Response of R McKitrick and N Nierenberg to *IJOC* decision to reject, Feb 18 2010.

Dear Dr. Comrie

We have received your response dated February 8, 2010 rejecting our manuscript. In your cover letter you indicated that your aim was to provide a fair and thorough review process. However, the result of the process, bearing in mind the context created by the publication of Schmidt's paper in 2009, and the fact that you did not give us the opportunity to respond to any of the referee comments, did not end up being fair to us.

The third referee supported consideration of a revised version, and raised points that could be addressed in a revision, so we will focus on the first two referees' comments.

Both the first and second referee critique our paper on grounds that would also have led to rejection of the Schmidt paper. Hence the decision to publish Schmidt but reject our paper amounts to a double standard.

The first referee rejects the whole approach of using statistical methods to test for biases in temperature data, and says our paper "does not contain anything new in mathematics or physics." But our argument does not require new derivations in mathematics or physics. The methodology is perfectly adequate for answering the questions at issue. To the extent a physical model is needed as part of the statistical framework, it is already present in the form of the GCM output as used in both our and Schmidt's paper. If this is not sufficient representation of the physical processes for the analysis we have done, it is equally insufficient to support the argument presented in Schmidt's paper.

Alternatively, if the referee is demanding a physical model of the ways in which land use change and local air pollution affects regional temperature trends, this is far beyond the scope of what we can reasonably be expected to provide. There is a longstanding practice of using even simpler models than ours in applied climatology, such as a regression against local population levels to estimate the required UHI adjustment in weather station data, or using statistical comparisons of urban and rural stations to estimate the urbanization component. In such cases population data are "translated into physical factors" via the estimation of model parameters, which is exactly what happens in our model as well. If estimation of empirical parameters is as unacceptable as the referee implies then every paper using empirical methods, and every paper using data processed by such methods, would have to be rejected. That would rule out countless papers, including all the ones cited by the IPCC (2007) in its defence of the CRU land data and many empirical papers published in the *IJOC* over the years.

The first referee also said our model does not have a "either a sound mathematical or physical background." This is untrue. Our model has a sound mathematical background: it is a properly-specified statistical model with well-known asymptotic distributional properties. The referee says it is "impractical" to use our model to "explain the complicated problem of climate data errors and model uncertainties." This statement is unfair given the scope of our analysis. We are testing an argument that was previously published in the *IJOC*, which presupposed that it is practical to use such models for the purpose. Contrary to the referee's assertion, it is quite practical to use this type of model to examine the issue, test hypotheses and form arguments. We are well within the established methodological bounds of this literature.

The referee says we should have referred to other studies in our introduction, but the list he or she provides is largely irrelevant. The Menne (2009) paper came out after our paper was submitted, and only applies to the US record, which, in the McKitrick and Michaels 2007 paper, we argued had relatively little bias compared to other regions anyway. The Jones et al. (2007) paper does not describe the specific homogeneity adjustments or test for their adequacy so it does not relate to the referee's argument. Reference to the IPCC (2007) report sections on model errors and attribution methodology (i.e. Chapters 8 and 9 of WG1) is both vague and off-topic. Even less relevant is the claim that we should have discussed the Woodward and Gray paper. They used time series methods to estimate ARMA processes in global temperature time series. Our paper uses spatial regression on a cross-sectional data base. ARMA methods have absolutely no relevance to our data set. The fact that the referee did not know this calls into question how thorough his or her reading of our manuscript was.

Overall, Referee #1 does not address, critique, refute or even acknowledge any of our arguments or findings. Instead he or she rejects the whole premise of statistical modeling of the underlying phenomena. The referee acknowledges that "The socialeconomic factors are important in the climate change studies" but insists that they "should be translated into physical factors to explain the physical consequences and implications." By this criterion many years' worth of published empirical papers would have to be rejected. Moreover the claim is extreme and implausible. Statistical modeling is entirely appropriate for testing hypotheses that have been presented in the literature in statistical form. On the other hand, if the *IJOC* now considers it to be inappropriate to publish statistical tests of empirical hypotheses, then it was equally inappropriate to have published the hypotheses in the first place. To change the criterion midstream is unfair and unwarranted.

The second referee appears to accept our findings regarding his or her listed points (2) and (3), thus refuting two key arguments in Schmidt's paper, but then changes the subject by deciding that the "argument has drifted too far from its main point." This is false: our argument is focused exactly on the main points raised by the publication of Schmidt's paper. If the points at issue are too tangential to merit publication, then so was the Schmidt paper. It is unfair, however, to publish an argument and then declare that the refutation of the argument should not be published because the topic is irrelevant.

In any case, the "main" point, according to the referee, is actually one that we address: "Are there any significant biases in the 1979-2002 observed land temperature trends, and, if so, what are they?" The referee compliments our analysis on this as "interesting and imaginative" but then resorts to the vague objection that our regression is overfit. The referee says: "to carry conviction the conclusions from it need to be validated on independent data. This paper rejects the idea of considering other datasets…" But this is false. MM07 did 500 cross-validation tests on randomly-withheld (independent) subsamples precisely to rule out overfitting. And as we discussed in our paper (page 21 lines 25-34) we re-did those tests in the new submission. Had the referee read the paper thoroughly he or she would not have accused us of refusing to do such tests. Also we re-did the analyses using three surface data sets and two satellite data sets, so the claim we refuse to use "other" data sets is untrue. What we "rejected" was doing the analysis on *sea surface data*, for the obvious reason that all economic development occurs on land. By claiming this is "disingenuous" of us, and saying we should use "other datasets" which, in the context, can only mean sea surface temperature, the referee is asking us to undertake the absurd, contradictory and impossible task of modeling land-use change on the open ocean.

Referee 2 writes:

> On page 4 it says: "the null hypothesis of no-socio-economic effects rejects at P=1.81*10-6 ...". The problem with this statement is that it depends on a lot of unstated (and often untestable) assumptions. To make a quantitative statement about real world probabilities based on a statistical model requires that the model chosen be correct - that it contains all the necessary regressors and no extraneous ones. It also requires a precise and quantitative probability model

> for the residuals, which has to be decided upon a priori. Using the MMO4 regression model and the null hypothesis of no-socio-economic effects, I would expect to see significant autocorrelation in the residuals, but coming up with an a-priori probability model for them is not easy. As there is a large uncertainty in the expected nature of the model residuals given the null hypothesis, it is impossible to calculate the probability of rejection with any precision.

Nothing in this paragraph disputes our work, it only amounts to saying that statistical modeling must be done carefully. We provided numerous specification tests that address all the specific challenges raised by the referee, including three different spatial models of the residuals to deal with the expected autocorrelation, yet the referee ignores them all. What the referee appears to be arguing is that the modeling challenges are so complex that *in principle* it is impossible to address them. If that is so, then, once again, the criterion would rule out publication of a long list of papers, including Schmidt's. We have actually done far more than other authors to address specification issues, so to apply such a nihilistic criterion only to our work is doubly unfair.

The referee's next paragraph begins

> I agree that GS09 confuses autocorrelation in the regressors and the residuals, but MN09 does not discuss the problem with much more sophistication: the assertion that a model with no SAC in the residuals is better than one with SAC is not defensible.

The first point alone justifies publication of a comment on Schmidt since the confusion about where the SAC correction needs to be applied goes to the heart of Schmidt's argument. The claim that we do not discuss the problem with "much more sophistication" is simply untrue. We provide a detailed, state-of-the-art model of the SAC problem and we provide ample testing support for our model selection. To suggest that disappearance of SAC from the residuals is not evidence of an improved specification is baffling.

Overall the referee's conclusion is that issues were badly handled in the original papers (including Schmidt's) but we do not resolve them successfully. But on every specific issue he or she raises, the referee apparently concedes our findings and does not provide a single technical objection to our methodology. He or she either ignores the specification testing we did do and then criticizes us as if we had not done it, or resorts to the position that the issues are so complicated it is impossible to deal with them, and concedes that the Schmidt paper was equally faulty. Either way, the decision to reject our submission, having previously published Schmidt's, is unfair.

As mentioned, the third referee is open to a resubmission so we do not need to go into detail on that report for the purpose of this letter. However it is worth pointing out that his or her main objection is the use of GCM runs and observational data in the same regression model. Once again, this is the framework Schmidt presented. If it is faulty in our context then it should not have been used in Schmidt's paper. That said, the referee is overstating the problem.

With regards to your other criterion of thoroughness in the review process, it is noteworthy that none of the referees acknowledged any of our quantitative conclusions, and indeed referee #2 claims we did not do certain tests that we did do, and which are discussed in our paper. None of them took note of the fact that Schmidt's coefficients fall outside the confidence intervals he claimed they needed to fall within to validate his argument, and in many cases even took the wrong sign (see our Table 5). None of them took note of the fact that Schmidt's argument rested on claims about SAC that he himself did not test for, and which conventional tests show affects Schmidt's results but not ours. And both seem unaware, or unconcerned, that they are recommending rejection of our paper on grounds that were waived when considering Schmidt's paper.

In sum, based on these referee reports we are unable to see how the process could be considered thorough or fair to us. We request that this submission be re-considered and that we be given the opportunity to provide a revision and response to the referee comments. There are prima facie grounds from the referee comments for concluding that we have at least established the case for publishing a comment on Schmidt's paper, if that is a preferred format for proceeding.

Yours truly,

Ross McKitrick and Nicolas Nierenberg.