was of the same order of magnitude as the apparent climate signal (McCarthy et al., 2008; Titchner et al., 2009; Thorne et al., 2011). A similar ensemble approach has also been applied to the RICH product (Haimberger et al., 2012). These various ensembles and new products exhibit more tropospheric warming / less stratospheric cooling than pre-existing products at all levels. Globally the radiosonde records all imply the troposphere has warmed and the stratosphere cooled since 1958 but with uncertainty that grows with height and is much greater outside the better-sampled NH extra-tropics (Thorne et al., 2011; Haimberger et al., 2012), where it is of the order 0.1°C per decade.

For MSU, AR4 considered estimates produced from three groups: UAH (University of Alabama in Huntsville); RSS (Remote Sensing Systems) and VG2 (now no longer updated). A new product has been created by NOAA labelled STAR, using a fundamentally distinct approach for the critical inter-satellite warm target calibration step (Zou et al., 2006a). STAR exhibits more warming/less cooling at all levels than UAH and RSS. For MT and LS, Zou and Wang (2010) concluded that this does not relate primarily to use of their inter-satellite calibration technique but rather differences in other processing steps. RSS also produced a parametric uncertainty ensemble (Box 2.1) employing a Monte Carlo approach allowing methodological inter-dependencies to be fully expressed (Mears et al., 2011). For large-scale trends dominant effects were inter-satellite offset determinations and, for tropospheric channels, diurnal drift. Uncertainties were concluded to be of the order 0.1°C per decade at the global mean for both tropospheric channels (where it is of comparable magnitude to the long-term trends) and the stratospheric channel.

SSU provides the only long-term near-global temperature data above the lower stratosphere, with the series terminating in 2006. Some AMSU-A channels have replaced this capability and efforts to understand the effect of changed measurement properties have been undertaken (Kobayashi et al., 2009). Until recently only one SSU data set existed (Nash and Edge, 1989), updated by Randel et al. (2009). Liu and Weng (2009) have produced an intermediate analysis for Channels 25 and 26 (but not Channel 27). Wang et al. (2012g), building on insights from several of these recent studies, have produced a more complete analysis. Differences between the independent estimates are much larger than differences between MSU records or radiosonde records at lower levels, with substantial inter-decadal time series behaviour departures, zonal trend structure, and global trend differences of the order 0.5°C per decade (Seidel et al., 2011; Thompson et al., 2012; Wang et al., 2012g). Although all SSU data sets agree that the stratosphere is cooling, there is therefore *low confidence* in the details above the lower stratosphere.

In summary, many new data sets have been produced since AR4 from radiosondes and satellites with renewed interest in satellite measurements above the lower stratosphere. Several studies have attempted to quantify the parametric uncertainty (Box 2.1) more rigorously. These various data sets and analyses have served to highlight the degree of uncertainty in the data and derived products.

2.4.4.2 Intercomparisons of Various Long-Term Radiosonde and MSU Products

Since AR4 there have been a large number of intercomparisons between radiosonde and MSU data sets. Interpretation is complicated, as most studies considered data set versions that have since been superseded. Several studies compared UAH and RSS products to local, regional or global raw/homogenized radiosonde data (Christy and Norris, 2006, 2009; Christy et al., 2007, 2010, 2011; Randall and Herman, 2008; Mears et al., 2012; Po-Chedley and Fu, 2012). Early studies focussed on the time of transition from NOAA-11 to NOAA-12 (early 1990s) which indicated an apparent issue in RSS. Christy et al. (2007) noted that this coincided with the Mt Pinatubo eruption and that RSS was the only product, either surface or tropospheric, that exhibited tropical warming immediately after the eruption when cooling would be expected. Using reanalysis data Bengtsson and Hodges (2011) also found evidence of a potential jump in RSS in 1993 over the tropical oceans. Mears et al. (2012) cautioned that an El Niño event quasi-simultaneous with Pinatubo complicates interpretation. They also highlighted several other periods of disagreement between radiosonde records and MSU records. All MSU records were most uncertain when satellite orbits are drifting rapidly (Christy and Norris, 2006, 2009). Mears et al. (2011) found that trend differences between RSS and other data sets could not be explained in many cases by parametric uncertainties in RSS alone. It was repeatedly cautioned that there were potential common biases (of varying magnitude) between the different MSU

2

records or between the different radiosonde records which complicate intercomparisons (Christy and Norris, 2006, 2009; Mears et al., 2012).

In summary, assessment of the large body of studies comparing various long-term radiosonde and MSU products since AR4 is hampered by data set version changes, and inherent data uncertainties. These factors substantially limit the ability to draw robust and consistent inferences from such studies about the true long-term trends or the value of different data products.

2.4.4.3 Additional Evidence from Other Technologies and Approaches

Global Positioning System (GPS) radio occultation (RO) currently represents the only self-calibrated SI traceable raw satellite measurements (Anthes et al., 2008; Anthes, 2011). The fundamental observation is time delay of the occulted signal's phase traversing the atmosphere. The time delay is a function of several atmospheric physical state variables. Subsequent analysis converts the time delay to temperature and other parameters, which inevitably adds some degree of uncertainty to the derived temperature data. Intercomparisons of GPS-RO products show that differences are largest for derived geophysical parameters (including temperature), but are still small relative to other observing technologies (Ho et al., 2012). Comparisons to MSU and radiosondes (Kuo et al., 2005; Ho et al., 2007, 2009a, 2009b; He et al., 2009; Baringer et al., 2010; Sun et al., 2010; Ladstadter et al., 2011) show substantive agreement in interannual behaviour, but also some multi-year drifts that require further examination before this additional data source can usefully arbitrate between different MSU and radiosonde trend estimates.

Atmospheric winds are driven by thermal gradients. Radiosonde winds are far less affected by time-varying biases than their temperatures (Gruber and Haimberger, 2008; Sherwood et al., 2008; Section 2.7.3). Allen and Sherwood (2007) initially used radiosonde wind to infer temperatures within the Tropical West Pacific warm pool region, then extended this to a global analysis (Allen and Sherwood, 2008) yielding a distinct tropical upper tropospheric warming trend maximum within the vertical profile, but with large uncertainty. Winds can only quantify relative changes and require an initialization (location and trend at that location) (Allen and Sherwood, 2008). The large uncertainty range was predominantly driven by this initialization choice, a finding later confirmed by Christy et al. (2010), who in addition questioned the stability given the sparse geographical sampling, particularly in the tropics, and possible systematic sampling effects amongst other potential issues. Initializing closer to the tropics tended to reduce or remove the appearance of a tropical upper tropospheric warming trend maximum (Allen and Sherwood, 2008; Christy et al., 2010). There is only *low confidence* in trends inferred from 'thermal winds' given the relative immaturity of the analyses and their large uncertainties.

In summary, new technologies and approaches have emerged since AR4. However, these new technologies and approaches either constitute too short a record or are too immature to inform assessments of long-term trends at the present time.

2.4.4.4 Synthesis of Free Atmosphere Temperature Estimates

Global-mean lower tropospheric temperatures have increased since the mid-20th century (Figure 2.24, bottom). Structural uncertainties (Box 2.1) are larger than at the surface but it can still be concluded that globally the troposphere has warmed (Table 2.8). On top of this long-term trend are superimposed short-term variations that are highly correlated with those at the surface but of somewhat greater amplitude. Global mean lower stratospheric temperatures have decreased since the mid-20th century punctuated by short-lived warming events associated with explosive volcanic activity (Figure 2.24a). However, since the mid-1990s little net change has occurred. Cooling rates are on average greater from radiosonde data sets than MSU data sets. This *very likely* relates to widely recognized cooling biases in radiosondes (Mears et al., 2006) which all data set producers explicitly caution are *likely* to remain to some extent in their final products (Free and Seidel, 2007; Haimberger et al., 2008; Sherwood et al., 2008; Thorne et al., 2011).

In comparison to the surface (Figure 2.22), tropospheric layers exhibit smoother geographic trends (Figure 2.25) with warming dominating cooling north of approximately 45°S and greatest warming in high northern latitudes. The lower stratosphere cooled almost everywhere but this cooling exhibits substantial large-scale structure. Cooling is greatest in the highest southern latitudes and smallest in high northern latitudes. There are also secondary stratospheric cooling maxima in the mid-latitude regions of each hemisphere.

Available global and regional trends from radiosondes since 1958 (Figure 2.26) show agreement that the troposphere has warmed and the stratosphere cooled. While there is little ambiguity in the sign of the changes, the rate and vertical structure of change are distinctly data set dependent, particularly in the stratosphere. Differences are greatest in the tropics and SH extra-tropics where the historical radiosonde data coverage is poorest. Not shown in the figure for clarity are estimates of parametric data set uncertainties or trend-fit uncertainties—both of which are of the order of at least 0.1°C per decade (Section 2.4.4.1).

Differences in trends between available radiosonde data sets are greater during the satellite era than for the full radiosonde period of record in all regions and at most levels (Figure 2.27; cf. Figure 2.26). The RAOBCORE product exhibits greater vertical trend gradients than other data sets and it has been posited that this relates to its dependency on reanalysis fields (Sakamoto and Christy, 2009; Christy et al., 2010). MSU trend estimates in the troposphere are generally bracketed by the radiosonde range. In the stratosphere MSU deep layer estimates tend to show slightly less cooling. Over both 1958–2011 and 1979–2011 there is some evidence in the radiosonde products taken as a whole that the tropical tropospheric trends increase with height. But the magnitude and the structure is highly data set dependent.

In summary, based on multiple independent analyses of measurements from radiosondes and satellite sensors it is *virtually certain* that globally the troposphere has warmed and the stratosphere has cooled since the mid-20th century. Despite unanimous agreement on the sign of the trends, substantial disagreement exists among available estimates as to the rate of temperature changes, particularly outside the NH extra-tropical troposphere, which has been well sampled by radiosondes.

Table 2.8 | Trend estimates and 90% confidence intervals (Box 2.2) for radiosonde and MSU data set global average values over the radiosonde (1958–2012) and satellite periods (1979–2012). LT indicates Lower Troposphere, MT indicates Mid Troposphere and LS indicates Lower Stratosphere (Figure 2.23. Satellite records start only in 1979 and STAR do not produce an LT product.

Data Set	Trends in °C per decade					
	1958–2012			1979–2012		
	LT	MT	LS	LT	MT	LS
HadAT2 (Thorne et al., 2005)	0.159 ± 0.038	0.095 ± 0.034	−0.339 ± 0.086	0.162 ± 0.047	0.079 ± 0.057	−0.436 ± 0.204
RAOBCORE 1.5 (Haimberger et al., 2012)	0.156 ± 0.031	0.109 ± 0.029	−0.186 ± 0.087	0.139 ± 0.049	0.079 ± 0.054	−0.266 ± 0.227
RICH-obs (Haimberger et al., 2012)	0.162 ± 0.031	0.102 ± 0.029	−0.285 ± 0.087	0.158 ± 0.046	0.081 ± 0.052	−0.331 ± 0.241
RICH-tau (Haimberger et al., 2012)	0.168 ± 0.032	0.111 ± 0.030	−0.280 ± 0.085	0.160 ± 0.046	0.083 ± 0.052	−0.345 ± 0.238
RATPAC (Free et al., 2005)	0.136 ± 0.028	0.076 ± 0.028	−0.338 ± 0.092	0.128 ± 0.044	0.039 ± 0.051	−0.468 ± 0.225
UAH (Christy et al., 2003)				0.138 ± 0.043	0.043 ± 0.042	−0.372 ± 0.201
RSS (Mears and Wentz, 2009a, 2009b)				0.131 ± 0.045	0.079 ± 0.043	−0.268 ± 0.177
STAR (Zou and Wang, 2011)					0.123 ± 0.047	−0.320 ± 0.175

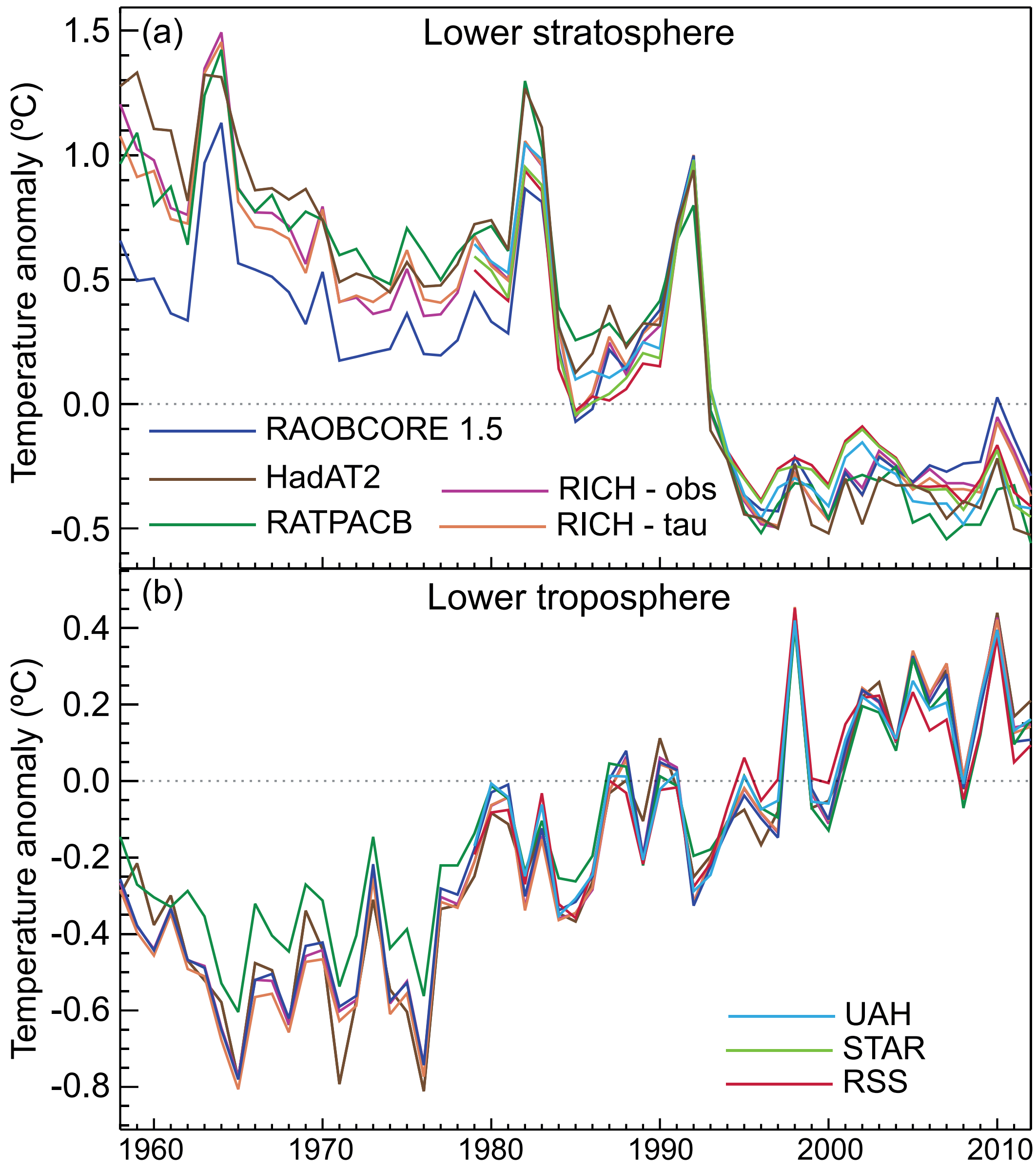


Figure 2.24 | Global annual average lower stratospheric (top) and lower tropospheric (bottom) temperature anomalies relative to a 1981–2010 climatology from different data sets. STAR does not produce a lower tropospheric temperature product. Note that the *y*-axis resolution differs between the two panels.

Docket No. E999/CI-14-643
DOC Ex. ___ KG-S-1
Page 12 of 15

Frequently Asked Questions

FAQ 2.1 | How Do We Know the World Has Warmed?

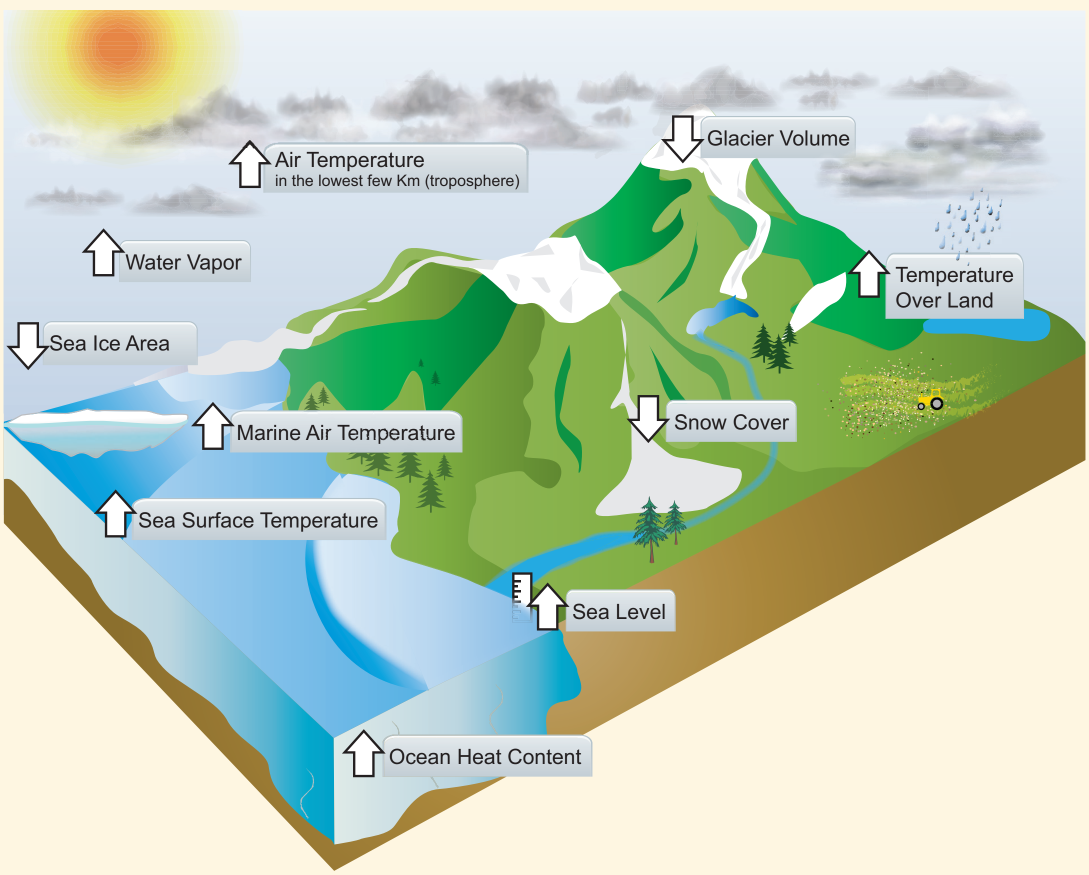
Evidence for a warming world comes from multiple independent climate indicators, from high up in the atmosphere to the depths of the oceans. They include changes in surface, atmospheric and oceanic temperatures; glaciers; snow cover; sea ice; sea level and atmospheric water vapour. Scientists from all over the world have independently verified this evidence many times. That the world has warmed since the 19th century is unequivocal.

Discussion about climate warming often centres on potential residual biases in temperature records from land-based weather stations. These records are very important, but they only represent one indicator of changes in the climate system. Broader evidence for a warming world comes from a wide range of independent physically consistent measurements of many other, strongly interlinked, elements of the climate system (FAQ 2.1, Figure 1).

A rise in global average surface temperatures is the best-known indicator of climate change. Although each year and even decade is not always warmer than the last, global surface temperatures have warmed substantially since 1900.

Warming land temperatures correspond closely with the observed warming trend over the oceans. Warming oceanic air temperatures, measured from aboard ships, and temperatures of the sea surface itself also coincide, as borne out by many independent analyses.

The atmosphere and ocean are both fluid bodies, so warming at the surface should also be seen in the lower atmosphere, and deeper down into the upper oceans, and observations confirm that this is indeed the case. Analyses of measurements made by weather balloon radiosondes and satellites consistently show warming of the troposphere, the active weather layer of the atmosphere. More than 90% of the excess energy absorbed by the climate system since at least the 1970s has been stored in the oceans as can be seen from global records of ocean heat content going back to the 1950s. *(continued on next page)*



FAQ 2.1, Figure 1 | Independent analyses of many components of the climate system that would be expected to change in a warming world exhibit trends consistent with warming (arrow direction denotes the sign of the change), as shown in FAQ 2.1, Figure 2.

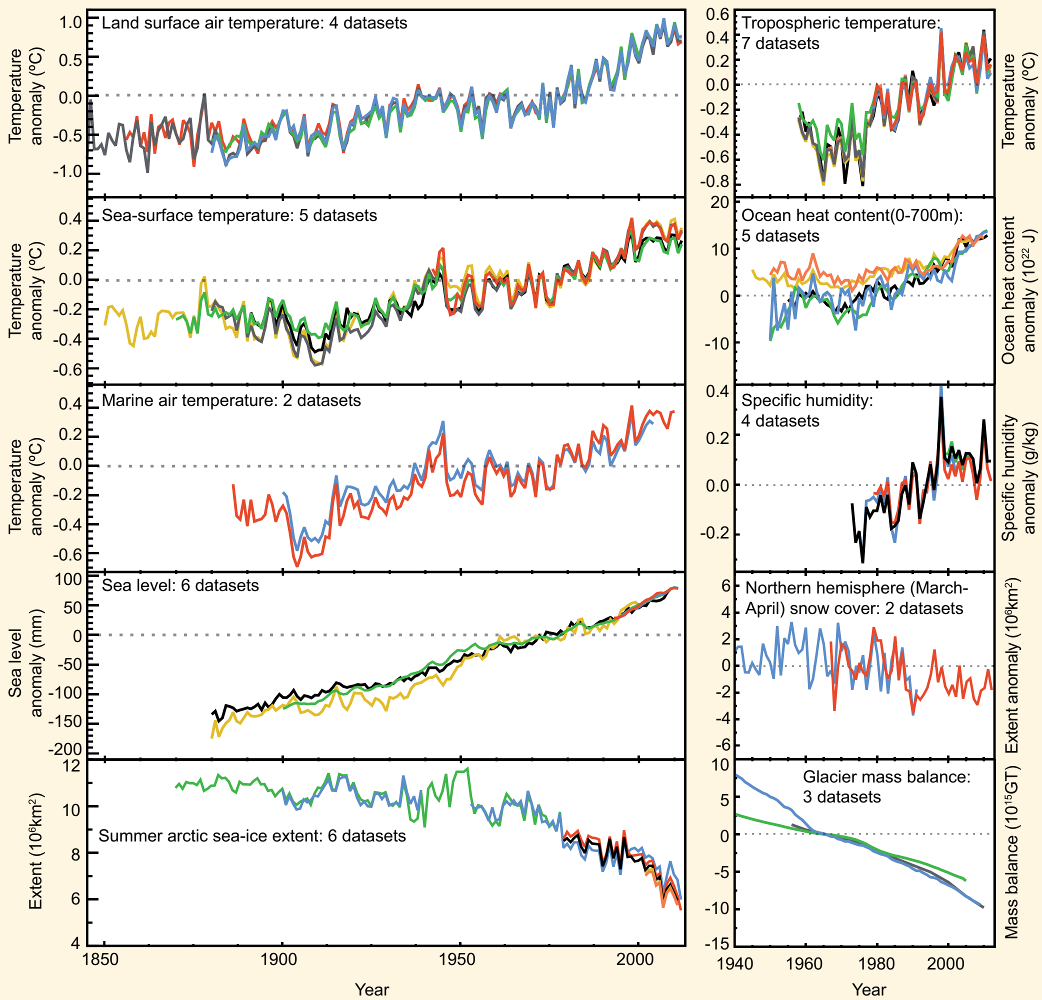
FAQ 2.1 (continued)

As the oceans warm, the water itself expands. This expansion is one of the main drivers of the independently observed rise in sea levels over the past century. Melting of glaciers and ice sheets also contribute, as do changes in storage and usage of water on land.

A warmer world is also a moister one, because warmer air can hold more water vapour. Global analyses show that specific humidity, which measures the amount of water vapour in the atmosphere, has increased over both the land and the oceans.

The frozen parts of the planet—known collectively as the cryosphere—affect, and are affected by, local changes in temperature. The amount of ice contained in glaciers globally has been declining every year for more than 20 years, and the lost mass contributes, in part, to the observed rise in sea level. Snow cover is sensitive to changes in temperature, particularly during the spring, when snow starts to melt. Spring snow cover has shrunk across the NH since the 1950s. Substantial losses in Arctic sea ice have been observed since satellite records began, particularly at the time of the mimimum extent, which occurs in September at the end of the annual melt season. By contrast, the increase in Antarctic sea ice has been smaller.

Individually, any single analysis might be unconvincing, but analysis of these different indicators and independent data sets has led many independent research groups to *all* reach the same conclusion. From the deep oceans to the top of the troposphere, the evidence of warmer air and oceans, of melting ice and rising seas all points unequivocally to one thing: the world has warmed since the late 19th century (FAQ 2.1, Figure 2).



FAQ 2.1, Figure 2 | Multiple independent indicators of a changing global climate. Each line represents an independently derived estimate of change in the climate element. In each panel all data sets have been normalized to a common period of record. A full detailing of which source data sets go into which panel is given in the Supplementary Material 2.SM.5.

Docket No. E999/CI-14-643
DOC Ex. ___ KG-S-1
Page 14 of 15

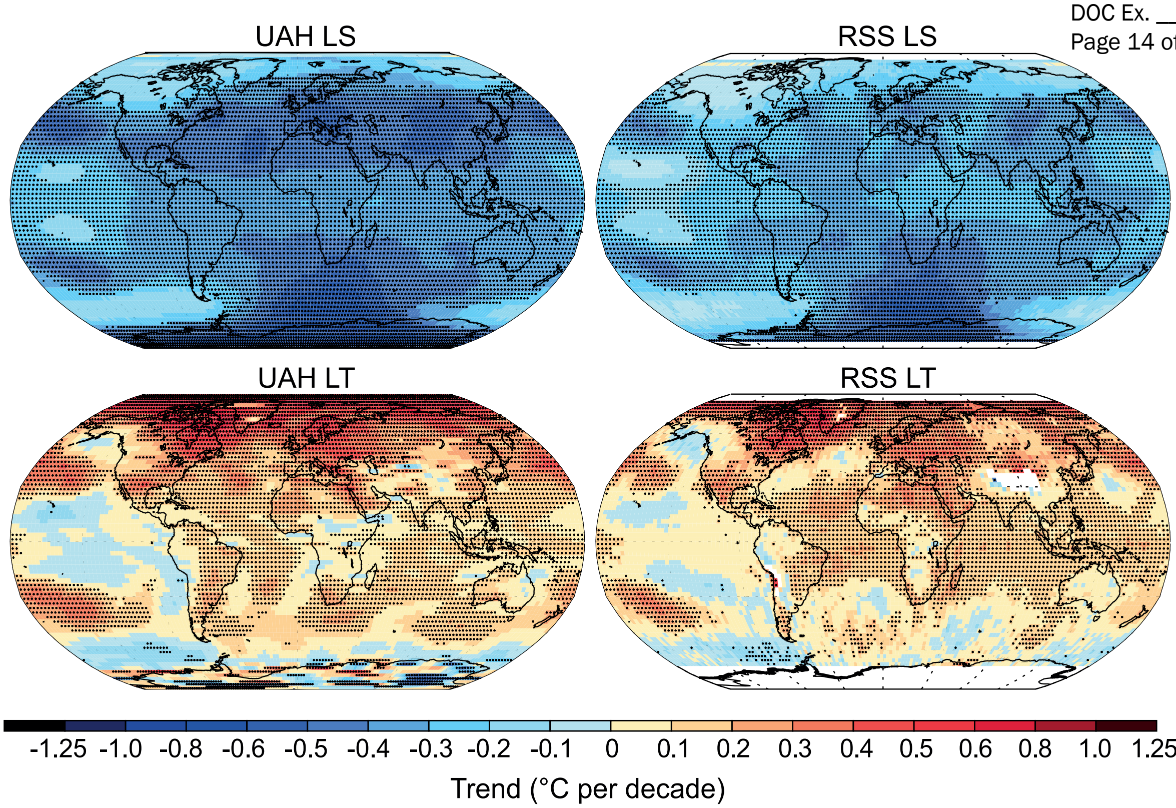


Figure 2.25 | Trends in MSU upper air temperature over 1979–2012 from UAH (left-hand panels) and RSS (right-hand panels) and for LS (top row) and LT (bottom row). Data are temporally complete within the sampled domains for each data set. White areas indicate incomplete or missing data. Black plus signs (+) indicate grid boxes where trends are significant (i.e., a trend of zero lies outside the 90% confidence interval).

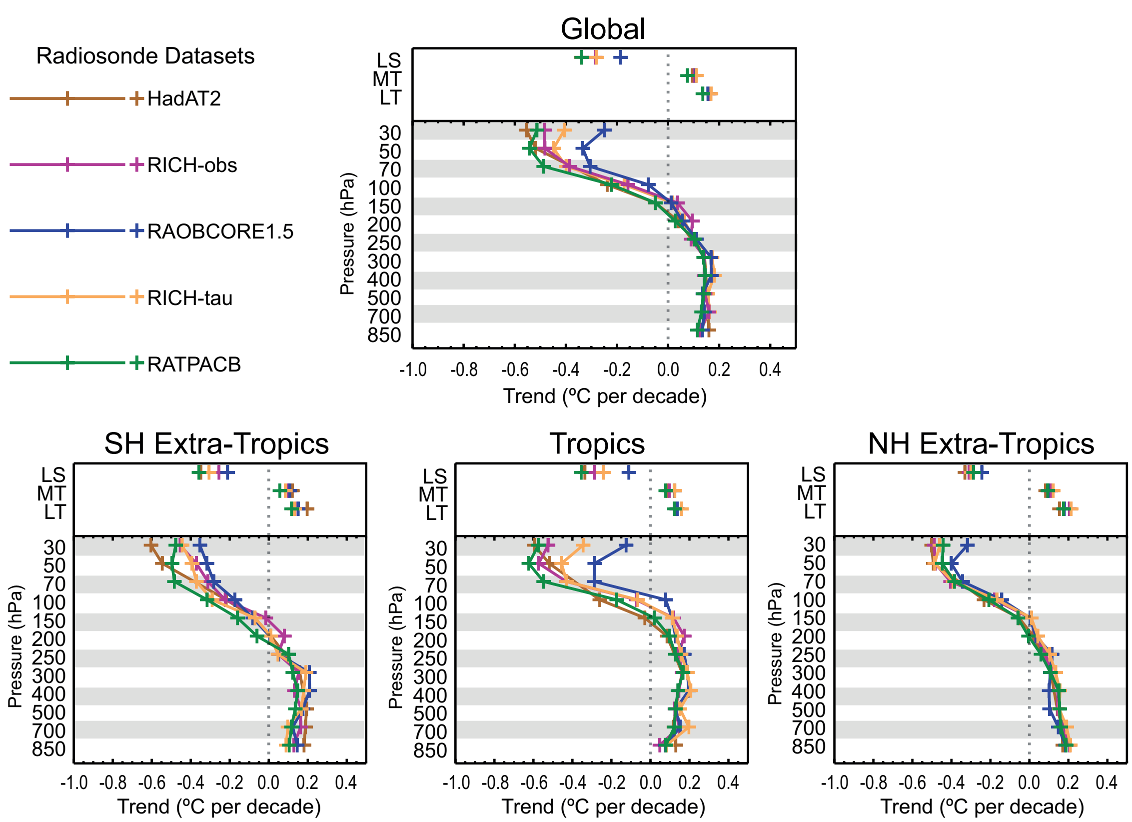


Figure 2.26 | Trends in upper air temperature for all available radiosonde data products that contain records for 1958–2012 for the globe (top) and tropics (20°N to 20°S) and extra-tropics (bottom). The bottom panel trace in each case is for trends on distinct pressure levels. Note that the pressure axis is not linear. The top panel points show MSU layer equivalent measure trends. MSU layer equivalents have been processed using the method of Thorne et al. (2005). No attempts have been made to sub-sample to a common data mask.

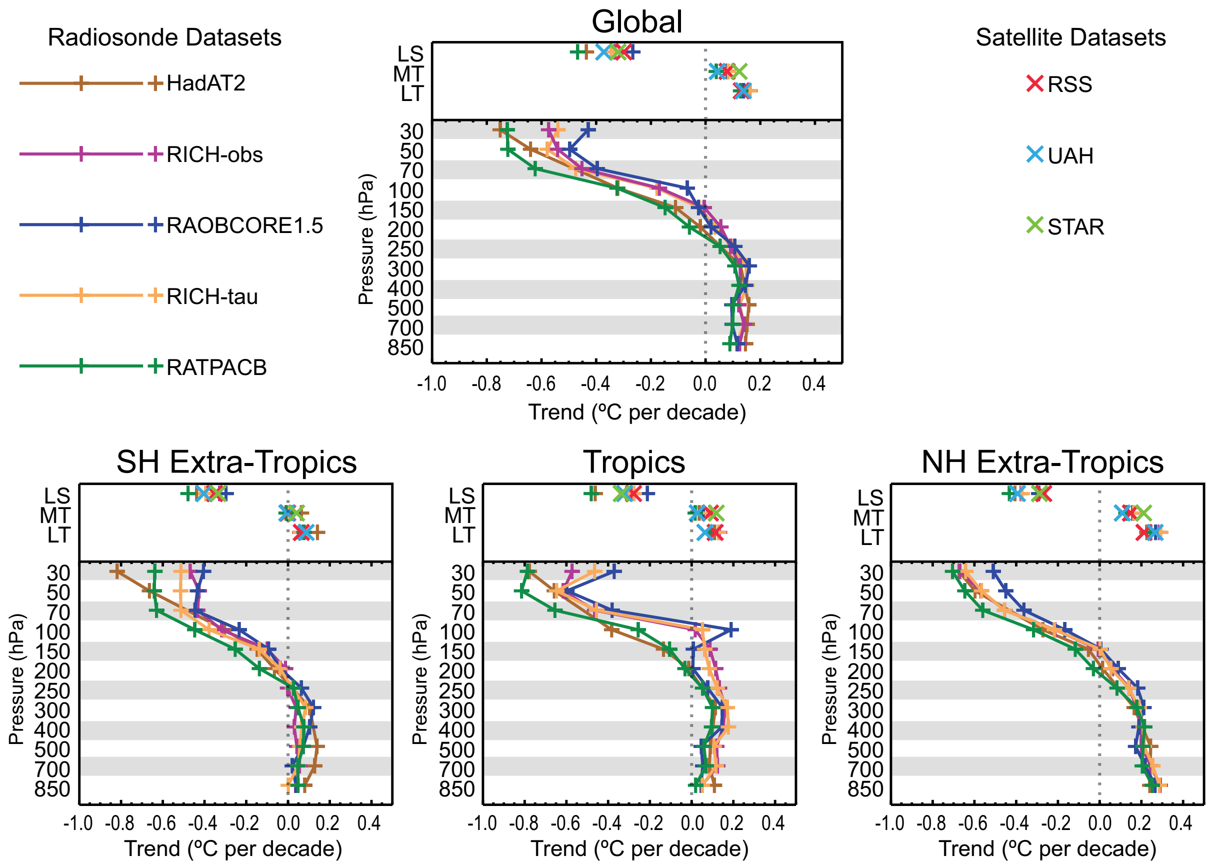


Figure 2.27 | As Figure 2.26 except for the satellite era 1979–2012 period and including MSU products (RSS, STAR and UAH).

Hence there is only *medium confidence* in the rate of change and its vertical structure in the NH extratropical troposphere and *low confidence* elsewhere.

2.5 Changes in Hydrological Cycle

This section covers the main aspects of the hydrological cycle, including large-scale average precipitation, stream flow and runoff, soil moisture, atmospheric water vapour, and clouds. Meteorological drought is assessed in Section 2.6. Ocean precipitation changes are assessed in Section 3.4.3 and changes in the area covered by snow in Section 4.5.

2.5.1 Large-Scale Changes in Precipitation

2.5.1.1 Global Land Areas

AR4 concluded that precipitation has generally increased over land north of 30°N over the period 1900–2005 but downward trends dominate the tropics since the 1970s. AR4 included analysis of both the GHCN (Vose et al., 1992) and CRU (Mitchell and Jones, 2005) gauge-based precipitation data sets for the globally averaged annual precipitation over land. For both data sets the overall linear trend from 1900 to 2005 (1901–2002 for CRU) was positive but not statistically significant (Table 3.4 from AR4). Other periods covered in AR4 (1951–2005 and 1979–2005) showed a mix of negative and positive trends depending on the data set.

Since AR4, existing data sets have been updated and a new data set developed. Figure 2.28 shows the century-scale variations and trends on globally and zonally averaged annual precipitation using five data sets: GHCN V2 (updated through 2011; Vose et al., 1992), Global Precipitation Climatology Project V2.2 (GPCP) combined raingauge–satellite product (Adler et al., 2003), CRU TS 3.10.01 (updated from Mitchell and Jones, 2005), Global Precipitation Climatology Centre V6 (GPCC) data set (Becker et al., 2013) and a reconstructed data set by Smith et al. (2012). Each data product incorporates a different number of station series for each region. The Smith et al. product is a statistical reconstruction using Empirical Orthogonal Functions, similar to the NCDC MLOST global temperature product (Section 2.4.3) that does provide coverage for most of the global surface area although only land is included here. The data sets based on *in situ* observations only start in 1901, but the Smith et al. data set ends in 2008, while the other three data sets contain data until at least 2010.

For the longest common period of record (1901–2008) all datasets exhibit increases in globally averaged precipitation, with three of the four showing statistically significant changes (Table 2.9). However, there is a factor of almost three spread in the magnitude of the change which serves to create *low confidence*. Global trends for the shorter period (1951–2008) show a mix of statistically non-significant positive and negative trends amongst the four data sets with the infilled Smith et al. (2012) analysis showing increases and the remainder decreases. These differences among data sets indicate that long-term increases