

Circling the Bandwagons: My Adventures Correcting the IPCC

by Ross McKittrick
Professor of Economics
University of Guelph

March 2010

Introduction

This is the story of how I spent 2 years trying to publish a paper that refutes an important claim in the 2007 report of the Intergovernmental Panel on Climate Change (IPCC). The claim in question is not just wrong, but based on fabricated evidence. Showing that the claim is fabricated is easy: it suffices merely to quote the section of the report, since no supporting evidence is given. But unsupported guesses may turn out to be true. Showing the IPCC claim is also *false* took some mundane statistical work, but the results were clear. Once the numbers were crunched and the paper was written up, I began sending it to science journals. That is when the runaround began. Having published several against-the-flow papers in climatology journals I did not expect a smooth ride, but the process eventually became surreal.

In the end the paper was accepted for publication, but not in a climatology journal. From my perspective the episode has some comic value, but I can afford to laugh about it since I am an economist, not a climatologist, and my career doesn't depend on getting published in climatology journals. If I was a young climatologist I would have learned that my career prospects would be much better if I never write papers that question the IPCC.

I am taking this story public because of what it reveals about the journal peer review process in the field of climatology. Whether climatologists like it or not, the general public has taken a large and legitimate interest in how the peer review process for climatology journals works, because they have been told for years that they will have to face lots of new taxes and charges and fees and regulations because of what has been printed in climatology journals. Because of the policy stakes, a bent peer review process is no longer a private matter to be sorted out among academic specialists. And to the extent the specialists are unable or unwilling to fix the process, they cannot complain that the public credibility of their discipline suffers.

The Background Issue: Temperature Data Contamination

Note: if you already know what the issues are you can skip this section.

Climate data is *supposed* to be a measure of something that does not, in most places, even exist. In most (inhabited) places around the world, the temperature outside would only be the climate if you happen to live where no one has ever changed the surroundings, either through deforestation or agriculture or road-building or urbanization or any other modification of the landscape. Otherwise, some of the temperature changes can be attributed to the local modifications. Even if you can find undisturbed locations, you might not have records that go back very far. Or you might have a long sequence of records collected using different instruments that need to be joined together somehow. Or you might have

records with gaps in them, or records that weren't all sampled at the same time of the day, other such problems. But more likely, you will have temperature data series collected in places that gradually got built up over time through urbanization. After all, we mostly collect temperature data in places where people live. Producing a long time series of climate data requires making a lot of assumptions about how the various dribs and drabs of temperature data around the world need to be adjusted and tweaked and modified in order to reveal the continuous "climate signal," as it is called. In other words, all the changes in the recorded temperatures that are caused by things other than climate change need to be filtered out of the data: urbanization, deforestation, equipment modification, etc. These are called "inhomogeneities."

It would be fine if the climate signal were large and the inhomogeneities were small. But it is the other way around. We are looking for changes measured in tenths or hundredths of a degree per decade, using data from weather stations where the inhomogeneities can easily shift the record by several degrees. So the adjustment rules matter. Whenever you see a climate data series, such as the so-called "global temperature", bear in mind that what you are seeing is the output of a model, not a reading from a scientific instrument. Thermometers produce some of the basic input data, but the models take over from there. Maybe the models are just right, applying perfect adjustments and not a titch more. Maybe they really do filter out all the inhomogeneities, yielding an output series so precise that we can talk meaningfully of changes on the order of a few hundredths of a degree per decade, confident that it's a pure climate signal.

The Intergovernmental Panel on Climate Change (IPCC) certainly thinks so. In the 4th Assessment Report, as in the previous three, the claim that their data are wonderfully adjusted to remove non-climatic contamination was invoked several times. It is an essential, bedrock assumption behind all key IPCC conclusions. Global temperature trends were presented in Table 3.2 on page 243 of the IPCC Report (Working Group I). The accompanying text (page 242) states that the 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 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. Make a note of that: Section 3.2 will turn out to be important.

IPCC Chapter 9 provides the summary of evidence attributing warming to greenhouse gases. The problem of 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 case for ignoring the issue of data quality problems consists of a citation to IPCC Chapter 3.

Lots of scientists believe the climate data are just fine. To take one example, a paper by Mikyoung Jun and coauthors in the *Journal of the American Statistical Association* in 2008 used surface climate data to test some properties of climate models. Like countless papers before them, they had to consider, if only for a brief moment, whether the climate data on which their analysis rested was garbage or not. They dispensed with the question as follows (page 935 of their paper):

Inhomogeneities in the data arise mainly due to changes in instruments, exposure, station location (elevation, position), ship height, observation time, urbanization effects, and the method used to calculate averages. However, these effects are all well understood and taken into account in the construction of the data set.

Ah, there you go. A little further on in their paper, after they had observed some discrepancies between model-generated and observed trends (which they label as D_i), they explain why they do not attribute them to data contamination as follows:

[Climate] scientists have fairly strong confidence in the quality of their observational data compared with the climate model biases. Therefore, we assume that the effect of observational errors to D_i is negligible.

There's your proof: the scientists are confident. But I am old-fashioned enough to want to know *why* they are so confident. Have they really drilled into the issue? Let's be clear what the stakes are in all of this. If the effects are *not* well understood and they are *not* taken into account in the construction of the data set then *all* the analysis using climate data over land may be flawed. That includes all the studies that claim to have detected global warming and attributed it to greenhouse gases.

A key basis for the IPCC's claim comes the UK's Climate Research Unit (CRU) itself. The CRU is famous for providing the CRUTEM surface climate data set relied upon by so many researchers and by the IPCC. The CRU web page <http://www.cru.uea.ac.uk/cru/data/hrg/> references several data products. One series is called CRU TS 1.x, 2.x and 3.x. The TS series are interesting for our purposes because they are not subject to adjustments for non-climatic influences. Users are explicitly cautioned not to use the TS data for the kind of climatic analysis found in IPCC reports. The 1.2 release of this product (CRU TS 1.2, http://www.cru.uea.ac.uk/cru/data/hrg/timm/grid/CRU_TS_1_2.html), had 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. 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.

In contrast, the primary purpose for which CRU TS 2.0 has been constructed is to permit environmental modellers to incorporate into their models as accurate a representation as possible of month-to-month climate variations, as experienced in the recent past. Therefore influences from urban development or land use change remain an integral part of the data-set. We emphasise that we use *all* available climate data.

If you want to examine the detection of anthropogenic climate change, we recommend that you use the [Jones](#) temperature data-set. This is on a coarser (5 degree) grid, but it is optimised for the reliable detection of anthropogenic trends.

The link attached to Jones' name leads to <http://www.cru.uea.ac.uk/cru/data/temperature/>, the home page for the HadCRUT data products (the land-only portion is CRUTEM). The clear implication is that users will find therein data that have been adjusted to remove non-climatic influences. Readers are referred to

some academic papers for the explanation of the process. Those papers don't actually tell you how it is done, they mostly tell you that *something* was done, and the remaining inhomogeneities are small. For instance, one of them (published in 1999) explains the adjustments as follows.

"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])."

The cited papers from the 1980s are technical reports to the US Department of Energy, referring to the construction of data sets that took place in the early 1980s. These are irrelevant since we are concerned with temperature data collected after 1980. In the early 1980s they would not have known what adjustments would be needed for data collected in the late 1990s. And in any case, another CRU paper published in 2005 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 possible since "no such complete meta-data are available." So in that paper the authors apply an assumption that the bias is no larger than 0.0055 degrees per decade. Round that up to 0.006 degrees and you have the basis for the IPCC claim I quoted earlier, from the Summary for Policymakers, that urban heat islands "have a negligible influence (less than 0.006°C per decade over land...)."

Absence of Evidence versus Evidence of Absence

Various researchers have looked at the question of whether CRU data are biased, and have come to mixed conclusions. The CRU cites a couple of papers that argue that the CRUTEM data are not contaminated with nonclimatic biases. Those papers are from a UK Met Office scientist named David Parker who has from time to time collaborated with CRU staff. So it is not exactly independent work. The papers base their conclusions on the failure to find a difference between warming on windy nights versus calm nights. But this is a problematic style of argument. Sometimes a researcher fails to find an effect because the effect really isn't there. But sometimes the effect is there, the study was just poorly-designed, or the statistical analysis was sloppy, or the researcher looked in the wrong place. That is why an argument based on the failure to find an effect is tentative at best.

What is needed in such contexts is to ask, first, whether the method that didn't find the effect is generally viewed as the right way to look for it, and second whether other researchers looked at the same problem in a different way and *did* find an effect. The answer to the first question is no. Papers have been published in good journals arguing that the method that didn't find the effect might fail to find it even if it is there. And the answer to the second question is yes: two different teams, working independently, published strong evidence of the effect. So taking scientific literature regarding the CRU surface temperature data as a whole, it is legitimate to point to the Met Office study that failed to find evidence of contamination. But it is not legitimate to stop there.

I was on one of the teams that published evidence of the effect. I worked with climatologist Patrick Michaels. Just before we published our results in 2004 a team of Dutch meteorologists (Jos de Laat and Ahilleas Maurellis) published a paper showing they had also asked the question in a different way and found the effect. They and we also published follow-up papers extending our results on new data sets.

Here is what Pat and I did. We started with two temperature data sets, each with observations from 1979 to 2000 for 218 locations around the world. The first data set, which we got from NASA, was, like the TS data, unadjusted for inhomogeneities. We fit a linear trend at each location. That gave us a spatial

pattern of temperature trends. Then we got data on climatological variables for the same 218 locations that would plausibly explain the pattern of trends. We obtained the spatial pattern of temperature trends as measured by weather satellites in the lower troposphere, as well as measures of local mean air pressure and a dryness indicator, latitude, and proximity to a coastline.

Then we added in data on socioeconomic variables. In the unadjusted data we expected to find that indicators of local industrial activity, data quality and other socioeconomic variables would have effects on the observed temperature trends. We used multiple regression analysis to show that this was, indeed, the case. The correlations were very strong, even after controlling for the climatic and geographic effects.

Then we took the CRU data—the adjusted series that everyone says is free from such signals. And guess what: the correlations were smaller but still large and statistically significant. The spatial pattern of trends in CRU temperature data does not appear just to be measuring climate, it is partly measuring the spatial pattern of industrialization.

Our paper was published in the journal *Climate Research* in 2004. There was some excitement when a blogger found a minor error in our computer code (we had released the code at the time of publication), but we sent a correction to the journal right away and showed that the results hardly changed. A comment was submitted to the journal claiming that our results failed a cross-validation test. But the test was overly extreme: it consisted of using only Southern Hemisphere data and a subset of explanatory variables and showing that the resulting estimation did not form very good predictions of the omitted Northern Hemisphere data. Nobody had ever used a test like that before. We were able to show that our model passed a more reasonable and conventional cross-validation test.

de Laat and Maurellis published a second paper in 2006 extending their earlier findings. Pat and I did as well, in 2007, but it did not appear until after the IPCC had issued a deadline for the appearance of new papers that could be cited in their report. Nonetheless, taking our paper and the two Dutch papers together, as well as the papers questioning the Met Office method that had failed to find evidence of contamination, the peer-reviewed literature now presented a strong case that the CRU surface temperature data was compromised by non-climatic biases. The studies also suggested that the contamination effects added up to an overstatement of warming.

Email, July 8 2004

The scientist who runs the CRU and takes primary responsibility for the quality of the data is Phil Jones. According to the climategate emails (eastangliaemails.com), on July 8 2004 Phil Jones wrote to Michael Mann as follows.

```
From: Phil Jones <p.jones@xxxxxxxxxxx.xxx>  
To: "Michael E. Mann" <mann@xxxxxxxxxxx.xxx>  
Subject: HIGHLY CONFIDENTIAL  
Date: Thu Jul 8 16:30:16 2004
```

```
Mike,  
Only have it in the pdf form. FYI ONLY - don't pass on. Relevant  
paras are the last 2 in section 4 on p13. As I said it is worded  
carefully due to Adrian knowing Eugenia for years. He knows the're  
wrong, but he succumbed to her almost pleading with him to tone it  
down as it might affect her proposals in the future !
```

I didn't say any of this, so be careful how you use it - if at all. Keep quiet also that you have the pdf.

The attachment is a very good paper - I've been pushing Adrian over the last weeks to get it submitted to JGR or J. Climate. The main results are great for CRU and also for ERA-40. The basic message is clear - you have to put enough surface and sonde obs into a model to produce Reanalyses. The jumps when the data input change stand out so clearly. NCEP does many odd things also around sea ice and over snow and ice.

The other paper by MM is just garbage - as you knew. De Freitas again. Pielke is also losing all credibility as well by replying to the mad Finn as well - frequently as I see it.

I can't see either of these papers being in the next IPCC report. Kevin and I will keep them out somehow - even if we have to redefine what the peer-review literature is !

Cheers
Phil

Prof. Phil Jones
Climatic Research Unit Telephone +44 (0) 1603 592090
School of Environmental Sciences Fax +44 (0) 1603 507784
University of East Anglia
Norwich Email p.jones@xxxxxxxxxxx.xxx
NR4 7TJ
UK

I have underlined the juicy part. 'MM' is McKittrick and Michaels, i.e. Pat's and my 2004 paper. De Freitas is Chris de Freitas, the editor who handled our paper. Jones is alluding to him because de Freitas had earlier handled a paper that questioned another doctrine of the IPCC, namely that the medieval era was colder than the present. The climategate emails contain many heated exchanges around that issue wherein people like Jones and Mann discuss whether to ~~submit a reasoned critique of the paper~~ launch a campaign to destroy the reputation of *Climate Research*. In the earlier instance the controversy prompted the publisher of *Climate Research* to conduct a review the handling of the file, which concluded that de Freitas had done a good job. But the well was poisoned by then and several editorial board members quit.

The refereeing process Michaels and I went through was long and detailed. We had four referees and initially none of them liked the paper. It took three rounds to settle all the objections before they approved publication, and de Freitas had made it clear he would not proceed without support of all the referees.

Not that the name of the editor has anything to do with anything. One of the patterns I have encountered in response to this work has been that critics begin by saying "I don't believe your results, because of X," where X is some technical objection. But when I respond either by showing that X is irrelevant, or that when I take it into account the results don't change, the critic replies "I don't care, I still don't believe them." In other words, the stated objection is usually just a red herring: the critics just hate the results. In 2006 I presented the findings of a follow-up study using a new and larger data base, which yielded nearly identical findings, at a conference at Los Alamos in the US. Chris Folland of the UK Met Office stood up and objected to the results, saying that the results are a fluke due to strengthened atmospheric circulation effects over Europe, which I hadn't controlled for. So I asked, if I take Europe out of my sample and I get the same results, would you believe them then. After a moment's thought he said No, I still wouldn't believe them.

When our *Climate Research* paper came out in mid-2004, many of our critics pounced on the programming error and said, in effect, we're *wrong* because there was an error in the calculation of the cosine of latitude. But when we fixed that and the results held up, they still refused to believe them. They moved on to saying we were wrong because we hadn't applied an adjustment for "error clustering," which is a legitimate concern for the kind of data we used, and because our sample was not truly global. So when I did the analysis for the follow-up study I used a clustering adjustment on the error term, and the results remained very strong. That paper was published in the *Journal of Geophysical Research* in 2007. The critics who had pounced on the cosine error and the error clustering issue moved on again, rejecting the new findings by saying that we had a problem of spatial autocorrelation, which means that the trends in one location are influenced by the trends in the surrounding region, which can bias the significance calculations. The source of that objection was Rasmus Benestad, a blogger at realclimate.org, but he didn't present any statistical proof of the problem. So I wrote a program that tested for spatial autocorrelation and ran it and showed that the results were not affected by it, and if I applied an adjustment for it anyway the findings remained the same. I even submitted it to the *Journal of Geophysical Research*, but the editor wrote back and said that he couldn't publish it because it was a reply to a comment that no one had submitted. Good point, I thought, so I wrote Rasmus an email, sending him my paper and the editor's note, and I suggested he write up his blog post as a proper comment and send it to the *JGR*, so his comment and my reply could be sent out for peer review. Rasmus wrote me back on December 28 2007 saying that he would think about it,

"but I should also tell you that I'm getting more and more strapped for time, both at work and home. Deadlines and new projects are coming up..."

(ellipsis in original).

Yes, yes, time pressures: I understand. I guess blogging takes up a lot of time. Well it's been three years and he still hasn't sent his comment to the *JGR*. You can forget about it Rasmus: I got the material published elsewhere.

So when Jones said my paper was "just garbage" and referred to "De Freitas again," it's not as if he'd have taken a different view had the editor been someone else. I can't find any indication that Jones has ever given a careful read to either of our papers on the subject. His reaction is purely emotional. Granted, a man is entitled to his emotions and biases, but the problem is that Jones had by this point accepted an invitation to serve as an IPCC Coordinating Lead Author. That means he was going to be in the position of reviewing the published evidence on, among other things, the question of whether the CRU data was contaminated with non-climatic inhomogeneities. Now it's true that the IPCC put him in a conflict of interest. The IPCC needed someone to write a section of the report that examined Jones' papers as well as those of Jones' critics and then offer a judgment on whether Jones was right or not. So they asked Jones to write it. Even though Jones is a common last name, you would think they could have found at least one person on Earth to do that job whose last name was not Jones. And you would think that an agency bragging about having 3,000 brilliant scientists involved with it could figure out how to avoid such conflicts of interests.

Nonetheless, once Jones accepted the invitation he gave up the right to be biased on behalf of the CRU. Yet in the email, sent a year before the IPCC Expert Review process would begin, he was already signaling his determination to block any mention of a paper that had provided significant statistical evidence that the data he supplied to the IPCC was potentially contaminated.

Keeping it Out of the IPCC

As it turns out he did keep it out of the IPCC Report, at least for a while. The first draft of the IPCC Report that went to reviewers contained no mention either of my paper or the deLaat and Maurellis papers. He was certainly aware of my paper since he mentioned it in the email a year earlier. I objected to the omission, as did another reviewer (Vincent Gray). In the second draft there was still silence on the subject. So I objected even more and wrote a longer challenge to the section. Expert review closed in June 2006. As of that point there was no mention of the MM paper in the IPCC Report.

Reviewers did not get to see the responses of the Lead Authors to our comments until long after final publication. Jones (and/or the other Coordinating Lead Author, Kevin Trenberth) wrote the following response to my first draft comments:

Rejected. The locations of socioeconomic development happen to have coincided with maximum warming, not for the reason given by McKittrick and Mihaels [sic] (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.

The bit about the strengthening of the Arctic Oscillation (AO) is bizarre on many levels. The AO is a multidecadal cycle in prevailing winds over the Arctic region, that can affect the severity of winters in northern regions. Our study used data on locations around the world, including the south end of South America and Australia. The IPCC Report doesn't even attribute warming patterns in the *Arctic* to the Arctic Oscillation, let alone patterns in the further reaches of the Southern Hemisphere. The comparison of land and ocean is irrelevant since our study only looks at the land areas—there isn't much industrialization over the open ocean.

Because all mention of my work and that of de Laat and Maurellis had been kept out of the IPCC drafts I assumed it would likewise be kept out of the published edition. So it was not until late 2007 that I became aware that the following paragraph had been inserted on page 244 of the Working Group I report.

McKittrick 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.

(I added the underlining.)

The first point to dispense with is the reference to Sections 3.2.2.7 and 3.6.4 in support of the claim that “the locations of greatest socioeconomic development are also those that have been most warmed by atmospheric circulation changes.” There is nothing whatsoever in either section that supports the point. In neither section is there any discussion of industrialization, socioeconomic development, urbanization or any related term. Section 3.2.2.7 presents a spatial map of warming trends since 1979. In the accompanying text they state that “Warming is strongest over the continental interiors of Asia and northwestern North America and over some mid-latitude ocean regions of the [Southern Hemisphere] as

well as southeastern Brazil.” These are the regions of greatest socioeconomic development? The continental interior of Asia suffered economic decline after 1990, and northwestern North America is sparsely-populated alpine forest, so the claim is rather unlikely to be true. Certainly Section 3.2.2.7 does not try to argue the point. Section 3.6.4 is a discussion of the North Atlantic Oscillation and the Northern Annular Mode, two oscillation patterns related to air pressure systems in the Northern Hemisphere. The section discusses seasonal weather patterns associated with these oscillation systems. Again there is no mention of spatial patterns of socioeconomic development, industrialization, urbanization or any related concept. Hence the citations to these sections serve only to mislead casual readers into thinking there is some kind of support for the statements.

The second point concerns the claim that the correlation in question “ceases to be statistically significant.” Statistical significance is a scientific term with a specific numerical interpretation. A statistical hypothesis test has an associated p value which indicates the probability that the score would be as large as it is if the hypothesis is false, i.e. there is no effect, only randomness, in the data. If a test score has a p value below 0.05, in other words less than a 5%, then the effect is said to be statistically significant. If p is greater than 0.05 but below 0.10 the effect is said to be weakly, or marginally significant. If p is greater than 0.10 then the effect is said to be statistically insignificant. The claim that a published result is statistically insignificant implies that the accompanying p value exceeds 0.10. These are standard, well-known statistical terms.

The effects reported in MM2004 had p values on the order of 0.002 or 0.2%, indicating significance. The sentence in the IPCC Report is worded awkwardly, but can be interpreted either as asserting that the correlations between socioeconomic development and temperature trends are statistically insignificant, or that upon controlling for the influence of atmospheric circulations they become statistically insignificant. On the first interpretation the statement is a plain old porkie since the p values reported in MM2004 are below 1%. On the second interpretation, the implication is that the relevant p value exceeds 0.1 upon introduction of variables controlling for the oscillation effects. Yet no p values are presented, nor is there a citation to any external source, peer-reviewed or otherwise, in which such information is presented, nor are readers supplied with any data, statistical tests, or evidence of any kind in support of the sentence. In other words the claim is a fabrication.

To my eyes it looks like the appropriate word to describe the new paragraph is either “lie” or “fabrication.” Evidence sufficient to disprove either accusation can be defined very precisely: it would consist of the p value supporting the claim of statistical insignificance, the peer-reviewed journal article in which it was presented, and the page number where the study is cited in the IPCC Report.

These things do not exist. Draw your own conclusions.

The Journal Game

The absence of supporting evidence for the IPCC claim was obvious enough, and I drew attention to it in a National Post op-ed in December 2007. However I knew that it was also incumbent upon me to show in a peer-reviewed article whether the claim was false or not.

Actually it ought to have been incumbent upon the IPCC to show that their claim was *true* before dismissing the evidence of contamination in the temperature data underpinning all their main conclusions. Unfortunately, the way the IPCC works, they are allowed to make stuff up, then it's their critics job to prove it is untrue.

So in late 2007 and early 2008 I wrote a paper that tested the claim that controlling for atmospheric circulation effects would overturn our earlier results. I obtained values of the effects of trends in the Arctic Oscillation, North Atlantic Oscillation, Pacific Decadal Oscillation and the El-Niño Southern Oscillation on the gridded surface trend pattern over the 1979-2002 interval. I re-did the regression results from my 2004 and 2007 papers after adding these variables into the model. I showed that the original results did not become statistically insignificant. I wrote the paper up and sent it out for publication.

The next part of the story involves a sequence of eight journals. I am not going to discuss them in exact chronological order, I am going to start with one of the later journals in the list, the *Journal of the American Statistical Association*. JASA is the top statistical journal in the world. My paper was reviewed by two referees, an Associate Editor and the Editor. The Editor told me that all the reviewers were impressed by the paper, but because the methods were so ordinary it lacked any cutting-edge statistical insights.

All of us agree that the paper presents a thoughtful and strong analysis. We also agree that the nature of the paper is very different from what usually appears in JASA. The referees recommended rejection on those grounds, and the AE was ambivalent. I am also torn---technically, a JASA paper need not be methodologically fresh, but essentially all of them are, and this trend has grown stronger over time...So, after some long thought, I believe everyone is better served if you submit this quite good paper to a different venue.

The Associate Editor wrote:

Both referees (and I) agree that the data analysis presented in the paper is carefully and well done. Both also state, however, that the paper would be best targeted in a scientific journal (e.g., Nature or Science) than in JASA. Their reasoning is that the methods used here are mundane, based primarily in linear models and t/F significance tests.

I agree with the referees that this paper has excellent prospects, and a likely greater impact, in the scientific literature. I also agree that the typical JASA A&CS paper brings to bear more sophisticated statistical techniques, and that the relatively mundane methods used in the present paper make it a less than ideal fit for the journal. Accordingly, I think it is reasonable to encourage the authors to submit this paper to another venue, especially a scientific journal. As it is, this is a fine paper, but it offers little in statistical direction, even in the sense of broadening understanding of the problem or area, and would fit much better elsewhere.

One of the reviewers said:

This is a careful data analysis of an important problem in climatology. The author makes a convincing case that gridded surface temperature data are contaminated by effects of urbanization notwithstanding the conclusions of the IPCC.

However, the statistical methods are mundane and quite standard. Thus, it is quite different than the usual JASA applications paper. In other words, this is a good paper for a scientific journal but less well suited for a statistics journal.

The other one said:

Although the scientific problem is interesting and important, the statistics in the paper may not be enough to fit a top statistical journal like JASA. I would suggest the author(s) try Nature/Science or a geophysical journal which might be a good fit.

This was a pretty encouraging set of responses. The Associate Editor and the second reviewer even suggested I should send it to *Nature* or *Science*, the most famous science journals in the world.

As it happens I had already done so. The first journal I had tried, back in March 2008, was *Science*. They thanked me but returned the paper without review, saying the topic was too specialized for them. I then sent it to *Nature* in April 2008. They too declined to send it out for review. They returned it with the comment that there have already been numerous papers published to date arguing that land use change has left persistent effects in the surface temperature record, and my analysis did not provide any major advance in determining the magnitude of this problem. As a result, while they did not have any doubts about the quality of my analysis (at least none that they mentioned), they did not think it was suitable for publishing in *Nature* and suggested I send it to a more specialized journal.

In early May I sent a pre-submission inquiry to the *Bulletin of the American Meteorological Society*. The *BAMS* website instructs authors to send an email to the editor describing the paper, prior to making a full submission. The editor will, presumably, advise on the suitability of the paper and will indicate whether a full submission is requested. My emailed proposal went in on May 2 2008.

A month passed without response. On June 4 2008, having heard nothing in reply I sent in a second email asking what was the timeline for hearing a response.

Another month passed without response.

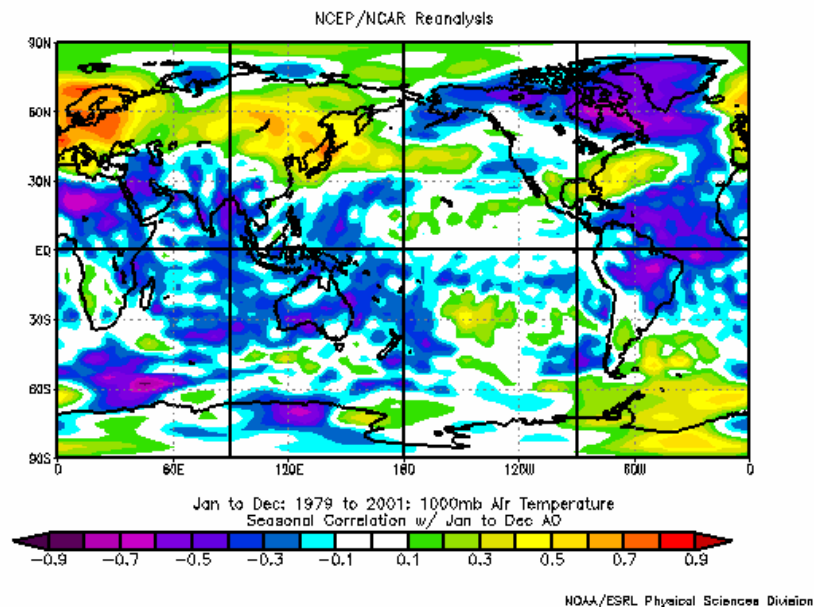
July came and I still had not heard anything, not even an acknowledgment of my emails. I had expected the editor to respond by telling me something along the lines of: this is the most boring thing ever written in the history of humanity and you are a bad person for having written it. I did not expect total silence. I sent another email on July 1st, stating that for two months I have been waiting for an acknowledgment of my proposal and I had heard nothing, so rather than wait any longer, I was sending my paper somewhere else.

I then sent it to *Theoretical and Applied Climatology*. They sent it to two reviewers. The responses were mixed, and the Editor asked me to prepare a revision that addressed the criticisms. One of the reviewers grasped the scope of the argument and wrote a brief review pointing out that the findings were important and the analysis was clear, so the paper should be published. The other reviewer got sidetracked on the general question of whether surface temperatures are affected by industrialization and decided that we had not provided a convincing proof of that point of view. What's more, the referee decided, the de Laat and Maurellis papers were not convincing, nor was my work with Michaels. There was no mention in the referee's report about the actual subject of the paper I had submitted, namely the failure of the IPCC's conjecture. Instead the referee decided that this submission would serve as a proxy for an entire literature that he disliked. While *Nature* had turned the paper down because so many others had already shown the existence of the problem, this referee recommended rejection because no evidence for the problem existed.

The referee made very approving remarks about the comment by Rasmus Benestad on the MM 2004 paper, and repeated Benestad's realclimate argument that our results were fatally undermined by spatial

autocorrelation. In a report that mostly consisted of sighs of subjective disbelief (“their studies are not convincing...” “I think the analysis is flawed...”, “I’m not convinced”) the only concrete technical objections were that we had not done cross-validation tests and we had not fixed the problem of spatial autocorrelation. The first claim was simply wrong: we had done cross-validation testing and it was written up in both papers. The second claim was reasonable. The version of the paper I had submitted did not contain a discussion of spatial autocorrelation. That material, remember, was held up because Rasmus Benestad still hadn’t sent his comment in to the *JGR*. It occurred to me at the time that the referee (who was anonymous) had a writing style rather similar in tone and phrasing to Rasmus Benestad’s. But, of course, there is no way that the referee could have been Benestad, since Benestad had already seen my unpublished notes showing that spatial autocorrelation was not a problem for the model results, and this referee was talking as if the problem existed. So just because the referee thought Benestad’s earlier writings were the last word on the subject, and just because he had Benestad’s choppy writing style and vague, disputatious way of arguing, it could not have been Benestad because the referee was writing as if he did not know that the autocorrelation issue had already been disproven.

I decided, however, that since the spatial autocorrelation material was unlikely to get into *JGR* any time soon, and since the referee had brought it up, I might as well insert a section discussing it. So in the revision I added a section providing detailed testing for spatial autocorrelation, as well as responding to all the other criticisms. The revision was submitted in October 2008. A month later the editor, Hartmut Grassl, wrote to say he was rejecting the paper because the referee said I had not addressed the problems. As it happens the referee had raised some new objections to the paper, such as claiming I had only used wintertime oscillation data and that I had not related it properly to temperature trends. Of course I hadn’t responded to these issues since they were not raised in the first round. Not that they had any merit. The only Figure in the manuscript is the following, which I copied directly from the source at the National Oceanic and Atmospheric Administration, where I obtained the data.



You can see in the original caption that the data is not wintertime-only, it is January to December. And the definition of the data, as I had explained in the paper, is the correlation between the oscillation index and the gridcell temperature, thus relating trends to trends, which is the appropriate metric.

The other objection of the referee was that I hadn't fixed the spatial autocorrelation problem. Here I realized that the referee didn't understand the technicalities. A regression model decomposes a dependent variable (\mathbf{y}) into a portion that can be explained by the independent variables (\mathbf{X}) using a set of estimated linear coefficients (\mathbf{b}), and a set of unexplained residuals (\mathbf{e}). The algebraic expression is the linear matrix equation $\mathbf{y}=\mathbf{X}\mathbf{b}+\mathbf{e}$. The significance tests are based on the ratios between the coefficients and the square roots of their estimated variances. Spatial autocorrelation can bias the variance estimates, *but only if it affects the residuals*. The formula for the variance matrix estimator \mathbf{V} is

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

If that looks confusing, rest assured that it is just a formula that take one batch of numbers and rearranges it to produce another list of numbers. If you can read this sentence you could, with a bit of explanation, understand what the formula does and why. For the present purpose, all that is required is that you notice that \mathbf{y} does not appear in it. The statistical properties of \mathbf{V} are inherited from the properties of \mathbf{e} , not \mathbf{y} , since \mathbf{y} is not in the formula. Also, in this type of model, \mathbf{X} does not contribute randomness to \mathbf{V} , it only acts as a scaling factor for the randomness in \mathbf{e} . So when statisticians and economists and scientists, and all others who know what they are doing, test models for things like spatial autocorrelation, they test the residuals \mathbf{e} , not the dependent variable \mathbf{y} .

This particular referee, however, noticed that I had tested the residuals \mathbf{e} , and he objected that I hadn't tested the dependent variable \mathbf{y} . And despite the fact that he had focused so much of his earlier comments on the autocorrelation issue, he confessed that he didn't understand the section in which I presented the standard, mundane methods for dealing with it.

I suspect most of the TAC-readership will not be able to follow the argumentation here, and I too find it hard to follow. What is the author trying to say? SAC is a problem if it means a lower degree of freedom than is apparent. Again, the discussion seems to be limited to SAC in the residual. Weighting will not resolve the problem of dependency – just give the nearby data lower weights – and again, I'm concerned about SAC in the temperature field and the socio-economic variables more than in the residuals (although the latter is also a matter of concern). Thus the specification test on pp. 8-12 is too limited to residuals and doesn't really address my concerns.

I told you this story would have its comic aspects. The referee recommended rejecting the paper, and the Editor, Hartmut Grassl, concurred.

Oh well, c'est la vie. Papers get rejected all the time, and there were other places I could send it. Getting turned down did not bother me because the referee had not found anything actually wrong with my analysis. However, I did feel that since the referee was wrong on the issues, I should write the Editor back and explain the situation and ask if he would be willing to reconsider his decision. I had received the letter rejecting my manuscript on November 5 2008. On November 7 Grassl had a letter from me explaining the problems in the referee's report. And then I waited for a reply.

And waited.

Two weeks later I wrote an email asking for confirmation that my letter had been received. Grassl's secretary confirmed that he did have it. A month later, still hearing nothing, I wrote again asking if they were considering the letter. Grassl's secretary wrote back to say she did not know why he had not responded, since he did have my letter. A month after that I still had not heard whether they were

considering the matter, and in the meantime a new paper had appeared in the literature by Rasmus Benestad's co-blogger Gavin Schmidt, repeating some of the referee's arguments against my earlier work. So I wrote a member of the editorial board and asked if he could check into where things stood. He was very apologetic that I had not received a reply and promised to look into it. But I still heard nothing after that. On April 15 2009, 5 months after sending in my response, still having heard nothing from Grassl, I re-sent him my letter and reminded him that I had heard nothing since sending it in.

I continued to hear nothing.

A week later, on April 20 2009, I emailed again saying that I had submitted my paper elsewhere and I did not want any further consideration from him and his stupid journal (or words to that effect). To this day I have never received a response from Grassl (or *BAMS* for that matter).

The outlet to which I sent my paper next was the *Journal of the American Statistical Association*. I was exasperated with the *TAC* referee's lack of understanding of basic statistics, and I decided to see what a real stats journal would say. My submission went in in April 2009 and their response came in August. As you saw earlier, the people who knew what they were talking about liked the paper and agreed that the results were solid. Rather than finding the methods confusing and hard to follow they found them too simple and mundane to merit appearing in *JASA*.

Taking up the *JASA* suggestion of a geophysical journal, in August 2009 I sent the manuscript to *Geophysical Research Letters*. On September 4, 2009, the editor of *GRL* returned the manuscript after having decided not to send it out for review. The stated reason was that

"[The] work is very narrowly focused around disputing a single sentence in the IPCC report. Indeed, you state this narrowness explicitly in the manuscript. Therefore, it is my determination that the work lacks sufficiently broad geophysical implications to meet the *GRL* criteria."

If only the IPCC had stretched their fabrications out over, say, a whole paragraph. I guess there is a policy at *GRL* against criticizing phony claims if they have been stated briefly. The fact that I was focusing my critique on only one IPCC sentence did not seem to me to make it a narrow issue, since most of the conclusions in the report depended, one way or another, on the truth of that one sentence. I wrote *GRL* a reply stating that my paper was being submitted elsewhere, adding:

The IPCC was presented with published, peer-reviewed evidence of a global bias in their surface temperature data, and the only counter-argument they offered relied on a fabricated statistical test result. The fact that they wrote with brevity while inventing non-existent test results does not diminish the necessity of correcting the record. I am taken aback by your claim that you cannot see any broader geophysical implications to this question.

At this point, to recap, I had spent 18 months submitting my paper to six journals. Three had refused to review it and one did not respond to my inquiries. Two had reviewed it, obtaining among them reports from six referees. Only one of those reviewers was negative. The reviewer had made obviously inaccurate statements, but the editor had cut off further communication so I could not respond.

On I went. I next submitted it to *Global and Planetary Change*, on September 4 2009. I had to edit the submission two weeks later because I had not included line numbers, so the review process did not begin until September 18.

The inevitable rejection came on December 2nd. The editor's cover letter read:

Dear Dr. McKitrick,

Unfortunately, we receive far more papers than we can publish. I regret, therefore, to inform you that we can not consider your paper for publication.

However, I wish you succes [sic] in preparing your manuscript for submission with another journal. And I am confident that these reviews (appended below) will be of much help during this process.

There was only one review attached, denoted "reviewer #2", which accurately summarized my argument, concurred with the results and concluded as follows:

This short paper is well written and well organized and given the clear research question and methodology, clearly deserves publication.

The only criticisms were of a minor editorial nature. Yet the journal had rejected the paper. I wrote the Editor back and asked if there were other reviews as well, but (wait for it...) to this day I have never received a reply.

However, I already knew who reviewer #1 was. Roger Pielke Sr., an emeritus professor of meteorology at Colorado State University, had written me in October to describe an unusual incident. On September 30th he received an email from the *GPC* editor asking him to review my manuscript. The email asked him to provide a review by October 30th. Roger went to the journal website where he was able to download the paper. He got an email acknowledging with thanks that he had agreed to supply a review, requesting it by October 30. But then the journal web site stopped recognizing him when he tried to sign in again to begin the review process. So on October 1st he sent an email to the *GPC* editor asking them to re-send his username and password. There was no reply to this email. On October 13th he received an email from *GPC* saying that he was being removed as a reviewer on the manuscript since he had taken too long. Roger objected that he had been unable to access the web site because they had not provided him with a working password, but he received no further response. And it made no sense because he had been given until October 30 to submit his review, but they removed him as a referee on October 13 for supposedly taking too long.

So to add to the remarkable history of this paper, I was now confronted with a journal that had solicited two reviews, blocked one reviewer before he could reply, received a positive response from the other reviewer, and then rejected the paper on the grounds that they could not publish every paper they receive.

Back to JASA

It was now clear to me that this paper was never going to be published in a climatology journal. True, I had not tried *every* possible journal, but at a certain point the pattern gets pretty clear. So I wrote to the editor of *JASA*, described what had happened at other journals, and asked if the paper might be reconsidered if I added some more complicated statistics (albeit at the risk of overkill), or whether he

could suggest an applied statistics journal. He discussed the first option with another editor but they decided the outcome would likely not change given the straightforward nature of the analysis required. However he pointed to a new journal that he and some colleagues had recently founded, called *Statistics, Politics and Policy*, which is dedicated to bringing rigorous statistical analysis to bear on important issues with policy implications. He said the paper would be a good fit, and encouraged me to submit it there. I did, and in due course my paper was accepted. It will appear in the inaugural issue this summer.

Conclusions

The paper I have talked about makes the case that the IPCC used false evidence to conceal an important problem with the surface temperature data on which most of their conclusions rest. In principle, one might argue that my analysis was wrong (though most reviewers didn't), but it would be implausible to say that the issue is unimportant or irrelevant.

Altogether I sent the paper to seven journals before it went to *SP&P*. From those seven journals I received seven reviews, of which six accepted the findings and supported publication. The one that rejected my findings contained some basic technical errors, but the journal editor would not respond to my letter pointing them out. *Nature*, *Science* and *Geophysical Research Letters* would not even review the paper, while the *Bulletin of the American Meteorological Society* never acknowledged the pre-submission inquiry. *Global and Planetary Change* received one review recommending publication, blocked another reviewer before he could submit a report and then turned the paper down.

In the aftermath of Climategate a lot of scientists working on global warming-related topics are upset that their field has apparently lost credibility with the public. The public seems to believe that climatology is beset with cliquish gatekeeping, wagon-circling, biased peer-review, faulty data and statistical incompetence. In response to these perceptions, some scientists are casting around, in op-eds and weblogs, for ideas on how to hit back at their critics. I would like to suggest that the climate science community consider instead whether the public might actually have a point.