Problems with the peer review underpinning the EPA Endangerment Finding

Ross McKitrick
Professor of Economics,
University of Guelph
ross.mckitrick@uoguelph.ca

November 10, 2017

Note: This report was prepared in 2017 in response to requests from various individuals for my input into the question of how or whether the Endangerment Finding should be reconsidered. It draws on material compiled earlier during the Endangerment Finding Process as well as a new survey of issues related to the Inspector General's report and the EPA's handling thereof which I undertook for the purpose of writing an op-ed.[1]

This document available online at rossmckitrick.com/submissionsresponses-to-govt-inquiries.html.

[1] https://www.nationalreview.com/2017/12/epa-clean-power-plan-basis-rules-may-be-faulty/

Introduction

This report outlines a case for re-doing the science behind the Endangerment Finding (EF). The main argument is that the EPA did not comply with orders from its Inspector General to classify the Technical Support Document (TSD) as a Highly Influential Scientific Assessment (HISA), which would have precluded relying on outside agencies like the IPCC to do the science review. The EPA specifically claimed its TSD was not a HISA, which precludes using it as a basis for rule-making costing more than $500 million. Since such rulemaking was clearly envisioned by the EPA, their position was dishonest and appears to have been contrived to avoid transparency and requirements for scientific rigour. The EPA position on endangerment cannot be supported with reference to assessments by the IPCC, NRC or GCRP either because such reports do not meet relevant OMB peer review requirements, and/or they do no show what the EPA claims they do.

The sections of this report are as follows.

1 Reliance on the IPCC review process in place of review conducted under EPA auspices.

Rebuttal: The IPCC review process does not comply with HISA requirements. Since the NRC and USGCRP reports rely on IPCC conclusions they are likewise deficient. The IPCC review process is compromised by arbitrary author selection procedures, conflicts of interest in which IPCC-selected authors review their own work and that of their critics, weak peer review procedures, the lack of a requirement to document the full range of opposing views, and a failure to fully disclose the outcome of its review process. The EPA Office of Inspector General (OIG) report set out peer review requirements required by the Information Quality Act (also known as the Data Quality Act (DQA)) that unambiguously disqualify IPCC products for the proposed rule. To get around this obstacle the EPA asserted that the Technical Support Document (TSD) on which it relied was not a "Highly Influential Scientific Assessment", a position that is preposterous on its face and clearly untrue in light of the relevant OMB criteria. To the extent it remains the EPA position, the EPA cannot rely on the TSD or the EF for any regulatory action costing more than $500 million. Since the EPA has already promulgated regulations costing more than that, the Agency's position is dishonest and counter to the goals of transparency and scientific rigour. The EF clearly needs to undergo a new HISA-compliant peer review.

1.1 Relevance: EPA admits the science behind the TSD was not peer-reviewed

When the EPA released its Proposed Endangerment Finding on greenhouse gases in April 2009[2] and its final Endangerment Finding in December 2010[3] it did not conduct any internal evaluation of the science, instead it relied on the IPCC assessment:

> However, the [EPA] Administrator is relying on the major assessments of the USGCRP, IPCC, and NRC as the primary scientific and technical basis of her endangerment decision for a number of reasons. …these assessment reports undergo a rigorous and exacting standard of peer review by the expert community, as well as rigorous levels of U.S. government review and acceptance. Individual studies that appear in scientific journals, even if peer reviewed, do not go through as

[2] US Federal Register 74 FR 18886.

[3] http://www.epa.gov/climatechange/endangerment/downloads/Federal_Register-EPA-HQ-OAR-2009-0171-Dec.15-09.pdf

> many review stages, nor are they reviewed and commented on by as many scientists. The review processes of the IPCC, USGCRP, and NRC (explained in fuller detail in the TSD and the Response to Comments document, Volume 1) provide EPA with strong assurance that this material has been well vetted by both the climate change research community and by the U.S. government.
>
> (US Federal Register 74 page 66510-66511).

This promotional description of the IPCC review process is inaccurate. As I will explain below, Lead Authors of Working Group reports have arbitrary authority to override reviewer comments, and even to rewrite the text after the review process has closed. Consequently the review process is much weaker than that which occurs in normal academic journals, where neither of these practices are allowed.

The OMB defines a Highly-Influential Scientific Assessment as follows:

> III. 1. Applicability: This section applies to influential scientific information that the agency or the Administrator determines to be a scientific assessment that:
>
> (i) Could have a potential impact of more than $500 million in any year, or
>
> (ii) Is novel, controversial, or precedent-setting or has significant interagency interest.[4]

In September 2010 the Office of the Inspector General (OIG) of the EPA issued a finding[5] that the TSD was a Highly Influential Scientific Assessment (HISA) and therefore the EPA was required to conduct its own peer review, rather than relying on the review undertaken by the IPCC. In 2011 the OIG directed the EPA to take corrective action but the EPA did not do so. Instead both the EPA and the OMB took the position that the TSD was not a HISA. The OIG rejected this position as being obviously deficient.[6] The EPA had told the OIG:

> the TSD consisted only of science that was previously peer reviewed and that these reviews were deemed adequate under the Agency's policy. (OIG Report p. 13)

---

[4] Federal Register https://www.ssa.gov/515/PeerReviewsFedRegNoticeForFinalBulletin.pdf p. 2675
[5] http://www.epa.gov/oig/reports/2011/20110926-11-P-0702.pdf
[6] The exchange is chronicled at https://climateaudit.org/2011/10/04/epa-the-endangerment-finding-was-not-a-highly-influential-scientific-assessment/

But it can be shown that the IPCC review process is not adequate under the Agency's policy. The OIG report (pp. 9—10) lists five criteria the EPA uses when assessing the quality of its scientific and technical information.

a. **Soundness**: The extent to which the scientific and technical procedures, measures, methods, or models employed to generate the information are reasonable for, and consistent with, the intended application.

b. **Applicability and utility**: The extent to which the information is relevant for the Agency's intended use.

c. **Clarity and completeness**: The degree of clarity and completeness with which the data, assumptions, methods, quality assurance, sponsoring organizations, and analyses employed to generate the information are documented.

d. **Uncertainty and variability**: The extent to which the variability and uncertainty (quantitative and qualitative) in the information, or in the procedures, measures, methods, or models, are evaluated and characterized.

e. **Evaluation and review:** The extent of independent verification, validation, and peer review of the information or of the procedures, measures, methods, or models.

This report in its entirety will show that the science used for the TSD fails to meet these criteria. Sections 1.2 and 1.3 show item e (Evaluation and review) to be deficient, drawing on critical assessments of the IPCC by the Inter-Academy Council and others. Item a (Soundness) will be shown to be deficient in Sections 2, 3, 5, 6, 7, 8 and 9. Item d (Uncertainty and variability) will be shown to be deficient in Section 4.

The OIG report (p. 11) also listed a key condition for evaluating whether another party's peer review processes is adequate for a highly influential scientific assessment. The EPA:

> . . . should examine closely the particulars of the peer review to ensure *independence and a conscious effort to incorporate the peer reviewers' comments into the final work product*. If there are perceived, or real, conflicts of interest, this may preclude the use of that peer review and, in those instances, another peer review would be needed.

(emphasis added). Section 1.2 below shows that the IPCC selects authors who are not independent and who are typically in conflicts of interest in the sense that they review their own work and that of their critics. It will also show that IPCC procedures allow Lead Authors arbitrary authority to ignore reviewer comments and rewrite text after the close of peer review. Sections 3 and 4 will give specific examples of how influential sections of the IPCC report were changed after the close of peer review so as to avoid incorporating reviewer comments in the final work product, even when they had been acknowledged to be correct.

In 2014 the EPA denied various petitions to reopen the Endangerment Finding. This report will present information that the EPA did not consider when making this finding. The EPA relied on backchannel information from NASA's Gavin Schmidt to EPA staffers to Jason Samenow, Marcus Sarofim and Rona Birnbaum, that was not included in the peer-review record and was not subject to independent cross-examination.[7]

1.2 IPCC Structure

As pointed out by Stephen McIntyre in his submission to the EPA in response to the release of the TSD[8], since the IPCC is an international organization it does not operate under any national-level transparency or Freedom of Information obligations, nor is it bound by US data quality requirements. To understand the many deficiencies of the IPCC review process it is necessary first to understand its structure. There are three administrative tiers in the IPCC.[9]

(i) The top level is called the Panel, consisting of representatives of the 195 member states, who meet in periodic plenary sessions to make decisions and review ongoing work. The documentary record shows[10] that Panel members provide only cursory and superficial input into IPCC operations, few members participate in the Assessment review process and most were not engaged with the recent reform process. For all practical purposes, the IPCC is directed and controlled by the IPCC Bureau.

---

[7] This matter is explained at https://climateaudit.org/2014/10/18/gavin-schmidt-and-the-epa-denial-decision/.
[8] McIntyre, Stephen, EPA-HQ-OAR-2009-0171.
Online at http://www.climateaudit.info/pdf/McIntyre_Submission_to_EPA.pdf
[9] This is based on the IPCC organizational chart at http://www.ipcc.ch/organization/organization_structure.shtml.
[10] See McKitrick, Ross R. "What Is Wrong With The IPCC? Proposals For a Radical Reform." Global Warming Policy Foundation Report 4, 2010. This and the next 3 sections are based on that document.

(ii) The IPCC Bureau (assisted by a 10-member IPCC Secretariat), is an administrative body elected by the Panel, consisting of a Chair, Vice Chairs, the Working Group Co-Chairs, and other Bureau members. The Bureau, as of the date of the TSD publication, consisted of 30 members elected at a meeting of the Panel in Geneva in September 2008.[11] 28 members were attached to the three Working Groups and 2 were Co-Chairs of the Task Force on Greenhouse Gas Inventories, which has its own 14-member Bureau. The Bureau and Secretariat have significant influence over the flow of information to the Panel, by structuring and presiding over the plenary meetings and overseeing the production of reports.

(iii) The next tier is divided into three Working Groups and one Task Force, where the work of preparing reports is conducted. Working Group Lead Authors are selected by the Bureau. Each Working Group produces a contribution to an assessment report, commonly known as IPCC Reports.

1.2.1 Arbitrary Author Selection and Conflict of Interest

The Bureau has complete control over the selection of Coordinating Lead Authors (CLAs) and Lead Authors (LAs) for the Working Groups. The CLAs and LAs then select contributing authors (CAs) at their own discretion to provide content to the chapters. While the Bureau recruits CLAs and LAs from author lists provided by member governments, it is not limited to using the names on those lists, instead it is allowed under IPCC rules to select anyone it wants.

One way that the IPCC's control over author selection leads to biases in the assessment process is that authors of key sections are typically in the position of assessing their own work and that of their critics. Some examples include:

- In the 2001 IPCC Report, Michael Mann was Lead Author of the paleoclimate chapter that assessed his own hockey stick graph and that of rival teams, and he steered the decision to give it prominence and suppress contradictory information in another graph by Briffa et al (see Section 5).
- In the 2007 IPCC Report, Phil Jones of the UK Climate Research Unit (CRU) was Lead Author of the chapter that assessed, among other things, the quality of CRU temperature data, and the findings of teams that had showed it is contaminated with a warm bias. He kept the critical information out of drafts shown to reviewers and then participated in inserting text after the

[11] See http://www.ipcc.ch/meetings/session30/doc5.pdf.

close of peer review that dismissed this evidence on the basis of a fabricated statistical test (see Section 3).

- The Lead Authors of IPCC 2007 Chapter 9 that critically assess the findings of "signal detection" literature, on which basis the attribution of climate change to GHG's is made, are themselves the authors of most of the signal detection studies on which their conclusions rest. This includes Gabriele Hegerl, Francis Zwiers, Peter Stott, Nathan Gillett, Myles Allan, Richard Betts, Reto Knutti and Simon Tett.

Lead Author selection is done behind closed doors using an opaque process that was much-criticized during the review of IPCC procedures conducted by the InterAcademy Council (IAC) in 2010.[12] The author selection criteria have since been revised slightly, without introducing any substantial changes, in response to the IAC review recommendations. There is a requirement to ensure representation of a wide range of views, but it is worded so weakly that it is in effect a dead letter:

> The composition of the group of Coordinating Lead Authors and Lead Authors for a section or chapter of a Report shall reflect the need to aim for a range of views, expertise and geographical representation.

In May 2011 the Panel responded to criticism by changing the wording from "shall reflect the need to aim for a range of views" to "shall aim to reflect a range of scientific, technical and socio-economic views,"[13] which is clearly a trivial change.

The centralized nature of the author selection process, and the absence of a meaningful requirement to include proponents of the full range of scientific views, means that the IPCC Bureau can predetermine the content of the report by selecting of CLAs and LAs they know to be committed to a particular point of view. Past IPCC Chairman Rajendra Pachauri denies that the author selection procedure is biased. In a 2007 interview he described the process in very idealized terms:

> These are people who have been chosen on the basis of their track record, on their record of publications, on the research that they have done. …They are people who are at the top of their profession as far as research is concerned in a particular aspect of climate change.[14]

---

[12] http://reviewipcc.interacademycouncil.net/
[13] See http://www.ipcc.ch/meetings/session33/ipcc_p33_decisions_taken_procedures.pdf p. 2.
[14] http://www.rediff.com/news/2007/jun/05inter.htm

But it is easy to find counterexamples that undermine this description.[15] One example is Sven Teske, a climate campaigner for Greenpeace who was selected by the IPCC as a Lead Author for its recent report on renewable energy (SRREN), which led to a non peer-reviewed Greenpeace report he coauthored becoming the basis for central claims in the report, which were subsequently highlighted in the press release announcing its publication.

Another particularly notable case is Sari Kovats, who was selected to serve as an IPCC Contributing Author in 1994 when she was 25 years old, had no Ph.D. and no academic publications, and was just starting a job as a research assistant at the London School of Hygiene and Tropical Medicine.[16] She began a part-time Ph.D. program in 2001, at which time she was promoted to a term as an IPCC Lead Author. The IPCC Bureau appointed her a third time as Lead or Contributing Author for a total of four chapters of the AR4, as well as expert reviewer. Her Ph.D. thesis wasn't completed until three years after the AR4 was published.[17]

These are not isolated cases. Past IPCC authors made many submissions to the IAC Review panel,[18] expressing concerns about the extent to which LAs are selected on political rather than scientific grounds. A common complaint was that the mandate to obtain geographic balance led to inclusion of many incompetent and untrained scientists, and political considerations often seemed to rank above scientific credentials.[19] Here are some excerpts from complaints filed by IPCC Lead Authors themselves about some of the people they were teamed with.

> There are far too many politically correct appointments, so that developing country scientists are appointed who have insufficient scientific competence to do anything useful. This is reasonable if it is regarded as a learning experience, but in my chapter in AR4 we had half of the [lead authors] who were not competent.
>
> The most important problem of the IPCC is the nomination and selection of authors and Bureau Members. Some experts are included or excluded because of their political allegiance rather than their academic quality. Sometimes, the "right" authors are put in key positions with generous

[15] The material on lead author selection is drawn from Laframboise, Donna (2011) *The Delinquent Teenager Who Was Mistaken for the World's Top Climate Scientist* Toronto: Amazon.
[16] http://www.webcitation.org/5xEHr8hDh
[17] See Donna Laframboise, http://nofrakkingconsensus.com/2011/03/16/the-strange-case-of-sari-kovats/ for a more detailed examination of the circumstances of this author's appointment.
[18] Available online at http://reviewipcc.interacademycouncil.net/Comments.pdf.
[19] See Laframboise (2011, op. cit.) for more excerpts and discussion.

> government grants to support their IPCC work, while the "wrong" authors are sidelined to draft irrelevant chapters and sections without any support.
>
> The whole process… [is] flawed by an excessive concern for geographical balance. All decisions are political before being scientific.
>
> …it is clearly noticeable that the [author nomination] process occasionally brings authors with poor knowledge or poor motivation into [lead author] positions.
>
> … I have experienced the addition of lead authors or [contributing] authors during the process who often seem to come with a political mandate – generally from developed countries and as such they can be very disruptive – let alone the dubious nature of the science they contribute!
>
> Since I have been selected for several IPCC reports, I have no personal prejudice (or grouse) on the process. However, regarding the selection of Lead Authors, I am more worried since the distortions, opaqueness and arbitrariness that is lately creeping into the process seems alarming. It seems that knowledge and scientific contributions are increasingly at discount in selection of authors compared to the personal connections, affiliations and political accommodations.
>
> IPCC works hard for geographic diversity. This is one valuable criterion, but it is not sufficient to choose a lead author. The result is that some of the lead authors (generally although not always from developing countries) are clearly not qualified to be lead authors and are unable to contribute in a meaningful way to the writing of the chapter.
>
> The team members from the developing countries (including myself) were made to feel welcome and accepted as part of the team. In reality we were out of our intellectual depth as meaningful contributors to the process.

These comments, and many more like them, came from past IPCC Lead Authors themselves, indicating that Pachauri's description of the author selection process is clearly untrue. He also failed to point out the most significant loophole in the process, namely that CLAs and LAs have a free hand in selecting Contributing Authors, who provide much of the text. The IPCC guidelines say only:[20]

[20] See http://www.ipcc.ch/pdf/ipcc-principles/ipcc-principles-appendix-a.pdf section 4.2.2.

> The Coordinating Lead Authors and Lead Authors selected by the Working Group/Task Force Bureau may enlist other experts as Contributing Authors to assist with the work.

This aspect of the process neutralizes the already weak requirements for balance, since no requirements for balance are imposed on the CA selection process—in fact no requirements of any kind are imposed on it. The IPCC does not even have to release the list of CAs during the report-writing process, the rules only stipulate that CAs should be named in the final, published report.

1.2.2 Influence of the World Wildlife Fund

Another illustration of the problems in the IPCC author selection process is the influence given to authors who are members of or advisors to the environmental activist organization World Wildlife Fund (WWF)[21]:

- 28 out of 44 chapters of the AR4 Working Group Reports had at least one author who is a campaign advisor for the World Wildlife Fund (WWF);
- WWF campaign advisors were involved in writing all 20 chapters of the Working Group II report and 6 of 11 chapters of the Working Group I report;
- WWF campaign advisors served as Coordinating Lead Authors for 15 of the 44 chapters in the AR4, and in three cases both the CLA's were WWF advisors;
- In one chapter, 8 of the authors were WWF campaign advisors.

This remarkable overlap between the IPCC and a powerful environmental activist organization cannot credibly be viewed as mere coincidence, and instead reveals a political bias in the author selection process, confirming the complaints heard by the IAC on this matter.

1.3 IPCC Report-Writing Process

The IPCC writing procedures involve preparing a series of versions of the report. A first version (the so-called Zero Order Draft) is prepared by the LAs and CLAs, drawing upon contributions from CAs. This is worked up into the First Order Draft which is then sent out for Expert Review.

[21] This information is documented in Laframboise (2011).

Selection of Expert Reviewers is generally open, and people can nominate themselves. Review comments are sent to the IPCC Secretariat, which provides them to chapter authors. Review Editors are supposed to ensure that all comments are taken into account. After this the Second Order Draft is released for another round of expert review and a round of government review. After these comments are received the report is returned to the Lead Authors for another complete rewrite prior to submission to the IPCC Bureau. This final draft is not itself subject to expert review. Sections 3 and 4 below will give examples of how material changes to the text were introduced at this stage over the explicit objections of peer reviewers as expressed in previous review rounds.

1.3.1 Failure to produce review annexes

The IPCC rules state[22] that Review Editors should supply annexes that explain significant unresolved differences of opinion. **But no such annexes have ever been produced.** One of the most contentious disputes in the AR4 between LAs and reviewers concerned the paleoclimate chapter, in particular the hockey stick. There were deep, unresolved disagreements between reviewers and LAs on how this matter should be presented.[23] Yet no annex was produced, and the Review Editors signed off on the Chapter nonetheless. In 2008, a UK citizen named David Holland sought information about how one of the Chapter 6 Review Editors (John Mitchell of the UK Met Office) had handled the controversies over the Mann et al. hockey stick. As part of his inquiries, Holland submitted a Freedom of Information Act request to the Met Office. The documents released in reply contained an email from IPCC Chair Susan Solomon to Mitchell[24] advising Mitchell on the limitations of his responsibilities as Review Editor. The email was dated March 14, 2008, and stated, in part:

> The review editors do not determine the content of the chapters. The authors are responsible for the content of their chapters and responding to comments, not REs. Further explanations, elaboration, or re-interpretations of the comments or the author responses, would not be appropriate.

[22] See http://ipcc.ch/pdf/ipcc-principles/ipcc-principles-appendix-a.pdf section 5.
[23] Documented in Holland, David (2007) "Bias and Concealement in the IPCC Process: The "Hockey Stick" Affair and Its Implications." *Energy and Environment* Volume 18, Numbers 7-8, December 2007 , pp. 951-983(33).
[24] Contained in emails from IPCC Review Editor Brian Hoskins that were released to Mr. David Holland in response to a UK Freedom of Information Act request, ICO decision notice FER0239225.

1.3.2 Less rigorous than academic journal peer review

Taking these points together, it is clear that RE's do not have the authority to stop Lead Authors who are determined to make arbitrary decisions about chapter content. For this reason the IPCC review process is fundamentally unlike the academic peer review process, in which the editor has the final right to accept or reject a paper and its contents based on review comments. Because the IPCC gives Lead Authors the final right to determine the content of their own chapters the EPA Administrator was wrong to declare the IPCC review process to be "rigorous and exacting," and the criteria cited by the EPA OIG are not met, implying that "this may preclude the use of that peer review and, in those instances, another peer review would be needed." Further evidence on this point will emerge in the subsequent sections.

2 Reliance on global surface temperature data products .

Rebuttal: The EPA relied heavily on the IPCC handling of questions about the soundness of surface temperature data. The IPCC ignored published evidence of a warm bias in CRU data, conspired to keep the evidence out of the AR4, and fabricated evidence to conceal the problem. The IPCC acknowledged in a report published after the EF that it had made an important judgment about its data products with no supporting evidence. Subsequent literature has confirmed that the land record is likely contaminated with a warm bias. The TSD thus relied on incomplete and fabricated claims which have since been refuted in peer-reviewed literature. The TSD also relied heavily on a single study (Schmidt, GA 2009 Int J Clim 10.1002/joc.1831) subsequent to the AR4 that has likewise been refuted.

2.1 Relevance

The EPA relied on IPCC data and modeling work as the basis of its scientific findings. They stated (April 2009 document, p. 59):

> Most of the observed increase in global average temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations. Global observed temperatures over the last century can be reproduced only when model simulations include both natural and anthropogenic forcings, that is, simulations that remove anthropogenic forcings are unable to reproduce observed temperature changes.

This “line of evidence” was part of the justification for the proposed rule.

This statement pre-supposes that there are no biases or contamination problems in the surface temperature record. The argument states that climate models running with only natural processes active cannot accurately reproduce the global average surface temperature series for the 20th century, but when a strong response to rising greenhouse gas levels is incorporated in the models, they generate a global average temperature that matches observed data much more closely. This only indicates that greenhouse gases are the cause of climate change if there are no other plausible changes that, were they to be programmed into the climate model, would also yield a match. Most of the observed warming in surface temperature data occurred over land, the records for which are known to be affected by urbanization and other forms of regional contamination, so the effects of such processes should be tested as well. But the IPCC does not do so, on the assumption that all such effects have been filtered out of the surface temperature data by the application of statistical “adjustments.” If this assumption is invalid, then the IPCC’s claims about both the magnitude and cause of surface warming cannot be accepted on the basis of the evidence they cite.

2.2 Lack of documentation establishing reliability of the surface temperature record

CRU, GISS and NOAA all publish global average temperature series over land based largely on the same underlying data source, the Global Historical Climatology Network (GHCN). It has been known for many decades that temperatures at land-based observational sites can be affected by changes in the land surface due to local deforestation, introduction of agriculture, road-building, urbanization, changes in monitoring equipment, measurement discontinuities, and so forth; as well as by local emissions of particulates and other air pollutants. These are considered *non-climatic* influences, since they cause purely local, and in principle reversible, changes in regional temperatures. Hence they must be filtered out of the local temperature record in order to reveal the *climatic* record. An ideal measurement of surface climatic changes would require a monitoring site untouched by human development, the equipment for which was consistent and perfectly maintained over the entire measurement interval. However such data do not exist. All measurement sites have undergone local surface changes, equipment has changed, and so forth. Consequently, data sets published as “climate” records are not mere observations: they are the outputs of models that take weather records as inputs, apply adjustments aimed at removing non-climatic influences, group the resulting records into regional grids and then translate the data into deviations from a local averages, yielding what are called gridded climate “anomalies”.

The problems with raw temperature data are widely recognized, including by the CRU. The CRU web page (http://www.cru.uea.ac.uk/cru/data/hrg/) references data compilations called CRU TS 1.x, 2.x and 3.x which are not subject to adjustments for non-climatic influences. **Users are explicitly cautioned not to use the TS data for measuring or analyzing climate change in the way apparently done in IPCC reports.** The 1.2 release of this product provided a list of FAQ's related to time series analysis (see http://www.cru.uea.ac.uk/cru/data/hrg/timm/grid/ts-advice.html). The first question, and its answer, are reproduced (in part) below.

> **Question One**
>
> **Q1**. Is it legitimate to use CRU TS 2.0 to 'detect anthropogenic climate change' (IPCC language)?
>
> **A1**. **No. CRU TS 2.0 is specifically *not* designed for climate change detection or attribution in the classic IPCC sense.[emphasis added]** The classic IPCC detection issue deals with the distinctly anthropogenic climate changes we are already experiencing. Therefore it is necessary, for IPCC detection to work, to remove all influences of urban development or land use change on the station data….If you want to examine the detection of anthropogenic climate change, we recommend that you use the Jones [sic, original contains link to CRUTEM data page] temperature data-set. This is on a coarser (5 degree) grid, but it is optimised for the reliable detection of anthropogenic trends.

The implication is that the Jones data (or the CRU surface record, called CRUTEM) has been adjusted "for the reliable detection of anthropogenic trends." Readers are referred to some academic papers for further explanation. The first is Brohan et al. (2006).[25] This paper does not explain how the data are adjusted, instead it focuses on defending the claim that the potential biases are very small. Two references are cited in support of this point. One is by US scientist Thomas Peterson, which refers to the contiguous US only. Another is by David Parker of the UK Hadley Centre, whose argument relied on an apparent similarity between trends on windy and calm nights. No references to papers critical of Parker's methods are cited. Section 2.3.3 of Brohan et al. states that to properly adjust the data would require a global comparison of urban versus rural records, but classifying records in this way is not

[25] Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548

possible since "no such complete meta-data are available" (p. 11), **so the authors instead make an arbitrary assumption (p. 11) that the bias is no larger than 0.06 degrees per century.** This *assumption* would later appear in the 2007 IPCC Summary for Policymakers as a research finding.

Brohan et al. refer to a 2003 paper in *Journal of Climate* by Jones and Moberg,[26] explaining the CRUTEM version 2 data product. This paper also has little information about the data adjustments. Reference is made to combining multiple site records into a single series, but not to removing non-climatic contamination. Moreover, the article points out (page 208) that it is difficult to say what homogeneity adjustments have been applied since the original data sources do not always include this information.

The other reference on the website is to a 1999 *Reviews of Geophysics* paper by Jones, New, Parker et. al.[27] This paper emphasizes that non-climatic influences (therein referred to as "inhomogeneities") must be corrected (Section 2, p. 37) for the data to be useful for climatic research. The part of the paper that provides information on the adjustments is Section 2.1, consisting of only 3 paragraphs, none of which explains the CRU procedures. The only explanatory statement is (page 174):

> "All 2000+ station time series used have been assessed for homogeneity by subjective interstation comparisons performed on a local basis. Many stations were adjusted and some omitted because of anomalous warming trends and/or numerous nonclimatic jumps (complete details are given by Jones et al. [1985, 1986c])."

Jones et al. [1985, 1986c] are technical reports that were submitted to the US Department of Energy, but they only cover data sets ending in the early 1980s, whereas the data now under dispute is the post-1979 interval. Even if the adjustments were adequate in the pre-1980 interval it is likely impossible to have estimated appropriate empirical adjustments in the early 1980s for changes in socioeconomic patterns that did not occur until the 1990s.

In sum, the CRU cautions that its unadjusted temperature data (the TS series) is inappropriate for the IPCC's purpose, and for climate detection and attribution analysis more generally. The CRU refers users instead to the adjusted CRUTEM products. Yet the accompanying documentation does not explain the

---

[26] Jones, P.D. and A. Moberg (2003). "Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001." *Journal of Climate* 16 203—223.

[27] Jones, P.D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, (1999) Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics.,* **37,** 173–199.

adjustments made to make the data products reliable for such usage, nor does it test if the adjustments are adequate.

Such studies have been done, and many show that the adjustments are inadequate, and the data are contaminated with non-climatic biases. The IPCC did not assess this material objectively, instead they assigned the topic to Jones and Parker, among a few others. It will be shown in the next section that these two exerted decisive influence on the text of the IPCC assessment of the data quality issue. Yet the IPCC had to have known that they lacked appropriate independence, and the effect of their intervention was to suppress information revealing quality problems in their own data. The EPA's unquestioned use of the CRUTEM data and its reliance on the IPCC in defence of doing so is a significant violation of its DQA responsibility.

2.3 EPA reliance on CCSP and IPCC assurances about data quality despite those sources providing no such evidence

In the April 2009 Technical Support Document (TSD) accompanying the Endangerment Finding, the EPA dismissed evidence of problems in the surface temperature record as follows (p. 22):

> Biases may exist in surface temperatures due to changes in station exposure and instrumentation over land, or changes in measurement techniques by ships and buoys in the ocean. It is likely that these biases are largely random and therefore cancel out over large regions such as the globe or tropics (Wigley et al., 2006). Likewise, urban heat island effects are real but local, and have not biased the large-scale trends (Trenberth et al., 2007).

Wigley et al. (2006) is a reference to the Executive Summary of the 2006 Climate Change Science Program Report "Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences." The review of the topic in the underlying CCSP report is found in Chapter 4, on pages 84-85. The report text provides no support for the TSD claim about random errors canceling out, nor does it address any of the published research after 2004 which demonstrates global-scale biases in the surface temperature network.

The CCSP report acknowledges the inherent problems in the underlying data but asserts without supporting evidence that the problems are removed by unspecified "adjustments" (emphasis added):

> **[Because] most land networks were not designed for climate monitoring, the data contain biases that data set creators address with different detailed methods of analysis.** The primary sources of uncertainty from a land-surface perspective are (a) the construction methods used in

> the analyses and (b) local environmental changes around individual observing stations (*e.g.,* urbanization) that may not have been addressed by the homogeneity assessments.
>
> Because the stations are not fully representative of varying-within-area land surface, coastal, and topographical effects, global data sets are produced by analyzing deviations of temperature from station averages (anomalies) as these deviations vary more slowly with a change in location than the temperatures themselves (Jones *et al.,* 1997). Random errors in inhomogeneity detection and adjustments may result in biased trend analyses on a grid box level. **However, on the relatively large space scales of greatest importance to this Report, such problems are unlikely to be significant in current data sets in the period since 1958 except where data gaps are still serious, *e.g.,* in parts of central Africa, central South America, and over parts of Antarctica.** Note that for the contiguous United States, the period 1958-2004 uses the greatest number of stations per grid box anywhere on the Earth's land surface, generally upwards of 20 stations per grid box. **For regions with either poor coverage or data gaps, trends in surface air temperature should be regarded with considerable caution, but do not have serious effects on the largest of scales as most of the spatial variability is well sampled**.

The absence of supporting evidence for repeated assertions about data quality is obvious in the above quotation. The test then appeals to the apparent *absence* of published studies testing for global-scale bias as evidence *for* the quality of the surface data. In doing so it only refers to global-scale studies published up to 1999 and thus ignores all the material discussed in Section 2.4 below. The authors were either unaware of it or, like the IPCC, knew about it but suppressed it (emphasis added):

> A variety of studies have documented that urbanization has a warming effect on the local microclimate; **however, no study has demonstrated that urban warming imparts a significant bias to multi-decadal trends over large areas**. In fact, the effect appears at most to be roughly an order of magnitude smaller than long-term trends (e.g., Jones et al., 1990). Several recent global (e.g., Easterling et al., 1997; Peterson et al., 1999) and national analyses (e.g., Li et al., 2004; Peterson et al., 2003) also indicate that urban and rural station networks had comparable trends since roughly the mid-20th century. In addition, minimum temperature trends since 1950 were similar on both windy and calm nights, the latter being more susceptible to urban warming (Parker, 2004).

The appeal to rural-urban comparisons is weakened by the fact that in much of the developing world there is either no rural data available, since the GHCN is heavily reliant on data from urban airports, or there is little meta-data to indicate the population of the areas where the sampling occurs. This is acknowledged in the CCSP report in the next paragraph, which conjectures that local biases cancel out at a large scale, but *notes that this has not been proven*:

> To insure that potential urbanization effects do not impact analyses, the NASA group adjusts the data from all urban stations so that their long-term trends are consistent with those from neighbouring rural stations (Hansen et al., 2001). It is generally accepted that local biases in trends mostly cancel through the use of many stations or ocean observations. **Because such a cancellation has not been rigorously proved**, partly due to the lack of adequate metadata, it is conceivable that systematic changes in many station exposures of a similar kind may exist over the land during the last few decades. If such changes exist, they may lead to small amounts of spurious cooling or warming, even when the data are averaged over large land regions.

The argument presented in Parker (2004) relied upon by the CCSP and subsequently by the EPA was eventually tested in a peer-reviewed paper by McKitrick (2013)[28] which showed it to be invalid, since it took the form of a test that could not reliably detect a contamination pattern even when it is known to be in the data.

In sum, the material in CCSP provides no basis for the TSD's dismissal of concerns about surface data quality, and on the specific claim that biases randomly cancel out, the CCSP report openly notes that this has not been proved and the metadata required to do so does not exist.

2.4 IPCC falsification of the evidence of data contamination

2.4.1 Jones conspires with Trenberth to suppress critical evidence in IPCC 2007 report

The other TSD source is Trenberth et al. (2007), which is a reference to Chapter 3 of the 2007 IPCC Report. Aside from the Wigley et al (2006) reference it is the sole basis for the assumption by the EPA that surface temperature data are uncontaminated.

[28] McKitrick, Ross R. (2013) Encompassing Tests of Socioeconomic Signals in Surface Climate Data. *Climatic Change* doi 10.1007/s10584-013-0793-5. Volume 120, Issue 1-2. See https://link.springer.com/article/10.1007%2Fs10584-013-0793-5

The CRU data was crucial for some of the main conclusions in the AR4. Global temperature trends are presented in AR4 Table 3.2 on page 243. The accompanying text (page 242) states that the CRU data uncertainties "take into account" biases due to urbanization. The Executive Summary to the chapter (page 237) asserts that "Urban heat island effects are real but local, and have not biased the large-scale trends…the very real but local effects are avoided or accounted for in the data sets used." The highly influential Summary for Policymakers stated:

> "Urban heat island effects are real but local, and have a negligible influence (less than 0.006°C per decade over land and zero over the oceans) on these values."

The supporting citation was to Section 3.2, which will be shown below to rely on fabricated evidence. IPCC Chapter 9 provides the summary of evidence attributing warming to greenhouse gases. The problem of CRU surface data contamination is set aside as follows (p. 693):

> Systematic instrumental errors, such as changes in measurement practices or urbanisation, could be more important, especially earlier in the record (Chapter 3), although these errors are calculated to be relatively small at large spatial scales. Urbanisation effects appear to have negligible effects on continental and hemispheric average temperatures (Chapter 3).

Again, the rationale for ignoring the issue of CRU data quality problems relies on a citation to Chapter 3, which in turn relied upon what turns out to be fabrications.

At the time of the 2007 IPCC report's preparation, evidence had been published by two independent teams in mainstream peer-reviewed journals,[29] each analyzing global-scale samples, showing statistically significant evidence that contamination in the surface temperature record due to industrialization and related land-use effects had not been adequately removed from climatic data sets and it added a clear warming bias at the global level in recent decades.

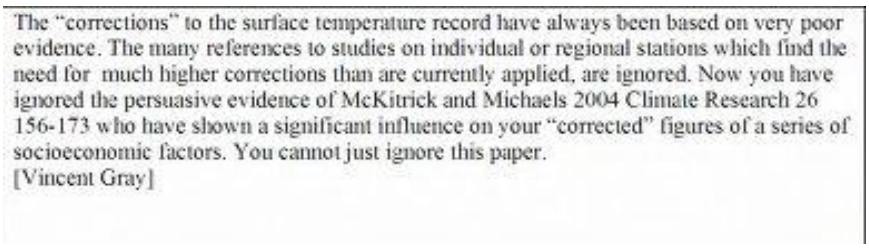
[29] De Laat, A.T.J., and A.N. Maurellis (2004), Industrial CO_2 emissions as a proxy for anthropogenic influence on lower tropospheric temperature trends, *Geophys. Res. Lett.* Vol. 31, L05204, doi:10.1029/2003GL019024. McKitrick, R.R. and P. J. Michaels (2004), A test of corrections for extraneous signals in gridded surface temperature data, *Climate Research* 26(2) pp. 159-173, Erratum, *Clim. Res.* 27(3) 265—268. De Laat, A.T.J., and A.N. Maurellis (2006), Evidence for influence of anthropogenic surface processes on lower tropospheric and surface temperature trends, *Int. J. Climatol.* 26:897—913.

One of the Climategate emails[30] is from IPCC Author Phil Jones to his colleague Michael Mann on July 8 2004, in which Jones confides that he and IPCC coauthor (Kevin) Trenberth were conspiring to keep this evidence out of the IPCC Report (emphasis added):

> "The other paper by MM is just garbage. [...] I can't see either of these papers being in the next IPCC report. **Kevin [Trenberth] and I will keep them out somehow — even if we have to redefine what the peer-review literature is!**"

'MM' in this email refers to McKitrick, R.R. and P. J. Michaels (2004), A test of corrections for extraneous signals in gridded surface temperature data, *Climate Research* 26(2) pp. 159-173. It is not clear what is the second paper to which Jones refers, but it might have been one by de Laat and Maurellis.

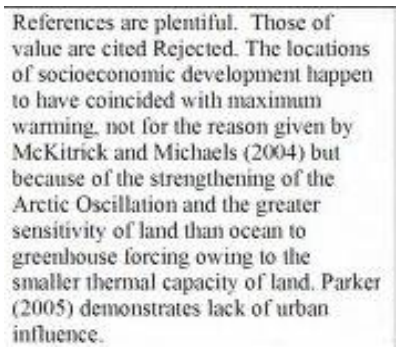
The IPCC released the First Order Draft in August 2005. Since this was over a year after Jones' email to Mann it is clear he was aware of the MM study. The relevant section of the IPCC First Order Draft was Chapter 3, pages 3-9 to 3-10. Consistent with the intent expressed in the email there was no mention of either MM2004 or the de Laat and Maurellis work and instead the section drew a conclusion that there was no evidence of large-scale bias in the data due to socioeconomic effects over the land surface. IPCC Expert Reviewer V. Gray of New Zealand criticized the omission as follows:


The "corrections" to the surface temperature record have always been based on very poor evidence. The many references to studies on individual or regional stations which find the need for much higher corrections than are currently applied, are ignored. Now you have ignored the persuasive evidence of McKitrick and Michaels 2004 Climate Research 26 156-173 who have shown a significant influence on your "corrected" figures of a series of socioeconomic factors. You cannot just ignore this paper.
[Vincent Gray]

(http://pds.lib.harvard.edu/pds/view/7795947?n=7&imagesize=1200&jp2Res=.25). Expert review comments by McKitrick also criticized the omission.

The IPCC response to the Gray comment was:

[30] Email 1089318616.txt http://www.ecowho.com/foia.php?file=1089318616.txt.



References are plentiful. Those of value are cited Rejected. The locations of socioeconomic development happen to have coincided with maximum warming, not for the reason given by McKitrick and Michaels (2004) but because of the strengthening of the Arctic Oscillation and the greater sensitivity of land than ocean to greenhouse forcing owing to the smaller thermal capacity of land. Parker (2005) demonstrates lack of urban influence.

It is noteworthy that the IPCC respondent acknowledges that "the locations of socioeconomic development happen to have coincided with maximum warming" but dismisses the obvious implication of this on the basis of an unsupported and implausible speculation that the effect is due a wind pattern at the north pole called the Arctic Oscillation. This claim would subsequently be refuted in the scientific literature, as will be explained below. The comparison of land and ocean was irrelevant since the dispute is not based on a comparison of land and ocean, but is only about the land record.

The IPCC respondent also relied entirely on a paper by Jones' colleague and coauthor David Parker of the UK Met Office as grounds to set the matter aside. Two points are of note here.

- First, the Parker paper did not address any of the evidence and analysis in the MM paper or the de Laat and Maurellis papers, instead it was a more narrowly focused study on the role of wind in urban heat island effects.
- Second, emails released to Mr. David Holland of the UK, referred to above as ICO decision notice FER0239225, show that the IPCC respondent who entered the above rejection was David Parker himself, hence he was not relying on independent evidence. Parker was a Lead Author for chapter 3, working under the direction of his collaborator Phil Jones, one of the two Coordinating Lead Authors. In a November 2005 email to Jones, Parker indicated that he had gone through the IPCC chapter draft and entered some of the needed responses to reviewers. Regarding the Gray comment, Parker boasted:

> I have rejected the McKitrick and Michaels 2004 with appropriate reasons!

Just as Jones was in a conflict of interest as the head of CRU evaluating the work of critics of CRU data quality, so Parker was in a conflict of interest relying on his own work to defend that of his close

colleague**. Thus in both cases there was no independence of the kind required by the EPA Inspector General.**

The IPCC Second Order Draft was released in March 2006. Again consistent with the intent revealed in Jones' email to Mann, and despite reviewer demands, there was still no mention of the MM findings or those of deLaat and Maurellis. Gray and McKitrick again provided feedback objecting to this omission. In June 2006 the expert review period closed.

2.4.2 IPCC fabricated evidence that the EPA relied upon, verbatim

After the close of expert review, and at some time between March and June 2006, a paragraph was inserted into the IPCC chapter that misrepresented the findings in the publications and made empirical claims with no supporting evidence (2007 IPCC Report, Chapter 3 page 244, emphasis added):

> McKitrick and Michaels (2004) and De Laat and Maurellis (2006) attempted to demonstrate that geographical patterns of warming trends over land are strongly correlated with geographical patterns of industrial and socioeconomic development, implying that urbanisation and related land surface changes have caused much of the observed warming. **However, the locations of greatest socioeconomic development are also those that have been most warmed by atmospheric circulation changes (Sections 3.2.2.7 and 3.6.4), which exhibit large-scale coherence. Hence, the correlation of warming with industrial and socioeconomic development ceases to be statistically significant.** In addition, observed warming has been, and transient greenhouse-induced warming is expected to be, greater over land than over the oceans (Chapter 10), owing to the smaller thermal capacity of the land.

Two points are of particular note:

- The first highlighted sentence is false: neither of the cited report sections (3.2.2.7 and 3.6.4) make any mention of or show any information on the spatial pattern of industrialization or its overlap with the warming record.
- The second highlighted sentence is a fabrication. The concept of "statistical insignificance" has a specific quantitative interpretation: it implies that an empirical test has been done yielding a *p* value greater than 0.1 (for marginal significance) or 0.05 (for significance). Both cited papers reported statistically *significant* correlations between warming patterns and the spatial distribution of industrialization. Specifically the effects reported in MM2004 had *p* values on the order of 0.002 or 0.2%, indicating significance. **The claim that the results were statistically**

> **insignificant is a direct falsification of the evidence. Nor did the IPCC present any supporting evidence to show that the results could be attributed to natural atmospheric circulation changes, nor. The claim was subsequently tested in a peer-reviewed paper and shown to be untrue.**[31]

The EPA relied verbatim on the IPCC fabrication quoted above in its dismissal of comments on the Endangerment finding:

> Commenters also point to recent papers (e.g., McKitrick and Michaels, 2007; de Laat and Maurellis, 2006) that attempt to demonstrate that geographical patterns of warming trends over land are strongly correlated with geographical patterns of industrial and socioeconomic development, implying that urbanization and related land surface changes have biased the temperature trends (and are, therefore, the cause of much of the observed warming). In the case of de Laat and Maurellis (2006) and an earlier paper by McKitrick and Michaels (2004), IPCC (Trenberth et al., 2007) assessed these papers and noted that the locations of greatest socioeconomic development coincided with those most warmed by atmospheric circulation changes, which are not limited to urban areas but rather have large-scale coherence. When this is taken into account, IPCC concludes that the correlation of warming with industrial and socioeconomic development ceases to be statistically significant.
>
> http://www.epa.gov/climatechange/endangerment/comments/volume2.html#2

The last two sentences repeat uncritically the untrue claims in the IPCC Report.

Since the CCSP report does not substantiate the EPA's position, the EPA's conclusions regarding the integrity of the surface temperature record can thus be seen to depend entirely on IPCC material that was untrue, fabricated and kept out of drafts shown to reviewers—an action which the Climategate emails showed to have been planned a year prior to the review process.

---

[31] McKitrick, Ross R. (2010) "Atmospheric Oscillations do not Explain the Temperature-Industrialization Correlation." Statistics, Politics and Policy, Vol 1 No. 1, July 2010.

2.4.3 IPCC subsequently withdrew the fabricated claim

The next IPCC report conceded that the claim relied upon by the EPA was made with no supporting evidence, and that the evidence of surface data contamination was robust. The IPCC AR5 (chapter 2 page 34) summarized the issue as follows (emphasis added):

> *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.***

2.4.4 UK Inquiry fails to find evidence for claim relied upon by EPA

In submissions to the 2010 UK House of Commons and Muir Russell Inquiries into the Climategate emails,[32] McKitrick asked both groups to demand of CRU scientist Phil Jones the *p*-value of the test that showed the MM results to be statistically insignificant. The House of Commons Inquiry did not do so but the Muir Russell panel (http://www.cce-review.org) did. They reported no answer, and inquiry documentation shows that Jones was unable to provide one.[33]

2.5 EPA relies upon subsequent non-independent research

Subsequent to IPCC 2007 there have been several papers published confirming the points raised in the McKitrick and Michaels (2004) and de Laat and Maurellis (2004, 2006) studies. McKitrick and Michaels (2007) expanded their analysis to include complete global coverage of the CRU land record over a longer time interval and obtained nearly identical results, but their work was not discussed in the IPCC (2007)

[32] Available at http://www.rossmckitrick.com/climategate.html.
[33] See Jones' responses at http://www.cce-review.org/evidence/15%20April%20Jones%20follow%20up.pdf.

report nor in CCSP reports. A 2009 paper by Gavin Schmidt[34] of NASA-GISS made a series of claims against the reliability of these results without subjecting them to formal statistical modeling and testing. The EPA relied upon Schmidt's paper in another section of its rejection of comments on the Endangerment finding:

> Neither IPCC nor CCSP assess McKitrick and Michaels (2007) which conclude that "that non-climatic factors, such as those related to land use change and variations in data quality, likely add up to a net warming bias in climate data, suggesting an overstatement of the rate of global warming over land." However we note a recent study by Schmidt (2009) that finds "The reported correlations [in McKitrick and Michaels, 2007]…are probably spurious (i.e. are likely to have arisen from chance alone). Thus, though this study cannot prove that the global temperature record is unbiased, there is no compelling evidence from these correlations of any large-scale contamination."

It is noteworthy that the TSD cites Gavin Schmidt as an expert reviewer (p. ii) but no one from the other side of the debate, indicating a lack of diligence on their part in obtaining a balanced group of reviewers on this issue.

Schmidt's claim, as stated by the EPA, were not actually demonstrated by the simulations in Schmidt's paper. As noted above, a peer-reviewed paper in 2010 by McKitrick and Nierenberg[35] rebutted Schmidt's claim and, as noted above, the IPCC in the 5th Assessment Report conceded the point.

2.5.1 Further independent evidence of data contamination

Klotzbach et al. (2009)[36] tested the same issue in a different way. They observed that if there is no contamination of surface data due to land use changes, the difference between surface trends and satellite-based measures of the lower troposphere should be constant over time. But instead the trends diverge, and the divergence runs opposite to the direction predicted by climate models. Consequently they conclude:

---

[34] Schmidt, GA (2009) Spurious correlation between recent warming and indices of local economic activity. *International Journal of Climatology* 10.1002/joc.1831

[35] McKitrick, Ross R. and Nicolas Nierenberg (2010) "Socioeconomic Patterns in Climate Data." *Journal of Economic and Social Measurement,* Vol 35 No. 3-4 pp. 149-175.

[36] Klotzbach, P.J., R.A. Pielke Sr., R.A. Pielke Jr., J.R. Christy, and R.T. McNider, (2009): An alternative explanation for differential temperature trends at the surface and in the lower troposphere. *J. Geophys. Res.*, 114, D21102, doi:10.1029/2009JD011841.

> The differences between surface and satellite data sets tend to be largest over land areas, indicating that there may still be some contamination because of various aspects of land surface change, atmospheric aerosols and the tendency of shallow boundary layer to warm at a greater rate.

They also point out that such findings indicate that it was incorrect for the 2006 CCSP Report (Karl et al. 2006) to claim that the surface and satellite data series have been reconciled as a result of corrections to the satellite record.

The Klotzbach findings were critiqued in blog posts by Gavin Schmidt of NASA-GISS, but a peer-reviewed paper[37] by independent authors in the *Journal of Time Series Analysis* upheld Klotzbach's findings. The authors formally assessed the ratio of troposphere to land trends and concluded "We find that amplification ratios typically associated with climate models are rejected by the observed temperature data confirming and extending the empirical findings of Klotzbach et al (2009, 2010). Allowing for a structural change at the end of 1998 to account for the so-called "hiatus" in warming gives results similar to Klotzbach et al (2009, 2010)."

3 Reliance on invalid trend uncertainty estimation methods.

> **Rebuttal:** The EPA asserts that modern temperature trends are too large to be consistent with natural variability. Ample peer-reviewed literature has long disputed this claim. The IPCC lacks the statistical competence to assess this and instead has suppressed the topic. In response to the first round of AR4 reviews they introduced a clearly-worded caution that their trend estimation methods likely understate uncertainties. But after the close of peer review this caution was removed and replaced with unsubstantiated text claiming the opposite. Notwithstanding this falsification of the record, numerous papers have shown the issue is real, and that it is not easy to conclude that modern climate trends are outside the range of natural variability.

[37] Vogelsang, Timothy and Nasreen Nawaz (2016) Estimation and Inference of Linear Trend Slope Ratios With an Application to Global Temperature Data. *Journal of Time Series Analysis* 10.1111/jtsa.12209.

3.1 Relevance: EPA reliance on IPCC handling of the issue

The EPA TSD discusses the issue of global warming trends in Section 4(b), pp 22-25. It relies entirely on data and analysis presented in the IPCC (2007) report. One of the critical aspects of such analysis, as illustrated by the prominence given to TSD Figure 4.2, is the apparent acceleration of recent trends. Trends are reported in the form of degrees C per decade, with 95% confidence intervals; for example they report a trend of 0.07 ±0.02 °C/decade globally over the 1906-2005 interval, meaning they are 95% confident the trend is between 0.05 and 0.09 °C/decade. They also report that the trend over the past 50 years is 0.13 ±0.03 °C/decade, indicating that the rate has almost doubled, and because the 95% confidence intervals that they computed do not overlap (0.05—0.09 versus 0.10—0.16) the increase in the warming rate is deemed statistically significant.

3.2 Technical background

Computing the trend term (the °C/decade rate) is straightforward. Computing the 95% confidence interval (the ± range) is much more complex. The issue is referred to in academic literature as "long term persistence." Data series, such as geophysical data like temperature and precipitation, often exhibit persistence, which mimics trend-like behavior. In order to compensate for the presence of persistence, the ± range must be calculated using specialized formulae, which typically yield a much wider 95% confidence interval compared to that generated using the simplistic IPCC methods.

The trend calculations in the IPCC report are not drawn from peer-reviewed literature directly, instead they are calculated especially for the report by the Lead Authors of Chapter 3, who do not use persistence methods. Instead they use a simplistic method called REML-AR1, which yields very small confidence intervals. During the review of the First Order Draft the IPCC was apprised by reviewers of some of the recent literature criticizing the method they were using, and were instructed to include a caution that, due to the particular features of data they were working with, their method might be overstating the significance of warming trends.

3.3 IPCC deletion of material inserted during the review process

In response to reviewer comments on the First Order Draft, the Second Order Draft of Working Group I, Chapter 3 of the AR4 inserted new text on page 3-9 as follows:

> Table 3.2 provides trend estimates from a number of hemispheric and global temperature databases. Determining the statistical significance of a trend line in geophysical data is difficult, and many oversimplified techniques will tend to overstate the significance. Zheng and Basher

> (1999), Cohn and Lins (2005) and others have used time series methods to show that failure to properly treat the pervasive forms of long-term persistence and autocorrelation in trend residuals can make erroneous detection of trends a typical outcome in climatic data analysis.

Reviewers apparently approved of its inclusion as there were no subsequent objections to it. A related statement was also included in the Appendix of the Second Order Draft (p. 3-116) cautioning that the method used by the IPCC authors to compute trends, REML AR1, yields statistical significance levels that are "likely to be overestimated," meaning that the temperature trends they reported may actually be attributable to natural variability rather than being indications of a sustained warming pattern:

> As some components of the climate system respond slowly to change, the climate system naturally contains persistence, so that the REML AR1-based linear trend statistical significances are likely to be overestimated (Zheng and Basher, 1999; Cohn and Lins, 2005). Nevertheless, the results depend on the statistical model used, and more complex models are not as transparent and often lack physical realism.

This was the wording in the final draft seen by reviewers. The draft of the IPCC Report that was sent out for government review on July 3 2006 (immediately after the close of expert review) also still included these statements.[38]

The Chapter Lead Authors had already acknowledged the underlying problem, as a November 2005 email from UK Met Office scientist (and IPCC Author) David Parker to IPCC Coordinating Lead Author Phil Jones made clear:[39]

> Maybe the biggest problem is Ross McKitrick and David Stephenson's remarks on trends; we only use an AR-1 and they may be correct in advocating a more complex model. Our software for restricted maximum likelihood does not cope with ARMA (1,1) and I may have to get John Kennedy to investigate new software using the cited references.

---

[38] This was confirmed by examination of attachments to emails from IPCC Review Editor Brian Hoskins that were released to Mr. David Holland in response to a Freedom of Information Act request, ICO decision notice FER0239225.

[39] This email is taken from a set of IPCC-related documents held by the University of Reading and released to UK resident Mr. David Holland in response to a Freedom of Information Act request, , ICO decision notice FER0239225.

The final version of the IPCC report that was published in May 2007 still applied the invalid AR-1 method. But at some point between circulation of the text for government review and release of the final, published edition, the statement warning of erroneous trend detection was deleted, and was replaced with text saying the opposite (p. 242):

> In Table 3.2, the effects of persistence on error bars are accommodated using a red noise approximation, which effectively captures the main influences. For more extensive discussion see Appendix 3.A

The text in the Appendix 3.A was changed to the following (p. 336):

> As some components of the climate system respond slowly to change, the climate system naturally contains persistence. Hence, the statistical significances of REML AR1-based linear trends could be overestimated (Zheng and Basher, 1999; Cohn and Lins, 2005). Nevertheless, the results depend on the statistical model used, and more complex models are not as transparent and often lack physical realism. Indeed, long-term persistence models (Cohn and Lins, 2005) have not been shown to provide a better fit to the data than simpler models.

No supporting evidence was provided for the last sentence. It is contradicted by a large set of empirical findings in the peer-reviewed literature. For instance, Rybski et al. (2006),[40] which was available to the IPCC prior to the release of AR4, discusses a suite of temperature reconstruction indexes and finds all of them characterized by long-term persistence, noting that this has also been shown in numerous studies of other temperature records:

> We show that all records are characterized by pronounced long-term persistence, similar to those found in real climate records [Koscielny-Bunde et al., 1998; Pelletier and Turcotte, 1999; Eichner et al., 2003; Blender and Fraedrich, 2003; Vyushin et al., 2004; Kiraly and Janosi, 2005; see also Cohn and Lins, 2005]. Here, the term ‘‘long-term persistence’’ refers to auto-correlation functions which decay by a power law and are characterized by an infinite correlation time. Due to the long-term correlations, the variability on long time scales is strongly enhanced.

Hence the changes made to the IPCC report after the close of peer review were as follows.

---

[40] Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006), Long-term persistence in climate and the detection problem, Geophys. Res. Lett., 33, L06718, doi:10.1029/2005GL025591.

- A caution about the likelihood of erroneous detection of warming trends, that had been inserted in response to expert review, was deleted.

- An unsupported claim was inserted into the chapter (p. 242) claiming that the chapter authors' trend detection method (REML AR1) "effectively captures the main influences," despite the chapter authors having said the opposite in an earlier draft, namely that their method likely overestimated the significance of trends.

- A caution in the Appendix that "linear trend statistical significances are likely to be overestimated" was changed to say merely that they "could be" overestimated.

- A sentence was added to the Appendix disputing the validity of long term persistence models, with no supporting citations, and contradicted by ample published evidence.

This provides a further illustration of the problem in the IPCC that Lead Authors exert so much arbitrary control. In this case they not only changed the text after the close of peer review, but they changed it to override and reverse material that had been inserted in response to expert review. This corruption of the IPCC report on a centrally-important topic shows that reliance on IPCC "findings" by the EPA is unacceptable.

3.4 Subsequent literature confirming evidence that the IPCC sought to suppress

Lennartz and Bunde (2009)[41] examined 45 temperature records, including 15 global or hemispheric averages and 30 individual city records, and found all exhibited long term persistence. Trend confidence intervals were therefore wider but still showed a significant warming had been observed over the 20th century in 13 of 15 global records and in most of the individual locations. However, over the past 50 years, the trend was found to be statistically *insignificant* in 25 out of 30 individual locations and 11 of 15 global averages, and they could not conclude that the recent trend was significantly higher than that for the century as a whole in any records except 4 of the 15 global series and 2 individual locations. The 4 global series that did exhibit a higher recent trend were the CRU Northern Hemisphere and global land records versions 2 and 3, namely the series most likely to exhibit a warm bias in recent decades due to urbanization, evidence of which the IPCC suppressed (see Section 2). By contrast, none of the series based chiefly on ocean surface temperatures exhibited evidence of a significant increase in trend, and

[41] Lennartz, S., and A. Bunde (2009), Trend evaluation in records with long-term memory: Application to global warming, Geophys. Res. Lett., 36, L16706, doi:10.1029/2009GL039516.

the Hadley global Sea Surface Temperature series (hadsst2) did not exhibit a significant trend at all over the entire 20th century or the most recent 50 years.

Mills (2010)[42] reviews several approaches to trend estimation and presents a general framework ("structural trend modeling") that decomposes long time series in to cyclical and trend components, thus embedding numerous specifications as special cases, including long persistence processes. He finds that the trend component of the 1850—2007 HADCRU global temperature series is not significantly different from zero. He also examines whether the trend confidence intervals are large enough to encompass the trend forecasts in the Hadley Centre climate model, and concludes that the model forecast "lies way outside any conventionally calculated confidence interval." Mills (2016)[43] provided a further update on this material and concluded again that the data did not provide support for model forecasts of significant future warming.

4 Reliance on hockey stick-style paleoclimate studies

Rebuttal: The TSD relied on the NRC (2006) report and IPCC (2007) to claim that modern temperatures are outside the bounds of historical ranges. The TSD offered no substantial discussion, merely cherry-picking a few concluding statements. It ignored what the NRC (2006) took to be one of their principal findings, that uncertainties of previous reconstructions have been underestimated. It also ignored the deletion of contrary evidence in the IPCC report. The IPCC and NRC both ignored the problem that the flawed hockey stick method is reused in other places and much of the underlying data is also reused, so the supporting studies are not independent. Subsequent literature has shown that proxy uncertainty makes it impossible to place much weight on the kinds of conclusions asserted in the TSD.

4.1 Relevance

If modern temperatures are within the range of natural variability, it is difficult to argue that the human influence on climate is overly large, or is likely to pose any greater dangers or hazards than ordinary

[42] Mills T. C. (2010) Skinning a cat: alternative models of representing temperature trends. *Climatic Change* 101: 415-426, DOI 10.1007/s10584-010-9801-1.

[43] http://www.thegwpf.org/statistical-forecasting-how-fast-will-future-warming-be/

climatic variability. Consequently, the IPCC has made a deliberate effort to find and promote results that appear to show the modern warming has taken us outside the bounds of natural variability. That led to the heavy use of the Mann hockey stick graph in the TAR, controversies over which led to the subsequent Wegman (2006)[44] and NRC reports (North et al.[45]). The TSD downplays the topic but nonetheless quotes IPCC and NRC statements that modern temperatures are likely higher than at any time in the past 1300 years.

4.2 NRC (2006) Conclusions

One of the main arguments made by McIntyre and McKitrick in their presentation[46] to the NRC panel was that that the hockey stick method, and the test statistics used to validate it, systematically underestimated the uncertainties in the data. The NAS panel agreed, concluding that this was a more general problem (p. 107):

> Regarding metrics used in the validation step in the reconstruction exercise, two issues have been raised (McIntyre and McKitrick 2003, 2005a,b). One is that the choice of "significance level" for the reduction of error (RE) validation statistic is not appropriate. The other is that different statistics, specifically the coefficient of efficiency (CE) and the squared correlation (r^2), should have been used (the various validation statistics are discussed in Chapter 9). Some of these criticisms are more relevant than others, but taken together, they are an important aspect of a more general finding of this committee, which is that uncertainties of the published reconstructions have been underestimated.

The NAS reported the failure of Mann's reconstruction to achieve statistical significance in a roundabout way, by referring to the problem in the results of Wahl and Ammann (2007),[47] who used the identical data and methods as Mann, rather than the original Mann paper itself (p. 91):

---

[44] Wegman, E. J., Scott, D. W. and Said, Y. H. (2006). Ad-hoc committee report on the 'hockey stick' global climate reconstruction. Report, House Committee on Energy and Commerce.
Available at http://republicans.energycommerce.house.gov/108/home/07142006_Wegman_Report.pdf

[45] North, G. et al. (National Research Council, NRC) (2006). *Surface Temperature Reconstructions for the Last 2,000 Years*. Washington: National Academies Press.

[46] Online at http://climateaudit.files.wordpress.com/2009/12/nas-mm.pdf

[47] Wahl, E.R. and C. M. Ammann (2007) "Robustness of the Mann, Bradley, Hughes reconstruction of Northern Hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence." Climatic Change (2007) 85:33–69 DOI 10.1007/s10584-006-9105-7.

> Reconstructions that have poor validation statistics (i.e., low CE) will have correspondingly wide uncertainty bounds, and so can be seen to be unreliable in an objective way. Moreover a CE statistic close to zero or negative suggests that the reconstruction is no better than the mean, and so its skill for time averages shorter than the validation period will be low. Some recent results reported in Table 1S of Wahl and Ammann (in press) indicate that their reconstruction, which uses the same procedure and full set of proxies used by Mann et al. (1999), gives CE values ranging from 0.103 to –0.215, depending on how far back in time the reconstruction is carried.

A few points here are noteworthy.

- First, the IPCC relied exclusively on the Wahl and Ammann analysis in its summary of the controversy over the hockey stick, yet did not point out that their test scores indicated a failure at statistical validation prior to 1450. This may be because, even though neither Wahl nor Ammann were authors of that section, the IPCC Lead Author who wrote it (Keith Briffa) made private and undisclosed use of Wahl as a coach on that section.[48] Climategate2 email #160 from Briffa to Wahl indicates the nature of the backchannel communication that went on.

> From: Keith Briffa [mailto:k.briffa@uea.ac.uk]
> Sent: Mon 7/24/2006 3:16 PM
> To: Wahl, Eugene R
> Subject: RE: confidential
>
> Gene
> here is where I am up to now with my responses (still a load to do) - you can see that I have "borrowed (stolen)" from 2 of your responses in a significant degree - please assure me that this OK (and will not later be obvious) hopefully.
> You will get the whole text(confidentially again) soon. You could also see that I hope to be fair to Mike - but he can be a little unbalanced in his remarks sometime - and I have had to disagree with his interpretations of some issues also.

[48] In his EPA submission, Steve McIntyre notes that this episode violates the requirement that the IPCC disclose the entire review record since this correspondence was never released by the IPCC.

Please do not pass these on to anyone at all.

Keith

- The failure to validate prior to 1450 means that the Wahl-Ammann/Mann reconstruction does not provide any statistically significant information as to whether the modern temperature record is warmer than during the medieval era.

The NRC panel also sided with McIntyre and McKitrick in concluding that the statistical method used in the hockey stick yielded biased results (p. 106):

> As part of their statistical methods, Mann et al. used a type of principal component analysis that tends to bias the shape of the reconstructions. A description of this effect is given in Chapter 9.

They went on to state that the problem with Mann's method was that it loaded too much weight on a small number of bristlecone pine series from the western US, which were inappropriate series for use as temperature proxies. The NRC concluded (p. 107):

> The more important aspect of this criticism is the issue of robustness with respect to the choice of proxies used in the reconstruction. For periods prior to the 16th century, the Mann et al. (1999) reconstruction that uses this particular principal component analysis technique is strongly dependent on data from the Great Basin region in the western United States.

And in their examination of the data in question, they warned that these strip-bark series should not be used in this type of research (p. 50).

4.3 EPA reliance on evidence ruled inadmissible by the NRC

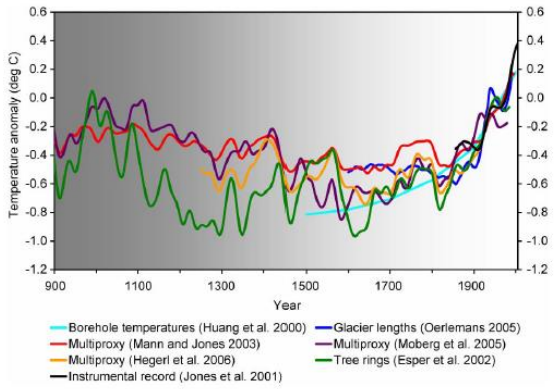
This latter finding is important since the graph relied on in the TSD document is itself heavily dependent on the bristlecone pine records and the faulty Mann statistical analysis (principal components).

Technical Support Document, U.S. Environmental Protection Agency

Global Surface Temperatures Over the Last 2,000 Years

Instrumental surface temperature records only began in the late 19th century, when a sufficiently large global network of measurements was in place to reliably compute global mean temperatures. To estimate temperatures further back in time, scientists analyze proxy evidence from sources such as tree rings, corals, ocean and lake sediments, cave deposits, ice cores, boreholes, glaciers, and documentary evidence. A longer temperature record can help place the 20th century warming into a historical context.

Figure 4.3: Reconstructions of (Northern Hemisphere average or global average) surface temperature variations from six research teams



Source: NRC (2006b) Reconstructions of (Northern Hemisphere average or global average) surface temperature variations from six research teams (in different color shades) along with the instrumental record of global average surface temperature (in black). Each curve illustrates a somewhat different history of temperature changes, with a range of uncertainties that tend to increase backward in time (as indicated by the shading).

The National Research Council conducted a study to describe and assess the state of scientific efforts to reconstruct surface temperature records for the Earth over approximately the last 2,000 years and the implications of these efforts for our understanding of global climate change. It found (NRC, 2006b):

TSD Figure 4.3, page 26.

The graph overlays proxy reconstructions by Mann and Jones (2003), Hegerl et al. (2006), Oerlemans (2005), Moberg et al. (2005) and Esper et al. (2002). The Hegerl and Oerlemans et al. graphs do not extend back to the medieval era and are uninformative with regards to the conclusion. The only ones that extend over the millennium are Mann and Jones, Esper and Moberg. As pointed out in Wegman (2006, p. 46), all three of these studies rely on the bristlecone data condemned by the NRC and they reuse the faulty Mann PC method singled out by the NRC for intense criticism.

4.4 Faulty calibration to modern temperatures: "Hide the Decline"

The Moberg and Esper results do not have higher values at the right-hand end of the graph compared to portions of the Medieval era. The visual effect of higher modern values arises from overlaying the 20th century thermometer record (in black) over the proxy reconstructions. Overlaying the proxy reconstructions with thermometer data is only valid if the underlying calibration models have yielded statistically significant validation scores. As noted above, the Mann hockey stick graph failed to do so,

and the NRC panel noted that the methodologies used in these studies generally yielded underestimates of uncertainties, or overestimates of the validation scores.

The weakness of the calibration between proxy records and modern thermometer records is at the core of the "hide the decline" scandal from Climategate. Evidence had emerged in the 1990s that after about 1960, tree ring proxy records exhibited a declining trend while temperatures indicated a rising trend. If proxies were failing to pick up the modern warmth, it is possible they also failed to pick up earlier warmth, in which case the height of the earlier portion of the graph might be underestimated. In 1998 CRU scientist Keith Briffa had published[49] a study using over 300 proxy records clearly showing a declining pattern in recent decades and a divergence from temperature trends. A problem arose for the IPCC when they included this reconstruction in a diagram with the Mann hockey stick, and the effect was to undermine confidence in all such reconstructions, including that by Mann. In Climategate email 3272, dated September 22 1999, IPCC Lead Author Chris Folland indicated the pressure that was on them to get the message right (emphasis added):

> >> >At 01:07 PM 9/22/99 +0100, Folland, Chris wrote:
>
> >> >>Dear All
>
> >> >>
>
> >> >>**A proxy diagram of temperature change is a clear favourite for the**
>
> **>> Policy**
>
> **>> >>Makers summary. But the current diagram with the [Briffa 1998] tree ring only data**
>
> **>> >>somewhat contradicts the [Mann] multiproxy curve and dilutes the message rather**
>
> **>> >>significantly.** We want the truth. Mike [Mann] thinks it lies nearer his result
>
> >> >>(which seems in accord with what we know about worldwide mountain
>
> >> glaciers
>
> >> >>and, less clearly, suspect about solar variations). The tree ring
>
> >> results
>
> >> >>may still suffer from lack of multicentury time scale variance. This is

[49] K. R. Briffa, F. H. Schweingruber, P. D. Jones, T. J. Osborn, I. C. Harris, S. G. Shiyatov, E. A. Vaganov and H. Grudd, 1998. Phil.Trans. R. Soc. Lond. B (1998) 353, 65-73

>> >>probably the most important issue to resolve in Chapter 2 at present.

>> >>

>> >>Chris

Right after this, Mann proposed dropping the Briffa series from the graph. Briffa objected in email #3272 (emphasis added):

>> At 04:19 PM 9/22/99 +0100, Keith Briffa wrote:

>> >

>> >Hi everyone

>> > Let me say that I don't mind what you put in the policy makers

>> >summary if there is a general concensus. However some general discussion

>> >would be valuable . First , like Phil , I think that the supposed

>> >separation of the tree-ring reconstruction from the others on the grounds

>> >that it is not a true "multi-proxy" series is hard to justify….

>> > There is still a potential problem with non-linear responses in the

>> >very recent period of some biological proxies (or perhaps a

>> fertilisation

>> >through high CO2 or nitrate input) . **I know there is pressure to present**

>> a

>> >nice tidy story as regards 'apparent unprecedented warming in a thousand

>> >years or more in the proxy data' but in reality the situation is not

>> quite

>> >so simple. We don't have a lot of proxies that come right up to date and

>> >those that do (at least a significant number of tree proxies) some

>> >unexpected changes in response that do not match the recent warming. I do

>> >not think it wise that this issue be ignored in the chapter.

>> > For the record, I do believe that the proxy data do show unusually

>> >warm conditions in recent decades. I am not sure that this unusual

>> warming

>> >is so clear in the summer responsive data. **I believe that the recent**

>> warmth

>> >was probably matched about 1000 years ago. I do not believe that global

>> >mean annual temperatures have simply cooled progressively over thousands

>> of

>> >years as Mike appears to and I contend that that there is strong evidence

>> >for major changes in climate over the Holocene (not Milankovich) that

>> >require explanation and that could represent part of the current or

>> future

>> >background variability of our climate.

The next day, facing increasing pressure to withdraw his data, Briffa again objected (email #2700, emphasis added).

cc: tkarl@ncdc.noaa.gov, mann@virginia.edu

date: Thu Sep 23 18:29:05 1999

from: Keith Briffa <k.briffa@uea.ac.uk>

subject: RE: IPCC revisions

to: "Michael E. Mann" <mann@multiproxy.evsc.virginia.edu>, "Folland, Chris" <ckfolland@meto.gov.uk>, 'Phil Jones' <p.jones@uea.ac.uk>

Dear Mike (and all)

Some remarks in response to your recent message…

Mike , I agree very much with the above sentiment. **My concern was motivated by the possibility of expressing an impression of more concensus than might actually exist . I suppose the earlier talk implying that we should not 'muddy the waters' by including contradictory evidence worried me . IPCC is supposed to represent concensus but also areas of uncertainty in the evidence.** Of course where there are good reasons for the differences in series (such as different seasonal responses or geographic bias) it is equally important not to overstress the discrepancies or suggest contradiction where it does not exist.

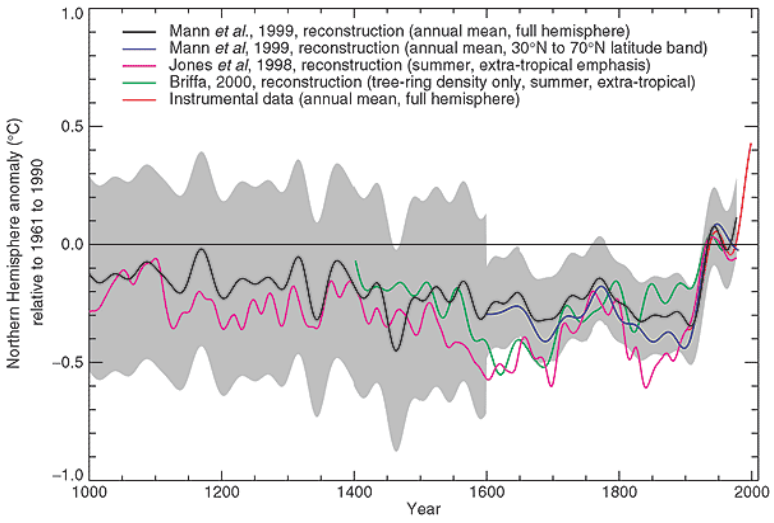
The solution, in the end, was to use Briffa's data, but delete the post-1960 portion (email 4105, emphasis added):

> cc: k.briffa@uea,p.jones@uea
>
> date: Tue, 05 Oct 1999 16:18:29 +0100
>
> from: Tim Osborn <t.osborn@uea.ac.uk>
>
> subject: Briffa et al. series for IPCC figure
>
> to: mann@virginia.edu,imacadam@meto.gov.uk
>
> Dear Mike and Ian
>
> Keith has asked me to send you a timeseries for the IPCC multi-proxy reconstruction figure, to replace the one you currently have. The data are attached to this e-mail. They go from 1402 to 1995, **although we usually stop the series in 1960 because of the recent non-temperature signal that is superimposed on the tree-ring data that we use.** I haven't put a 40-yr smoothing through them - I thought it best if you were to do this to ensure the same filter was used for all curves.

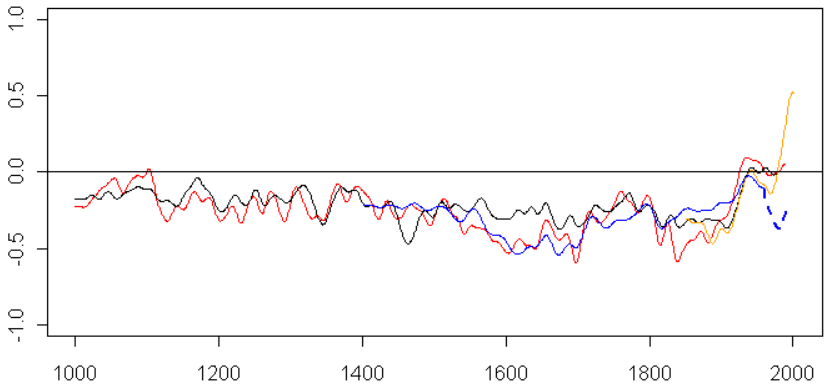
The Figure that appeared in the TAR and again in the AR4 achieved the appearance of consensus by removing the post 1960 Briffa data. No notice was given to the reader, and the matter was not discovered until blogger Steve McIntyre write about it in 2005.[50]

The IPCC (TAR) published the following graph:

[50] See http://www.climateaudit.info/data/climate2003/pages/blog/briffa.mxd.htm.



Had they not deleted the post-1960 Briffa data, the Figure would have looked like this:



Thus a truthful version of the diagram, rather than the one actually published, would have shown that over the interval where we can observe both proxies and thermometer data, the proxies do a poor job of following temperatures, , making the earlier portion questionable as a temperature history.

4.5 Recent evidence on the invalid nature of the proxy reconstructions

Mann et al. (2008),[51] like earlier studies, appends modern temperature records to historical proxies and appears to show the modern world is exceptionally warm in the context of the past 2,000 years.

[51] Mann ME, *et al*. (2008) Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc Natl Acad Sci USA* 105:13252–13257.

McIntyre and McKitrick (2009)[52] showed that a key data series on which the Mann et al. graphs depended, the Tiljander sediment series, was used upside down, and that the statistical calibration of their proxy model is so weak that finite uncertainty limits cannot be derived, in other words, their data do not provide any more information about the historical climate than do random numbers. This was confirmed by a lengthy study in *Annals of Applied Statistics* in 2011 by two statisticians completely independent of the paleoclimate community (McShane and Wyner)[53] who found, among other things, that diametrically opposite conclusions about the warmth of the modern era relative to the medieval era can be derived using the same Mann et al. data by statistical models with equal claims to accuracy.

In sum, the EPA claims about the historical uniqueness of modern temperatures rest on selective of information in the NRC report, data shown by the NRC to be unreliable, and the unreliable evidence in the IPCC report. Detailed consideration of the sources shows that proxy reconstructions remain too uncertain to support claims one way or the other.

5 Reliance on climate models that are not able to represent accurately the atmospheric response to GHG's.

Rebuttal: The TSD dismisses the lack of a troposphere hotspot by relying entirely on the Santer et al. (2009) study, which was refuted by MMH(2010) and others. Climate models are in serious error where it matters most, namely the tropical troposphere.

5.1 Relevance

A key region for modeling the climatic response to greenhouse gases is the vast section of atmosphere in the tropics up to an altitude of 16 km, spanning 20 degrees North to 20 degrees South of the equator. Ever since the first climate models were produced, and in all the modeling work done since, including for the IPCC in its 2007 Report, the theory of amplified greenhouse gas-induced warming implies that

[52] McIntyre, Stephen and Ross R. McKitrick (2009) Proxy inconsistency and other problems in millennial paleoclimate reconstructions *Proceedings of the National Academy of Sciences* February 2, 2009. 106:E10; doi:10.1073/pnas.0812509106

[53] McShane, B.B. and A.J. Wyner, (2011). "A Statistical Analysis of Multiple Temperature Proxies: Are reconstructions of surface temperatures over the last 1000 years reliable?" *Annals of Applied Statistics* Volume 5, Number 1 (2011), 5-44.

warming trends should reach a maximum there, specifically in the mid- and upper-troposphere over the tropics. As emphasized in Fu et al. (2011, p. 1):[54]

> This feature [enhanced warming in the tropical upper troposphere] has important implications to the climate sensitivity because of its impact on water vapor, lapse rate, and cloud feedbacks… It is therefore critically important to observationally test the GCM-simulated maximum warming in the tropical upper troposphere.

A recent survey article by Thorne et al. (2011)[55] summarizes the point as follows:

> Since the earliest attempts to mathematically model the climate system's response to human-induced increases in greenhouse gases, a consistent picture of resulting atmospheric trends has emerged. The surface and troposphere (the lowest 8—12 km) warm with a local maximum trend in the upper levels in the tropics, while the stratosphere above cools.

The IPCC also emphasizes that,[56] according to climate model predictions, warming due to greenhouse gases reaches a maximum in the upper troposphere over the tropics, and that all model runs suggest this pattern ought to be observable in current data.

Fu et al. point out that, since the 1970s, the trend has been for models to predict a larger and larger differential between the warming rate in the tropical lower troposphere compared to the tropical mid- or upper-troposphere. Hence as the theory of greenhouse induced-warming has developed over recent decades, the expectation of enhanced warming in the tropical troposphere has emerged as a central prediction.

But there is considerable empirical evidence that no such warming "hotspot" has been observed since the advent of satellite monitoring in 1979. Many commenters on the EPA Endangerment Finding pointed to the empirical evidence that the combined records from weather balloons and satellites does not support the model predictions of amplified warming in the tropical troposphere.[57] **A significant discrepancy between models and observations on this point would imply a major failure on the part**

---

[54] Fu, Qiang, Syukuro Manabe and Celeste M. Johanson (2011) "On the warming in the tropical upper troposphere: Models versus observations" *Geophysical Research Letters* VOL. 38, L15704, doi:10.1029/2011GL048101, 2011.
[55] Thorne, P. W., J. R. Lanzante, T. C. Peterson, D. J. Seidel, and K. P. Shine (2011), Tropospheric temperature trends: History of an ongoing controversy, Wiley Interdisciplinary Rev. Clim. Change, 2, 66–88, doi:10.1002/wcc.80.
[56] IPCC WGI pp. 763-764; also Figure 9.1.
[57] http://www.epa.gov/climatechange/endangerment/comments/volume3.html

of climate models, directly undermining the soundness of, among other things, the EPA's position. Indeed the 2006 CCSP Report on surface and satellite records, mentioned above, pointed to this problem, as follows:

> A potentially serious inconsistency, however, has been identified in the tropics. Figure 4G shows that the lower troposphere warms more rapidly than the surface in almost all model simulations, while, in the majority of observed data sets, the surface has warmed more rapidly than the lower troposphere. In fact, the nature of this discrepancy is not fully captured in Fig. 4G as the models that show best agreement with the observations are those that have the lowest (and probably unrealistic) amounts of warming.
>
> (Wigley et al. 2006, p. 11)

In 2007, papers by two teams of authors (Christy, Norris, Spencer and Hnilo, and Douglass, Christy, Pearson and Singer) showed that observed data sets contained much less warming than even the lowest model-based predictions. The Douglass et al. paper[58] specifically asserted that the model-data discrepancy is statistically significant. The EPA Response to comments on the Endangerment Finding (3-7) reveals some hesitation on their part concerning this matter:

> EPA is aware of the emerging literature on this issue and the challenges in identifying the anthropogenic fingerprint in the tropics. The TSD's characterization of this issue is consistent with the assessment literature as well as the most recent studies, which find that when uncertainties in models and observations are properly accounted for, newer observational data sets are in agreement with climate model results.

This last statement, it will be shown, is untrue.

5.2 Model-observational discrepancies in the tropical troposphere

The EPA responded to the evidence in the Douglass et al. paper by citing three sources. First, they refer to a paper by Haimberger et al. (2008)[59] which uses a weather balloon series called RAOBCORE version 1.4, which apparently agrees with some model projections. However, Haimberger has since revised the

[58] Douglass, D. H., J. R. Christy, B. D. Pearson, and S. F. Singer (2008), A comparison of tropical temperature trends with model predictions, Int. J. Climatol., 28, 1693–1701, doi:10.1002/joc.1651.

[59] Haimberger, L., C. Tavolato, and S. Sperka (2008), Towards elimination of the warm bias in historic radiosonde records—Some new results from a comprehensive intercomparison of upper air data, J. Clim., 21, 4587–4606, doi:10.1175/2008JCLI1929.1.

RAOBCORE version 1.4 data to remove a spurious warming influence from an input data source.[60] The trend in the lower tropical troposphere in RAOBCORE 1.4 set is now 0.117 degrees C per decade whereas the average predicted trend in climate models for the same region is 0.272 degrees C per decade, more than twice as high. Clearly this data set cannot be the basis for setting aside the commenters' concerns about models overstating warming.

The second paper cited by the EPA is Allen and Sherwood (2008),[61] who use *windspeed* data collected by weather balloons to infer temperature trends. They find higher trends than studies using thermometers to measure temperature trends. The EPA does not provide a discussion of the problems associated with using wind data to infer temperatures. A 2010 paper by John Christy and 8 coauthors[62] in the journal *Remote Sensing* points out that until the advent of modern GPS systems, weather balloons tended to drift out of radio range at high altitudes on the windiest days, leading to an artificial depression of the highest windspeeds in the earlier years of the record, introducing a known source of bias in the trend over time. Also, windspeed data is very limited in the tropics compared to temperature data, and as Christy et al. point out, the temperature trend calculations by Thorne et al. imply windspeeds in the interpolated regions would have to be much higher than those observed in regions that do have data. Consequently, it was inappropriate for the EPA to place greater reliance on this study than on the many studies using direct temperature observations, especially since its method is new and rather speculative.

The third study cited by the EPA, and arguably the one that is key to their position, is a 2008 paper by Ben Santer et al.,[63] asserting that uncertainties in climate models and observations are sufficiently large with regards to trends in the tropical troposphere as to rule out a finding of inconsistency. They reach this conclusion by arguing that Douglass et al. used an incorrect statistical methodology to compare modeled and observed trends, and in the Santer et al. analysis they propose a slight improvement in methods, which they apply to data ending in 1999. They report the uncertainties in the model trends to be sufficiently large as to partially overlap with the uncertainties in the observed trends, leading Santer et al. to conclude that the models-data differences are not statistically significant.

---

[60] The problem apparently was in the ERA-40 reanalysis data.

[61] Allen, R. J., and S. C. Sherwood (2008), Warming maximum in the tropical upper troposphere deduced from thermal winds, Nat. Geosci., 1,399–403, doi:10.1038/ngeo208

[62] Christy, John, et al. (2010) "What Do Observational Datasets Say about Modeled Tropospheric Temperature Trends since 1979?" Remote Sensing 2010, 2, 2148-2169; doi:10.3390/rs2092148

[63] Santer, B. D., et al. (2008), Consistency of modelled and observed temperature trends in the tropical troposphere, Int. J. Climatol., 28, 1703–1722, doi:10.1002/joc.1756.

McKitrick, McIntyre and Herman (2010)[64] showed that the Santer et al. conclusions fail on two grounds. First, neither Douglass et al. nor Santer et al. used appropriate statistical modeling techniques for comparing trends in data sets of the kind under dispute. McKitrick et al. applied two different state of the art statistical methods for trend comparisons, both of which are well-established in the econometrics literature. Second, they extended the data up to the end of 2009 (the maximum extent available at the time of the analysis). Ending the data at 1999, as Santer et al. did, biases the results because there was a large El Nino event in 1998, temporarily boosting the observed trend so much that it appears to match models.

McKitrick et al. found that on the full sample up to 2009, the satellite and weather balloon data sets were not significantly different from each other, but were significantly different from models. In particular, the models predicted two to four times more warming, on average, than is observed in the data, and the differences are statistically very significant.

In light of these updated findings, the EPA's reliance on Santer et al. (2008) is unsound, as is their claim that

> "when uncertainties in models and observations are properly accounted for, newer observational data sets are in agreement with climate model results."

Furthermore, when McKitrick et al. did their analysis on a sample ending in 1999, to match that of Santer et al., they found the model-observation difference marginally significant, an indication of the bias in the Santer et al. method. But in that case they also noted that there is no significant warming trend in the balloon and satellite series when the data are truncated at 1999, something not mentioned by the EPA in its reliance on the Santer et al. results. When the data are extended up to 2009, some of the observational series indicate a significant warming trend, but it is very small compared to model predictions, and the model-observation discrepancy is statistically significant. Thus McKitrick et al. confirm the 2006 observation of the CCSP Report of a "potentially serious inconsistency" between models and data.

The continuing importance of this issue is attested by the Thorne et al. (2010) review, which points out that if observations fail to support the tropospheric warming projected by models this would have

---

[64] McKitrick, Ross R., Stephen McIntyre and Chad Herman (2010)Panel and Multivariate Methods for Tests of Trend Equivalence in Climate Data Sets. *Atmospheric Science Letters* DOI: 10.1002/asl.290.

"fundamental and far-reaching implications for understanding of the climate system." The Thorne et al. review article asserts that the models and observations are in general agreement, but like the TSD it relies for this conclusions entirely on the Santer et al. study and makes no mention of the McKitrick et al. findings.

Additionally, in a peer-reviewed journal article in 2014, Ross McKitrick and Tim Vogelsang extended the analysis back to the late 1950s using weather balloon data and concluded that the model-observation mismatch is highly significant over the entire post-1958 interval. After allowing for a level shift in the observations around 1977, they found is no significant upward trend in the tropical troposphere.[65]

5.3 Other indications of climate model failure

Fu et al. (2011) found that not only do the IPCC climate models exaggerate warming at the surface and each layer above, but they also exaggerate the rate of amplification of warming with height. Fyfe et al. (2011)[66] sought to replicate the 1961-2006 observed global average surface temperature using a prominent climate model, allowing their model to re-initialize with observed outcomes every five years. They reported having to apply repeated corrections to the model trend since it regularly drifts away from reality. Kaufmann and Stern (2004)[67] analysed climate model predictions of the global surface average temperature and asked whether the model had any more information in it than a small number of GCM input series, namely observations on greenhouse gas concentrations, solar irradiance, volcanic dust and atmospheric aerosols, rescaled to represent forcing units on temperature change processes. They could not reject the hypothesis that the GCM added no information to the forecast other than that inherited from the observed forcing series. In other words, the detailed structure of the climate model was itself uninformative; the only information was contained in the input data.

---

[65] McKitrick, Ross R. and Timothy Vogelsang (2014) "HAC-Robust Trend Comparisons Among Climate Series with Possible Level Shifts" *Environmetrics* DOI: 10.1002/env.2294.

[66] Fyfe, J.C., G.J. Boer, G. J., V. Kharin, W.S. Lee, W. J. Merryfield, and K. von Salzen (2011) "Skillful predictions of decadal trends in global mean surface temperature" *Geophysical Research Letters* VOL. 38, L22801, doi:10.1029/2011GL049508, 2011

[67] Kaufmann, Robert K. and David I. Stern (2004) "A Statistical Evaluation of Atmosphere -Ocean General Circulation Models: Complexity vs. Simplicity." Rensselaer Polytechnic Institute Department of Economics Working Paper 0411, May 2004.

6 Reliance on ability of models to project changes in temperature and precipitation at the regional level

Rebuttal: There is no defence of this point in the AR4. Testing of the spatial pattern of regional trends is not done, or is only done using informal visual comparisons. Recent literature shows models fail to explain spatial trend patterns in both temperature and precipitation.

6.1 Relevance

The whole basis of claiming endangerment to US citizens is the assumption that the forecasts generated by climate models are accurate at the regional level within a single country. If that claim is untrue, there are no grounds for concluding that climatic changes will be harmful. Leaving aside, for the sake of argument, the question of whether climate models are correct in saying that greenhouse gas emissions will result in a large global-scale change, a global-scale change may result in: few or no changes to climatic patterns at individual locations; or changes to climatic patterns that yield net benefits locally; or changes to climatic patterns that are potentially costly but are not measurable because the changes that would have occurred over the same period from purely natural causes cannot accurately be forecast. To rule all these possibilities out requires demonstration that climate models can accurately predict temperature and precipitation at the regional level.

6.2 IPCC and CCSP silence on this topic

The discussion of the evaluation of climate models in AR4 Chapter 8 is dominated by *a priori* process checks, that is, whether certain known meteorological processes are coded into the models. Model evaluation at regional levels focuses on static reproduction tests, that is, the ability to reproduce the distribution of mean temperature and precipitation levels, and diurnal temperature ranges, but not temperature trends, around the world. In essence, models are tested on whether they get cold poles and hot tropics. But success on this measure, to the extent it is achieved, is not an indication of accurate ability to forecast the spatial pattern of trends over time, that is, to put warming where warming has been observed and cooling where cooling has been observed. The IPCC report notes (p. 594) that relatively few studies have looked at whether empirical fidelity between model simulations of historical

periods and observations improves the accuracy of climate trend forecasts. Gleckler et al. (2008)[68] note that the ability of a climate model to replicate a mean climate state has little correlation to measured fidelity on interannual trend measures.

Knutti (2008)[69] argues that testing model accuracy over both space and time is necessary for evaluating their credibility. Berk et al. (2001)[70] warned that quantitative comparison of model outputs to observed data was rare and "relies very heavily on eyeball assessments" (Berk et al p. 126). Since then, neither the 2007 IPCC report nor the Climate Change Science Program 2008 review of climate models[71] provided quantitative tests of how well climate models reproduce the spatial pattern of temperature trends in recent decades, relying instead on "eyeball assessments." This validation of the models is a key element of the DQA requirements that EPA was obligated to independently confirm.

AR4 Chapter 9 presents a diagram and accompanying discussion (Figure 9.6, pp. 684-686) of the averaged output from 58 GCM runs and the spatial pattern of temperature trends over land from 1979-2005, comparing model runs under the assumption that greenhouse gases do not warm the climate versus runs that assume they do. It is asserted that the latter assumption fits the data better, but no quantitative evidence is provided. CCSP (2008) presents a visual comparison of the fit between observed trend patterns over 1979-2003 and those generated by a single model, the GISS ModelE. Again the discussion is entirely qualitative—readers were given no statistical scores testing whether the model attains statistically significant validity.

CCSP (2008) reports a 95-98% correlation between modeled and observed temperatures over space and time. However this is not a test of regional trend accuracy. The underlying study is Covey et al. (2003).[72] The tests were not of historical reproduction of observations, but instead ahistorical, no-forcing control runs. Covey et al. were merely testing the ability to reproduce the annual temperature range in each

---

[68] Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), "Performance metrics for climate models," *J. Geophys. Res.*, *113*, D06104, doi:10.1029/2007JD008972.

[69] Knutti, R. (2008) "Why are climate models reproducing the observed global surface warming so well?" *Geophysical Research Letters* 35, L18704, doi:10.1029/2008GL034932, 2008.

[70] Berk, Richard A., Robert G. Fovell, Frederic Schoenberg and Robert E. Weiss (2001) "The use of statistical tools for evaluating computer simulations." *Climatic Change* 51: 119-130.

[71] CCSP (2008): *Climate Models: An Assessment of Strengths and Limitations*. A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research [Bader D.C., C. Covey, W.J. Gutowski Jr., I.M. Held, K.E. Kunkel, R.L. Miller, R.T. Tokmakian and M.H. Zhang (Authors)]. Department of Energy, Office of Biological and Environmental Research, Washington, D.C., USA, 124 pp

[72] Covey, C., K.M. AchutaRao, U. Cubasch, P. Jones, S.J. Lambert, M.E. Mann, T. J. Phillips, K.E. Taylor (2003) "An overview of results from the Coupled Model Intercomparison Project." *Global and Planetary Change* 37 103—133.

region. They compared the 12 monthly means from GCM control runs on a gridcell-by-gridcell basis to late 20th century monthly means in CRU data. The models did a good job reproducing the spatial pattern of the mean across grid cells, the amplitude of the seasonal cycle within each grid cell and the different seasonal amplitudes across grid cells. However the ability to predict *trends* in the gridded monthly or annual means over time in response to observed forcing changes was not tested.

6.3 Published evidence of the lack of climate model validity at regional levels

Knutson et al. (2006)[73] present a comparison of the spatial trend pattern between an ensemble average of simulations from a single model for 1949—2000 and corresponding observations. In their Figure 5d (p. 1635) the differences are denoted as significant or not based on a *t*-test. They report that in 31% of the locations, the *t*-statistic rejects the hypothesis that the modeled and observed trends are the same. However, the apparent failure to reject differences elsewhere does not imply that the models are accurate. Consider for example, a test for different trends among two lists of random numbers. A t-test would typically fail to reject such a difference, but that does not imply that one explains the other. Jun et al. (2008, Figure 7) contrast observed and model trends but the significance of the mismatches is not reported.

McKitrick and Nierenberg (2010), cited above, found that a vector of socioeconomic variables had significant explanatory power for the spatial pattern of temperature trends over land. But the GISS-E climate model, as well as an average formed over all climate models used for the AR4, failed to reproduce the correlation pattern and typically generated the opposite pattern to that observed in the data. They also found that models exhibit a regional pattern of spatial autocorrelation in temperature trends that is not found in a regression model using observational data. While the focus of that study was on detecting non-climatic contamination of surface temperature data, the implication is equally relevant to the present discussion, namely that climate models have been shown to fail to predict the spatial pattern of warming and cooling trends over land, whereas a simple model using data on the spatial pattern of socioeconomic activity has been shown to successfully predict such a pattern.

---

[73] Knutson, T. R.; Delworth, T. L.; Dixon, K. W.; Held, I. M.; Lu, J.; Ramaswamy, V.; Schwarzkopf, M. D.; Stenchikov, G.; Stouffer, R. J (2006) "Assessment of Twentieth-Century Regional Surface Temperature Trends Using the GFDL CM2 Coupled Models." *Journal of Climate* **19**, 1624–1651.

McKitrick and Tole[74] (2012) evaluated three categories of variables for explaining the spatial pattern of warming and cooling trends over land: predictions from 22 general circulation models (GCMs) used by the IPCC for the AR4; geographical factors like latitude and pressure; and socioeconomic influences on the land surface and data quality. As explained in Section 2, the IPCC strongly insists that socioeconomic measures have no explanatory power, a position also maintained by the EPA. Statistical tests showed 20 of 22 GCMs individually contributed either no significant explanatory power or yielded a trend pattern opposite to the observations. Further testing showed that socioeconomic variables have considerable unique explanatory power. McKitrick and Tole used advanced computational methods to examine all possible linear combinations of explanatory variables, showing that only three of the 22 GCMs would ever exhibit any explanatory power, but that the use of socioeconomic measures of temperature data contamination is essential for yielding a valid model of the spatial pattern of trends in land surface temperature records. Climate models from Norway, Canada, Australia, Germany, France, Japan and the UK, as well as American models from GFDL, NOAA and GISS, failed to exhibit any explanatory power for the spatial pattern of surface temperature trends in any test, alone or in any combination. The findings thus show both that the evidence of temperature data contamination is robust and that most climate models fail to provide meaningful data for the purpose of establishing EPA findings.

Some other relevant studies in recent years are shown below.

Koutsoyiannis, D., A. Efstratadis, N. Namassis and A. Christofides (2008) "On the credibility of climate predictions" *Hydrological Sciences,* **53**(4) August 2008

Abstract Geographically distributed predictions of future climate, obtained through climate models, are widely used in hydrology and many other disciplines, typically without assessing their reliability. Here we compare the output of various models to temperature and precipitation observations from eight stations with long (over 100 years) records from around the globe. The results show that models perform poorly, even at a climatic (30-year) scale. Thus local model projections cannot be credible, whereas a common argument that models can perform better at larger spatial scales is unsupported.

[74] McKitrick, Ross R. and Lise Tole (2012) "Evaluating Explanatory Models of the Spatial Pattern of Surface Climate Trends using Model Selection and Bayesian Averaging Methods" *Climate Dynamics*, June 2012 DOI 10.1007/s00382-012-1418-9.

Anagnostopoulos, G. G., D. Koutsoyiannis, A. Christofides, A. Efstratiadis & N. Mamassis (2010). "A comparison of local and aggregated climate model outputs with observed data." *Hydrological Sciences Journal*, 55(7) 2010.

Abstract We compare the output of various climate models to temperature and precipitation observations at 55 points around the globe. We also spatially aggregate model output and observations over the contiguous USA using data from 70 stations, and we perform comparison at several temporal scales, including a climatic (30-year) scale. Besides confirming the findings of a previous assessment study that model projections at point scale are poor, results show that the spatially integrated projections are also poor.

Stephens, G. L., T. L' Ecuyer, R. Forbes, A. Gettlemen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes (2010), "Dreary state of precipitation in global models," J. Geophys. Res., 115, D24211, doi:10.1029/2010JD014532.

Abstract New, definitive measures of precipitation frequency provided by CloudSat are used to assess the realism of global model precipitation. The character of liquid precipitation (defined as a combination of accumulation, frequency, and intensity) over the global oceans is significantly different from the character of liquid precipitation produced by global weather and climate models. Five different models are used in this comparison representing state-of-the-art weather prediction models, state-of-the-art climate models, and the emerging high-resolution global cloud "resolving" models. The differences between observed and modeled precipitation are larger than can be explained by observational retrieval errors or by the inherent sampling differences between observations and models. We show that the time integrated accumulations of precipitation produced by models closely match observations when globally composited. However, these models produce precipitation approximately twice as often as that observed and make rainfall far too lightly. This finding reinforces similar findings from other studies based on surface accumulated rainfall measurements. The implications of this dreary state of model depiction of the real world are discussed.

Fildes and Kourentzes (2011)[75] used standard forecasting evaluation tests to compare the validity of GCM regional temperature forecasts over 1—10 year-ahead horizons. The testing approach compares a forecasting system against an uninformative "random walk" alternative consisting simply of using the last period's value as the forecast for the next period's value. The resulting score ranges from 0 for a perfect forecast up to 1.0 for a forecast method that is no better than the random alternative. A forecasting method receiving a score above 1.0 is deemed worse than uninformed guesses. Simple statistical models typically yielded scores between 0.805 and 0.973, indicating slight improvements on the random walk, though in some cases their scores went above 1.0, in one case as high as 1.762. The GCMs did extremely poorly, however, with scores ranging from 2.386 to 3.732, indicating a complete failure to provide valid forecast information at the regional level. The authors comment (p. 990):

> This implies that the current GCM models are ill-suited to localised decadal predictions, even though they are used as inputs for policy making.

In sum the EPA reliance on the IPCC and CCSP for assertions about endangerment to US regions was inappropriate because those agencies present no evidence to support the view that regional temperature and precipitation forecasts from climate models are accurate. What testing has been done in recent years has shown poor results.

7 Reliance on speculative claims about extreme weather

Rebuttal: Since climate models are not able to predict temperature and precipitation at regional levels, any assertions about trends in extreme weather are purely speculative. The InterAcademy Council found the work of IPCC WG2, which looks at the impacts of climate change, to be full of unsubstantiated conclusions. The recent IPCC report on weather extremes pulled back significantly from suggesting a link between global warming and extreme weather.

7.1 Relevance

The TSD (pp. 61—63) presents a lengthy list of projected worsening climate extremes, relying heavily on the AR4 and a 2008 CCSP product. But the IPCC handling of this issue was heavily criticized by the IAC.

[75] Fildes, Robert and Nikolaos Kourentzes (2011) "Validation and Forecasting Accuracy in Models of Climate Change *International Journal of Forecasting* 27 968-995.

7.2 IAC Report slams the IPCC for its handling of impacts research and its reliance on non peer-reviewed material

The IAC was deeply critical of the way the IPCC, particularly Working Group II, handled and reported on uncertainty, especially in regards to statements about the impacts of climate change. Since Working Group II handles the topic of impacts, this is directly pertinent to the endangerment question. The IAC said:

> The Working Group II Summary for Policy Makers in the Fourth Assessment Report contains many vague statements of "high confidence" that are not supported sufficiently in the literature, not put into perspective, or are difficult to refute.
>
> (p. 37)

The IAC found that the guidance for explaining uncertainty is not itself adequate, and is often not followed anyway (p. 4).

> Many of the 71 conclusions in the "Current Knowledge about Future Impacts" section of the Working Group II Summary for Policy Makers are imprecise statements made without reference to the time period under consideration or to a climate scenario under which the conclusions would be true….In the Committee's view, assigning probabilities to imprecise statements is not an appropriate way to characterize uncertainty. If the confidence scale is used in this way, conclusions will likely be stated so vaguely as to make them impossible to refute, and therefore statements of "very high confidence" will have little substantive value.
>
> (pp. 33-34).

More generally, the IAC noted that in some cases

> [IPCC] authors reported high confidence in statements for which there is little evidence, such as the widely-quoted statement that agricultural yields in Africa might decline by up to 50 percent by 2020. Moreover, the guidance was often applied to statements that are so vague they cannot be falsified. In these cases the impression was often left, quite incorrectly, that a substantive finding was being presented.
>
> (p. 36)

The IAC concluded that "many of the conclusions in the "Current Knowledge about Future Impacts" section of the Working Group II Summary for Policy Makers are based on unpublished or non-peer-reviewed literature" (p. 33). They also found that many conclusions stated with "High Confidence" by Working Group II had little or no scientific basis:

> [By] making vague statements that were difficult to refute, authors were able to attach "high confidence" to the statements. The Working Group II Summary for Policy Makers contains many such statements that are not supported sufficiently in the literature. (p. 4).

The IAC concludes that had Working Group II used a level-of-understanding scale, rather than their "confidence" scale, it would have made clear the "weak evidentiary basis" for many of their conclusions (p. 33).

7.3 IPCC Special Report on Extreme Weather (SREX) more accurately revealed a lack of certainty

The 2012 SREX report provides many striking indications of the lack of certainty around projections of worsening climatic conditions (emphasis added in some places).

Attribution to human causation: IPCC SREX (p. 161) says "The AR4 concluded that it is more likely than not that anthropogenic influence has contributed to increases in the frequency of the most intense tropical cyclones (Hegerl et al., 2007). **Based on subsequent research that further elucidated the scope of uncertainties in both the historical tropical cyclone data as well as the physical mechanisms underpinning the observed relationships, no such attribution conclusion was drawn in the recent WMO assessment** (Knutson et al., 2010). The present assessment regarding detection and attribution of trends in tropical cyclone activity is similar to the WMO assessment (Knutson et al., 2010): the uncertainties in the historical tropical cyclone records, the incomplete understanding of the physical mechanisms linking tropical cyclone metrics to climate change, and the degree of tropical cyclone variability – comprising random processes and linkages to various natural climate modes such as El Niño – provide only **low confidence for the attribution of any detectable changes** in tropical cyclone activity to anthropogenic influences.

Drought frequency - IPCC SREX (page 170): From a paleoclimate perspective recent droughts are not unprecedented, with severe 'megadroughts' reported in the paleoclimatic record for Europe,

North America, and Australia (Jansen et al., 2007). Recent studies extend this observation to African and Indian droughts (Sinha et al., 2007; Shanahan et al., 2009): much more severe and longer droughts occurred in the past centuries with widespread ecological, political, and socioeconomic consequences. Overall, these studies confirm that in the last millennium several extreme droughts have occurred (Breda and Badeau, 2008; Kallis, 2008; Büntgen et al., 2010). In North America, there is medium confidence that there has been an overall slight tendency toward less dryness (wetting trend with more soil moisture and runoff; Table 3-2), although analyses for some subregions also indicate tendencies toward increasing dryness. This assessment is based on several lines of evidence, including simulations with different hydrological models as well as PDSI and CDD estimates (Alexander et al., 2006; Andreadis and Lettenmaier, 2006; van der Schrier et al., 2006a; Kunkel et al., 2008; Sheffield and Wood, 2008a; Dai, 2011). The most severe droughts in the 20th century have occurred in the 1930s and 1950s, where the 1930s Dust Bowl was most intense and the 1950s drought most persistent (Andreadis et al., 2005) in the United States, while in Mexico the 1950s and late 1990s were the driest periods. Recent regional trends toward more severe drought conditions were identified over southern and western Canada, Alaska, and Mexico, with subregional exceptions (Dai, 2011).

Climate extremes projections - IPCC SREX Page 11: Confidence in projecting changes in the direction and magnitude of climate extremes depends on many factors, including the type of extreme, the region and season, the amount and quality of observational data, the level of understanding of the underlying processes, and the reliability of their simulation in models. **Projected changes in climate extremes under different emissions scenarios generally do not strongly diverge in the coming two to three decades, but these signals are relatively small compared to natural climate variability over this time frame. Even the sign of projected changes in some climate extremes over this time frame is uncertain.** For projected changes by the end of the 21st century, either model uncertainty or uncertainties associated with emissions scenarios used becomes dominant, depending on the extreme.

Disaster-related losses – IPCC SREX pp. 268-269: There is high confidence, based on high agreement and medium evidence, that economic losses from weather- and climate-related disasters have increased (Cutter and Emrich, 2005; Peduzzi et al., 2009, 2011; UNISDR, 2009; Mechler and Kundzewicz, 2010; Swiss Re 2010; Munich Re, 2011). A key question concerns whether trends in such losses, or losses from specific events, can be attributed to climate change. In this context,

changes in losses over time need to be controlled for exposure and vulnerability. Most studies of long-term disaster loss records attribute these increases in losses to increasing exposure of people and assets in at-risk areas (Miller et al., 2008; Bouwer, 2011), and to underlying societal trends – demographic, economic, political, and social – that shape vulnerability to impacts (Pielke Jr. et al., 2005; Bouwer et al., 2007). Some authors suggest that a (natural or anthropogenic) climate change signal can be found in the records of disaster losses (e.g., Mills, 2005; Höppe and Grimm, 2009), but their work is in the nature of reviews and commentary rather than empirical research. Attempts have been made to normalize loss records for changes in exposure and wealth. **There is medium evidence and high agreement that long-term trends in normalized losses have not been attributed to natural or anthropogenic climate change** (Choi and Fisher, 2003; Crompton and McAneney, 2008; Miller et al., 2008; Neumayer and Barthel, 2011).

The statement about the absence of trends in impacts attributable to natural or anthropogenic climate change holds for tropical and extratropical storms and tornados (Boruff et al., 2003; Pielke Jr. et al., 2003, 2008; Raghavan and Rajesh, 2003; Miller et al 2008; Schmidt et al., 2009; Zhang et al., 2009; see also Box 4-2). Most studies related increases found in normalized hurricane losses in the United States since the 1970s (Miller et al., 2008; Schmidt et al., 2009; Nordhaus, 2010) to the natural variability observed since that time (Miller et al., 2008; Pielke Jr. et al., 2008). Bouwer and Botzen (2011) demonstrated that other normalized records of total economic and insured losses for the same series of hurricanes exhibit no significant trends in losses since 1900. The absence of an attributable climate change signal in losses also holds for flood losses (Pielke Jr. and Downton, 2000; Downton et al., 2005; Barredo, 2009; Hilker et al., 2009), although some studies did find recent increases in flood losses related in part to changes in intense rainfall events (Fengqing et al., 2005; Chang et al., 2009). For precipitation- related events (intense rainfall, hail, and flash floods), the picture is more diverse. Some studies suggest an increase in damages related to a changing incidence in extreme precipitation (Changnon, 2001, 2009), although no trends were found for normalized losses from flash floods and landslides in Switzerland (Hilker et al., 2009). Similarly, a study of normalized damages from bushfires in Australia also shows that increases are due to increasing exposure and wealth (Crompton et al., 2010).

Increasing exposure of people and economic assets has been the major cause of long-term increases in economic losses from weather- and climate-related disasters (high confidence). The attribution of economic disaster losses is subject to a number of limitations in studies to date: data availability (most data are available for standard economic sectors in developed countries); type of hazards studied (most studies focus on cyclones, where confidence in observed trends and attribution of changes to human influence is low; Section 3.4.4); and the processes used to normalize loss data over time. Different studies use different approaches to normalization, and most normalization approaches take account of changes in exposure of people and assets, but use only limited, if any, measures of vulnerability trends, which is questionable. Different approaches are also used to handle variations in the quality and completeness of data on impacts over time. Finding a trend or 'signal' in a system characterized by large variability or 'noise' is difficult and requires lengthy records. These are all areas of potential weakness in the methods and conclusions of longitudinal loss studies and more empirical and conceptual efforts are needed. Nevertheless, the results of the studies mentioned above are strengthened as they show similar results, although they have applied different data sets and methodologies.

7.3.1 Lack of evidentiary basis for EPA's claim of endangerment

Many more such examples can be found. **In light of the general failure of climate models to accurately make predictions at the regional level, and the admission of the IPCC that there is no evidence of trends in damages from extreme weather and/or a link to greenhouse gas emissions, there is no scientific basis for the EPA's position that greenhouse gas emissions will have specific, endangering effects on US regions is groundless.**

8 Reliance on unsubstantiated assertions about higher illness and mortality from ground-level ozone

Rebuttal: No evidence is presented in the TSD that climate models provide an accurate basis for forecasting changes in local air quality, and the material in Section 6 shows this is unlikely to be true. The evidence as presented in the TSD indicates that: **the projected changes in ozone formation, if true, are nonetheless small; that the increase in atmospheric water vapour**

elsewhere predicted by the IPCC is expected to decrease ozone formation, offsetting the predicted increase from higher temperatures; and a contributor to ozone formation is increased atmospheric stability and reduced severe storm formation, yet the TSD does not account for the offsetting benefits of such changes, should they happen. Statistical analysis shows that changes in ground level ozone within the range cited by the TSD has no plausible health effects.

8.1 Selective use of evidence in Jacob and Winner (2009)

Much of the discussion of air pollution and health in the TSD is boilerplate and not specifically related to climate change, or makes only vague assertions about climate change "affecting" it. The main assertion that climate change will worsen air pollution is made with reference to a single study from 2009.[76]:

> There is now consistent evidence from models and observations that 21st-century climate change will worsen summertime surface ozone in polluted regions of North America compared to a future with no climate change (Jacob and Winner, 2009).

What Jacob and Winner actually say is (p. 56, emphasis added):

> The most important climate variables affecting tropospheric ozone on a global scale are stratosphere-troposphere exchange, lightning NOx, and water vapor. These three variables are all expected to increase in the future climate; the first two cause an increase in ozone and the third a decrease. Different models thus project changes in the global tropospheric ozone burden over the 21st century ranging from -5% to +12% (Wu et al., 2008b). Despite this disagreement in sign, **the models agree that climate change will decrease the ozone background in the lower troposphere** where the water vapor effect is dominant (stratosphere-troposphere exchange and lightning are more important in the upper troposphere). An ensemble analysis of 10 global GCM-CTMs by Dentener et al. (2006) indicates **a decrease of annual mean surface ozone in the northern hemisphere** by 0.8 ±0.7 ppb for 2000–2030 climate change, with the standard deviation describing the spread between models.

[76] Jacob, D.J., and D.A. Winner (2009) Effect of climate change on air quality , Atmospheric Environment., 43: 51-63.

They then point out that the only places where ozone is expected in climate models to increase is in selected cities where ozone is already elevated (p. 57):

> A general finding among models is that the ozone increase from climate change is largest in urban areas where present-day ozone is already high…Most models also find that the sensitivity of ozone to climate change is highest during pollution episodes (Hogrefe et al., 2004; Tagaris et al., 2007;Wu et al., 2008a), although some studies do not find such an effect (Murazaki and Hess, 2006; Lin et al., 2008a).

But they do not indicate if this covers a large geographic area or a large fraction of the US population or a large fraction of the year. **The TSD fails to point out that peak ozone episodes are brief, lasting only** a few days typically, so even if the effect is real, it only affects a small number of days during the year.

8.2 Small scale of changes

The increases in ozone levels are projected to be small. The TSD (p. 77) says (emphasis added):

> The EPA Interim Assessment (2009) suggests that climate change effects on ozone grow continuously over time, with evidence for significant increases emerging as early as the 2020s. The EPA Interim Assessment (2009) and the IPCC (Field et al., 2007; Wilbanks et al., 2007) cite a study which evaluates the effects of climate change on regional ozone in 15 U.S. cities, finding that **average summertime daily 8-hour maximum ozone concentrations could increase by 2.7 ppb in the 2020s and by 4.2 ppb in the 2050s under the A2 (high-end) scenario**.

These amounts are very small. It is already the case that current ozone exposure levels have been shown to be unlikely to be associated with major health effects or mortality. Koop and Tole (2004)[77] conducted a thorough analysis of Toronto Ontario, which has an air pollution profile similar to many US cities, and found that previous studies had likely overstated the effect of ozone due to failure to control for model uncertainty, or the "cherry-picking" effect when researchers get to pick and choose which variables will go into their explanatory models. Koop and Tole employed a numerical search method called Bayesian Model Averaging that searches over all variable groupings and weighs each model by the support it gets in the data. The distribution of coefficient estimates in a regression relating ozone levels to mortality turned out like this:

---

[77] Koop, Gary and Lise Tole (2004) "Measuring the Health Effects of Air Pollution: To What Extent Can We Really Say That People are Dying from Bad Air?" *Journal of Environmental Economics and Management* 47: 30—54.

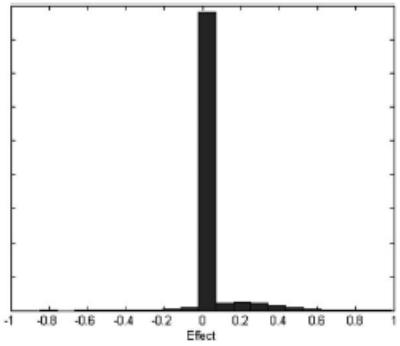


Fig. 1. Posterior of cumulative effect of O_3.

Most of the probability is associated with models in which the effect of ozone is zero, or in other words, models that assume ozone has no relation to mortality obtain the highest support in the data. However, a small tail of probability attaches to models that assign a positive effect on mortality, and if a researcher were to impose that assumption it might be possible to generate statistical findings that "prove" a link between ozone and health.

Koop and Tole also investigate the probability of other forms of air pollution affecting mortality and again find that the result is usually an artifact of using a statistical model that imposes the effect by assumption. Once the prior assumption is relaxed, the effects disappear (p. 42):

Table 4
Probability of including each group of explanatory variables

Explanatory variable	SO_2	CO	O_3	$PM_{10-2.5}$	$PM_{2.5}$	PRESSURE	TEMP
SO_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CO		0.000	0.000	0.000	0.000	0.000	0.000
O_3			0.000	0.000	0.000	0.000	0.000
$PM_{10-2.5}$				0.000	0.000	0.000	0.000
$PM_{2.5}$					0.000	0.000	0.000
PRESSURE						0.807	0.000
TEMP							1.000

The TSD claims (p. 78), relying solely on IPCC model forecasts, that

> Holding population, dose-response characteristics, and pollution prevention measures constant, ozone-related deaths from climate change in the New York City metropolitan area are projected to increase by approximately 4.5% from the 1990s to the 2050s (under the high-end IPCC A2 scenario) (Field et al., 2007).

Viewed in the appropriate context, **namely the lack of ability of climate models to generate meaningful forecasts at the scale of a single city, and the offsetting effects of climate change on ozone formation with a net negative effect overall, and the small scale of changes involved, and the likelihood that current levels are not associated with mortality effects, this statement can be seen as meaningless.** The sheer repetition of unsubstantiated claims such as this in the TSD marks out the entire document as tendentious and unreliable as a basis for major policy decisions.

9 The claim that changes are likely to cause other problems in the US that dominate any benefits.

Rebuttal: At this point the entire discussion is speculative, relies on the IPCC at face value, ignores evidence of adaptive capacity and fails to take proper account of the uncertainties and limitations of climate modeling.

10 Summary

The TSD on which the EPA Endangerment Finding was based is a Highly-Influential Scientific Assessment in the meaning of the US OMB guidelines, since it has led to more than $500 million in regulatory action. But the EPA never conducted a HISA-compliant review of the science, instead it relied on reports of the IPCC, NRC and USGCRP. This was invalid since the IPCC review process is not compliant with HISA requirements, nor has its review record been adequately disclosed, nor do its reports provide evidentiary bases for many of the positions asserted by the EPA. In one case the EPA relies on material known to have been fabricated and subsequently retracted by the IPCC. The EPA also uses data and methods specifically condemned by the NRC in the report it supposedly relies on.

Since it is unrealistic to suppose that climate policy will never cost more than $500 million, the Administrator must rectify the situation by suspending the Endangerment Finding and directing EPA staff to undertake a HISA-compliant review of the science.