INTERNATIONAL JOURNAL OF CLIMATOLOGY
*Int. J. Climatol.* **29**: 2041–2048 (2009)
Published online 7 January 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1831



# Spurious correlations between recent warming and indices of local economic activity†

Gavin A. Schmidt*
*NASA Goddard Institute for Space Studies and Center for Climate Systems Research, Columbia University, New York, NY, USA*

**ABSTRACT:** A series of climate model simulations of the 20th Century are analysed to investigate a number of published correlations between indices of local economic activity and recent global warming. These correlations have been used to support a hypothesis that the observed surface warming record has been contaminated in some way and thus overestimates true global warming. However, the basis of the results are correlations over a very restricted set of locations (predominantly western Europe, Japan and the USA) which project strongly onto naturally occurring patterns of climate variability, or are with fields with significant amounts of spatial auto-correlation. Across model simulations, the correlations vary widely due to the chaotic weather component in any short-term record. The reported correlations do not fall outside the simulated distribution, and 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. Published in 2009 by John Wiley & Sons, Ltd.

KEY WORDS global warming; economic activity; spurious regression

*Received 16 May 2008; Revised 2 September 2008; Accepted 8 November 2008*

## 1. Introduction

Two recent papers (de Laat and Maurellis, 2006; McKitrick and Michaels, 2007) (henceforth dLM06 and MM07) have independently asserted that records of surface warming are likely contaminated due to 'anthropogenic surface processes'. This could be interpreted as either asserting that the 'contamination' is an artifact and the surface records, therefore, give a misleading picture of 'true' global warming, or that unaccounted-for processes need to be incorporated into climate model hindcasts. These two studies looked at correlations over the 23-year period (1979–2001) (in both cases) of local warming with anthropogenic $CO_2$ emissions in the dLM06 case and with satellite data and various econometric measures (Gross Domestic Product, literacy and educational levels, population etc.) in the MM07 case. While the specific analyses differ in the two papers, the net effect of the choices made is to correlate very restricted spatial patterns of economic activity with the pattern of warming.

There are a number of reasons why these analyses are unlikely to have come to the correct conclusions (that either an unaccounted-for process is significant or that the data are contaminated). Firstly, there is significant independent evidence for warming in the oceans, snow cover, sea ice extent changes, phenological records etc. which are consistent with the land station analyses (Bindoff *et al.*, 2007; Lemke *et al.*, 2007; Trenberth *et al.*, 2007). Secondly, alternative hypotheses, that the correlations are related to patterns of climate variability, or related to known local forcing agents (such as tropospheric ozone, black carbon etc.) were not considered. Finally, the proposed processes to explain these correlations appear inadequate. In dLM06, the so-called contamination is also detected in satellite records, ruling out small-scale local problems with the measuring network of surface stations (urban heat island effects or micro-site issues). Other explanations of their results include potentially real features of the climate system, but are either already included in climate models to some extent (land use change), or are generally considered to be negligible on the global scale (direct waste heating). It is certainly possible that these last features are not correctly specified, or are more significant than generally thought, however, before any attribution to a cause can be made, we have to assess whether the correlations actually have any statistical power to detect an anomaly.

The climate system has a great deal of unforced weather 'noise' that has significant decadal variations and complex spatial structure which is uncorrelated with any external climate driver. This leads to the well known phenomenon that the variability of trends over specific regions is a strong function of their spatial extent. The smaller the area selected, the greater the spread in the observed trends. For instance, the individual grid box trends in the HadCRUT3v dataset over the period 1979–2001, range from −0.6 to 1.2 °C/decade compared

* Correspondence to: Gavin A. Schmidt, NASA Goddard Institute for Space Studies and Center for Climate Systems Research, Columbia University, New York, NY, USA. E-mail: gschmidt@giss.nasa.gov

to the trend of $0.17 \pm 0.01\,^{\circ}$C/decade in the global mean (Brohan *et al.*, 2006). There is a similar variance in trends at any one grid box over the same period seen in individual simulations in climate models ensembles. For instance, in an ensemble of 20th Century simulations with the Goddard Institute of Space Studies (GISS) ModelE-R at the grid point centered on $37.5\,^{\circ}$E, $50\,^{\circ}$N (in Eurasia, picked at random), trends go from $-0.17$ to $0.5\,^{\circ}$C/decade in five ensemble members with identical forcing over the same 1979–2001 period (Hansen *et al.*, 2007).

This spread across space in the observations or any specific simulation and across simulations, implies that there is a spatial scale below which attribution of short-period trends to specific forcings is impossible (since any forced signal is drowned out by the unforced 'noise'). That scale will depend on the duration of the record and the strength of the forcing. For the late 20th Century, attribution of trends is possible at the continental level, but not generally at smaller scales (IDAG, 2005; Zhang *et al.*, 2006). Thus, it might be expected that correlations based on regional patterns that are less than continental in extent will exhibit a great deal of variability and will not be particularly useful in detecting and attributing climate change nor in detecting potential problems in specific datasets.

Whether the reported correlations in dLM06 and MM07 can be used to determine if there is an unaccounted-for contamination of the surface data therefore depends on how significant those correlations really are given the magnitude of the unforced variability in the system. Note that this is a completely separate source of 'noise' than the uncertainty in actual measurements or in the fitting of a linear trend and, in fact, is much larger than either.

There is a relatively easy way to assess whether there is any true significance to these correlations. We can take fully consistent model simulations for the same period and calculate the distribution of the analogous correlations. Those simulations contain no unaccounted-for processes (by definition!) but plenty of internal variability, locally important forcings and spatial correlation. If the distribution encompasses the observed correlations, then the null hypothesis (that there is no contamination) cannot be rejected. This calculation does not imply that the modelled distribution is a complete simulation of the same distribution in the real world. However, since the real world is likely to be more variable (due to unresolved sub-grid-scale processes in the models, or more complex forcings than are accounted for), showing that any observed correlation is truly reflecting something outside the modelled framework requires, at minimum, that it be outside the existing models' range.

There are a number of suitable model simulations that can be assessed. First, we use a set of simulations using fully coupled models driven by 20th Century forcings. These runs have consistent changes in the sea surface temperature and land temperature and capture the long-term trends well, but because of their dynamic ocean component the internal variability will be uncorrelated with that in the real world (Hansen *et al.*, 2007). For instance, El Niño-like variability is a large part of the internal variability of the system and over a short time-period (as considered here), the random nature of its occurrence, or its lack of fidelity in amplitude and phasing might compromise any comparison with the real world. Thus, a further set of simulations are those forced with ocean sea surface temperatures and with the external 20th Century forcings (AMIP-style runs, Gates *et al.*, 1999). These runs have correctly timed El Niño events and volcanoes as in the real world and similar large-scale long-term trends. However, they are not fully consistent with the forcings and may show biases in ocean/land contrasts.

Analyses were made of various simulations using GISS ModelE archive for the 20th Century, using estimates of all appropriate forcings (see Schmidt *et al.* (2006) and Hansen *et al.* (2007) for details). Specifically, 4 ensemble members from an AMIP-style simulation including all forcings (AMIPf8) were used, and 5 ensemble members of the coupled atmosphere–ocean model GISS-ER 20th Century simulations. The forcings in each case are all the same, but the specific sequence of weather and internal variability is uncorrelated between the runs. The 5 independent members of the coupled model ensemble each start from initial conditions 20 years apart from their pre-industrial control run so that the spread in response due to uncertain ocean initial conditions can be estimated.

Whether the model simulations show qualitatively similar correlations to those reported for the real world will allows us to assess whether the observed correlations significantly detect a real pattern. If they do not, then that would imply that some process (potentially a contamination of the surface record) that was not included in these models could be responsible. However, if these patterns do show up, we need to determine whether this is related to internal variability or whether they arise in response to various forcings. If we can show either, then we may be in a position to attribute the observed correlations.

## 2. de Laat and Maurellis (2006)

The analysis method in dLM06 is to split a particular temperature dataset into two classes according to a local threshold based on 1990 EDGAR2 anthropogenic $CO_2$ emissions (in kgC/m/yr) (van Aardenne *et al.*, 2001) (building off an earlier paper de Laat and Maurellis, 2004). They used both the HadCRUT2 temperature data (Jones and Moberg, 2003) and a satellite-derived estimate for the lower troposphere (UAH version 5.0) (Christy *et al.*, 2003). Here, I will just use the more up-to-date HadCRUT3v product (Brohan *et al.*, 2006). It is noted that the EDGAR2 1990 emissions estimates have not changed appreciably in later versions and so the same version for consistency will be used. The emissions used in dLM06 did not include emissions from aircraft, international shipping, or biomass burning, thereby restricting the emissions to industrial emissions from land areas (Jos de Laat, personal communication).

Neither the update to HadCRUT3v nor the exclusion of some sources of $CO_2$ emission make any appreciable difference to the results.

For each class of points, the area-weighted mean temperatures are calculated and the trend estimated using ordinary least-squares regression. This is denoted ‘method 2’ in dLM06 and is what was used in that paper despite the text mistakenly indicating that ‘method 1’ (the area-weighted mean of the trends) was used (Jos de Laat, personal communication). There is only a minor difference in the resulting trends (and that depends solely on missing values), but in the ‘method 2’ case, the uncertainties are easier to calculate. The distinction is irrelevant to the calculations that follow. The trend for all points above and below the threshold are plotted as a function of increasing threshold. The error bars are the $1\sigma$ uncertainty in the trend estimate and do not include uncertainties in the underlying dataset, nor do they correct for possible auto-correlation and so underestimate the true uncertainty (Santer *et al.*, 2000).

In Figure 1, is reproduced the key calculation in dLM06 for the observed temperature data (1979–2001). Their result was that there is a correlation between higher trends and local industrial activity and a similar phenomenon is seen here (the uptick in the black curve). Also plotted is the same calculation for each of the 4 AMIPf8 ensemble members for the same time-period along with the minimum and maximum spread in the error bars associated with determining a linear trend. The results for the coupled model ensemble fall within the envelope of the AMIP results and so are not shown for clarity. As can be seen, there is a large spread in the trends as the $CO_2$ emission threshold increases, all of which is attributable to internal atmospheric variability. One ensemble member coincidently reproduces almost exactly the observed situation. The standard deviation of the right-most calculations is greater than 0.1 °C/decade.

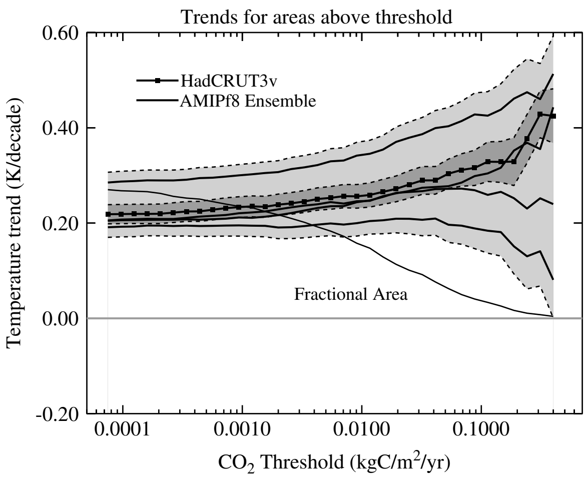


Figure 1. Correlations of temperature trends with $CO_2$ emissions (as a proxy for anthropogenic activity) as a function of the emissions threshold. The observed data are in black while the identical analysis for 4 AMIP-style runs using GISS ModelE are shown in blue. The shading denotes the $1\sigma$ uncertainty in the least-squares regression.

Thus, a two-sigma uncertainty on the high threshold trends is at least 0.2 °C/decade, placing the observed values well within the expected distribution.

This is a very different comparison to that shown in dLM06 where they showed results from two models each with a single simulation, neither of which showed any influence from the $CO_2$ emissions thresholding. The explanation is however clear. The model analysis in dLM06 used 100 years of data from each model (model years 2000–2099) and the simulations had a very strong 1% increasing $CO_2$ forcing (Jos de Laat, personal communication). Both factors serve to reduce the impact of any internal variability on the analysis, as could be deduced from the much smaller error bars on the trends for the models compared to the observations. Those previous comparisons were not, therefore, useful in assessing the role of internal variability in contributing to the dLM06 result on much shorter timescales and under milder forcing conditions.

### 2.1. Sensitivity to time-period

We can use the emissions in 1990 as a reasonable estimate for the pattern of industrial activity since at least the 1940s, thus we can also test the importance of time-period and start date within the observational record and in the simulations. If the surface data are contaminated then one would expect a consistent pattern through time, whereas if the correlation is spurious, one would expect a spread of results. Using a start date every 10 years from 1930 onwards, a calculation of the above- and below-threshold trends for each subsequent 23-year period (the same length as for the 1979–2001 period considered above) was made. Since the mean global trend varies considerably over each sub-period, the results are more clearly shown as the difference between the above- and below-threshold trends.

Figure 2 shows that the spread of results is indeed large, in striking support of the internal variability hypothesis. The most striking example is for the 1950–1972 case where the trends are significantly more negative in the high $CO_2$ threshold points. If the cause of the higher trends for the high threshold points in the more recent time-period was caused by urbanization or local land use change, there is little reason to expect a switch in sign from the 1950s to the present, and so these potential causes are inconsistent with this result. This might also have been deduced from the result reported in dLM06, that the thresholding of the satellite data showed a similar pattern.

## 3. McKitrick and Michaels (2007)

The approach of MM07 is slightly more direct. They simply looked for correlations between land surface trends and a rather eclectic mix of climate and econometric variables. More specifically, they assumed that the surface trends can be modelled as a linear function of the Microwave Sounding Unit (MSU) lower

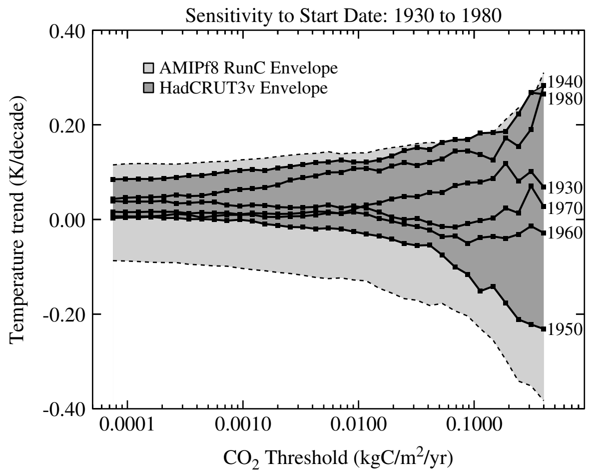


Figure 2. Correlations of 23-year temperature trends with $CO_2$ emissions using the HadCRUTv3 using different starting dates. The difference between the above-threshold trend and the below-threshold trend is plotted for clarity. The equivalent spread from a single realization from the model is shown in the background for comparison.

tropospheric (LT) trends ('trop') plus additional variables. The MSU-LT records came from the University of Alabama Huntsville (UAH) analysis (Spencer and Christy, 1990) [updated to version 5.2]. Alternative versions of the same product exist from Remote Sensing Services (www.remss.com version 3) which has larger global long-term trends than the UAH version, the differences due primarily to decisions made in tying together different satellite records (Mears and Wentz, 2005). The basic assumption in MM07 is that the satellite trends are a more robust measure of the true surface warming, and that economic correlations to the residuals should be a fingerprint of potential contamination.

There are a number of problems with this basic hypothesis. First, the lower troposphere (centred ∼4 km above the surface) is physically expected to have more spatially homogeneous temperature trends than the surface due to the greater mixing of lower tropospheric air masses compared to the surface layers. Thus, even in a perfect situation there will be significant residuals that could well be correlated to the pattern of human development since that is not randomly distributed. The attempts in MM07 to account for that appear rather simplistic.

More importantly however, as previous criticism (Benestad, 2004) of an earlier version (McKitrick and Michaels, 2004) of MM07 pointed out, the significance of the correlations is likely over-estimated since spatial correlations were not taken into account. Adjacent grid boxes in both economic and climate data are not independent (Jones *et al*., 1997), and assuming that they are leads to over-estimating the significance of any correlation and the potential for over-fitting any statistical model. Some indication of this is given by the fact that the largest 'contamination' deduced from their methodology are in very remote polar regions such as Svalbard or the South Orkneys, hardly sites of significant industry (Rasmus Benestad, pers. communication).

The same economic indicators as MM07 (see that publication for sources and full definitions) have been used. The most important of these (according to the results in MM07) are 'g' (1979 Gross Domestic Product (GDP) divided by land area), 'e' educational attainment (literacy levels and secondary schooling), 'x' missing months in the temperature records, 'p' population growth (1979–1999), 'm' income growth (1979–1999), 'y' GDP growth (1979–1999), 'c' coal consumption growth (1980–2000).

For simplicity, only two tests used in MM07 – their 'SURF' multiple-regression model, where they include 'trop', other physical variables and 'g', 'e', 'x', 'p' 'm' 'y' and 'c' as dependent variables, and 'G3', which uses only 'g', 'e', and 'c' among the economic variables are reproduced. These are chosen because these are the ones used for the claims of high nominal significance ($P < 1 \times 10^{-14}$) and for their subsequent 'correction' of the surface temperature data. Note that the use of these variables and methodology is only so that it is possible to examine the fragility of the previously published correlations. No support for the original experimental design is implied and none should be inferred.

### 3.1. Sensitivity to observational records

Reproduced first are the results from Table II in MM07 for tests SURF and G3 (Table I in this paper). The emulation, using the multiple-linear regression routines from Legendre (2002) gives exactly the same regressions, and while the tests for significance are different (two-tailed tests with 10 000 Monte Carlo permutations of the residuals), there is a one-to-one match with the nominally significant terms as derived by MM07. This demonstrates that there is no effective difference in the statistical procedures used.

Second, the same calculation with updated temperatures (CRUTEMv3) and with the RSS version of the MSU-LT data are performed. As would be expected from the greater global trend in the RSS product, the regression coefficients to 'trop' increase when compared to the UAH product. More interestingly, the significance of correlations to population, income and GDP growth disappear, pointing clearly to the fragility of these relationships. This loss of significance is related to the switch of satellite data and should raise concerns that the original significance was overstated, possibly because of the spatial correlation issue mentioned above. One other minor difference is that the correlation to the absolute latitude becomes significant. The economic indices 'g', 'e' and 'c' do, however, remain nominally significant (under the MM07 assumptions).

Performed next are these same regressions using surface temperature and emulated MSU-LT data from five realizations, and the ensemble mean from 20th Century coupled simulations, again using data from 1979 to 2001. When using output from model simulations, the 'x' variable is not used (since all records within the models are complete). The results show that the only consistently significant regression across all simulations is with

Table I. Regression coefficients for the fit between the surface temperature record, the MSU-LT record and various climate and economic indices as defined by MM07 as follows: 'trop' is the MSU-LT temperature trend, 'slp' is mean climatic sea level pressure, 'dry' is a dummy variable indicating a mean dewpoint below 0 °C, 'dslp' is dry' times 'slp', 'water' is a flag for a coastline in the grid cell, and 'abslat' is the absolute latitude. Economic variables are described in the text. Bold denotes nominal significance as defined in MM07 (without adjusting for spatial correlation) (compare with MM07 Table I).

| Variable | CRUv2 + UAH MSU | | CRUv3 + RSS MSU | |
|---|---|---|---|---|
| | SURF | G3 | SURF | G3 |
| Intercept | −4.20808 | −3.78891 | −2.22913 | −1.11246 |
| Trop | **0.86307** | **0.88841** | **0.98721** | **1.00240** |
| Slp | 0.00442 | 0.00408 | 0.00239 | 0.00122 |
| Dry | 0.57042 | 1.58466 | 0.19861 | −0.29344 |
| Dslp | −0.00046 | −0.00146 | −0.00013 | 0.00036 |
| Water | −0.02892 | −0.02448 | −0.01688 | −0.01214 |
| Abslat | 0.00062 | −0.00030 | **0.00281** | **0.00274** |
| g | **0.04316** | **0.04803** | **0.04445** | **0.05510** |
| e | **−0.00269** | **−0.00280** | **−0.00252** | **−0.00212** |
| x | 0.00412 | 0.00288 | −0.00035 | −0.00032 |
| p | **0.38391** | | 0.15135 | |
| m | **0.40932** | | 0.26628 | |
| y | **−0.30475** | | −0.21597 | |
| c | **0.00618** | | **0.00762** | |
| n | 440 | 440 | 434 | 434 |
| r | 0.53 | 0.51 | 0.52 | 0.51 |

the MSU-LT values (Table II). Although each simulation has some significant regression to one or more economic indicators, no economic indicator is significant for all the simulations, and significant correlations exist for the economic variables even with the ensemble mean. Correlation coefficients (and regressions between 'surf' and 'trop') are uniformly higher as would be expected in the absence of any observational uncertainty or sub-grid-scale variability in either the surface temperatures or MSU-LT trends. One point possibly worth noting is that the correlations with 'g' and 'e' are of the opposite sign to those in the regressions with the observed data.

In the G3 test, the correlations in the model to the economic variables 'g' and 'e' is even more striking (Table III). Three out of five have nominally significant regressions to 'e', and two out of five runs have significant regressions to both 'g' and 'e', as does the ensemble mean (as with the observed data). The $t$-test values are of comparable size (though slightly smaller) to those seen in the original tests.

Also performed were the same calculations with the AMIPf8 runs and their ensemble mean (not shown) and it was found that 'e', 'p' and 'g' were each significant in the SURF experiment in one specific run, but no other. In the G3 experiment, one run shows that 'g' is significant, and one run has 'e' significant, each with a positive regression.

Since there is no economic or other contamination in the models, the preponderance of nominally significant correlations certainly implies that the reported $F$-test values are not a fair assessment of the hypothesis put forward by MM07. We find that supposedly 95% significant correlations to 'g' and 'e' (in experiment G3) occur in 3 and 4 (respectively) simulations out of 9, roughly 7 times as often as should be expected if the 'significance' test used by MM07 had even its minimum reported power. This clearly demonstrates that there are far fewer degrees of freedom in these correlations than they assumed.

## 4. Discussion

The two sets of tests used here are very different in nature, and yet within the consistent model context, our results are similar to those seen in the observations.

Table II. As in Table I, column 1, but with output from GISS-ER (Schmidt *et al*., 2006).

| Variable | SURF test (GISS-ER) | | | | | |
|---|---|---|---|---|---|---|
| | RunA | RunB | RunC | RunD | RunE | Ensemble |
| Intercept | 2.26641 | 2.05760 | 2.13344 | **6.85362** | **−4.33633** | 1.11121 |
| Trop | **1.44023** | **1.43273** | **1.32420** | **1.48376** | **1.12618** | **1.22160** |
| Slp | −0.00222 | −0.00222 | −0.00222 | **−0.00694** | **0.00417** | −0.00118 |
| Dry | −1.57550 | 0.05045 | 0.17500 | 0.93299 | 3.79676 | 0.82598 |
| Dslp | 0.00153 | −0.00006 | −0.00015 | −0.00099 | −0.00375 | −0.00083 |
| Water | −0.02033 | **−0.03532** | **−0.04576** | **−0.03065** | **−0.04219** | **−0.03610** |
| Abslat | −0.00005 | **0.00128** | **−0.00078** | **0.00195** | **0.00115** | **0.00074** |
| g | −0.01171 | −0.00208 | −0.00330 | **−0.01793** | −0.01107 | −0.00763 |
| e | −0.00030 | **0.00057** | **0.00087** | **0.00076** | **0.00061** | **0.00036** |
| p | **−0.17525** | **−0.09212** | −0.05862 | −0.06243 | 0.03953 | **−0.06663** |
| m | −0.15285 | **−0.16996** | −0.05782 | −0.03183 | −0.02871 | **−0.08630** |
| y | 0.08856 | **0.13842** | 0.05174 | 0.03535 | −0.00528 | 0.05728 |
| c | 0.00060 | −0.00114 | −0.00118 | 0.00067 | 0.00188 | 0.00032 |
| R | 0.69 | 0.79 | 0.70 | 0.73 | 0.65 | 0.68 |

Table III. As in Table I, column 2, but with output from GISS-ER.

| Variable | G3 test (GISS-ER) | | | | | |
|---|---|---|---|---|---|---|
| | RunA | RunB | RunC | RunD | RunE | Ensemble |
| Intercept | 3.33518 | 2.24250 | 1.58085 | **6.33386** | −2.62692 | **1.85955** |
| Trop | **1.37771** | **1.41803** | **1.31042** | **1.46890** | **1.08848** | **1.22844** |
| Slp | −0.00336 | −0.00234 | −0.00167 | **−0.00643** | 0.00249 | **−0.00193** |
| Dry | −3.04618 | −0.29150 | 0.41295 | 1.09698 | 2.49981 | 0.03369 |
| Dslp | 0.00297 | 0.00028 | −0.00039 | −0.00114 | −0.00248 | −0.00005 |
| Water | **−0.02771** | **−0.03490** | **−0.04526** | **−0.02716** | **−0.04670** | **−0.03757** |
| Abslat | 0.00054 | **0.00097** | **−0.00074** | **0.00192** | **0.00099** | **0.00080** |
| g | −0.01335 | −0.01053 | −0.00294 | **−0.01463** | **−0.01814** | **−0.01119** |
| e | 0.00009 | 0.00031 | **0.00077** | **0.00072** | **0.00072** | **0.00043** |
| r | 0.67 | 0.79 | 0.69 | 0.72 | 0.62 | 0.67 |

Since there is no extraneous contamination possible in the model runs, and since the results are not consistent across all runs (which could be indicative of a forced response to perhaps, aerosol, ozone forcing or land use change, all of which are represented (albeit imperfectly) within these simulations), the only remaining conclusion is that these correlations are due to intrinsic 'weather' variability within the models.

It is therefore necessary to look a little deeper into what the source of these correlations are. In both the above cases, the calculations boil down to spatial correlations on extremely restricted domains or with smooth fields. The general problem is associated with the concept of 'field significance' (Livezey and Chen, 1983), i.e. how to determine whether a spatially correlated field is significantly related to a time series or other field when the number of separate point-to-point correlations is much larger than the true number of the degrees of freedom. Monte Carlo tests with suitably generated data that conforms to the null hypothesis are a standard approach. However, in this case it is not at all obvious how to synthetically generate such data in a meaningful way.

Figure 3(a) shows the spatial extent of the grid points used in the right-most few points in Figure 1. It should be clear that their areal extent is small (~1% of the globe, 3% of the land area, for the threshold above 0.2 kgC/m/yr) and very specific to the Northern Hemisphere mid-latitudes (a few boxes each from California, northeastern USA, central Europe, eastern China, Japan, northern India). For this area, given the distribution of trends in individual grid boxes and the spatial correlation among them, one would expect the range of trends to span 0.3 to 0.4 °C/decade (assuming roughly 10 degrees of freedom), as indeed they do (Figure 1).

Similarly, the structure of the fields that appear significant in MM07 ('g' and 'e', Figure 3(b) and (c)) are either very focused or have a very low number of degrees of freedom, certainly not the 400+ assumed in the significance tests used by MM07 (and above). The temperature fields of 'surf' and 'trop' (updated to CRUTEM3v and RSS MSU-LT, Figure 3(d) and (e)) also have significant structure. Their difference (scaled as suggested by the regression in column 4 of Table I) is shown in Figure 3(f) and appears to have more degrees of freedom though still not the 400+ assumed.

However, it may be possible to estimate the true number of degrees of freedom in these fields (while noting that this is a rather ill-posed question). Figure 4 shows the normalized semi-variogram for the key data used in MM07. For each radius of influence ($r$ on the x-axis, measured in great circle radians), is calculated the average $\mathrm{E}[(\mathrm{x_{ij}} - \mathrm{x_{kl}})^2]/2$ (including areal weighting) for each point $(k,l)$ within a distance $r$ of point $(i,j)$ (normalized so that variance of the whole field is 1). For fields with spatial correlation, this value will be significantly less than the total field variance for a small radius of influence. At $r = 2\pi$, the area encompasses the whole globe, and the semi-variogram is equal to the total variance. The grey lines correspond to the same calculation but for synthetic data at the same resolution. We start off with random uncorrelated data at each grid point (drawn from a Gaussian distribution with mean 0, $\sigma = 1$). We then progressively smooth this data spatially with an increasing radius of influence. We estimate the degrees of freedom (dof) in this smoothed data as the number of times the smoothing area fits on the globe ($= 2/(1 - \cos(r))$ where (as above) $r$ is the great circle angle in radians). For a smoothing radius of about 2000 km ($r = \pi/10$), there are about 41 dof.

There is some variability in the semi-variograms for different realizations of the random field, but the results in Figure 4 are typical. By comparing the actual data to the synthetic data, it can be seen that the 'surf' and 'trop' variables have patterns that are similar to fields with around 20 to perhaps 60 dof. This is consistent with recent estimates for the surface temperature patterns which suggest around 50 dof on the annual timescale (and less for multi-decade trends) (Wang and Shen, 1999; Brohan *et al.*, 2006). For the economic variables, 'g' has perhaps 100 dof, but a somewhat abnormal structure, while 'e' has more like 15. There is, of course, some uncertainty here, but the point to be

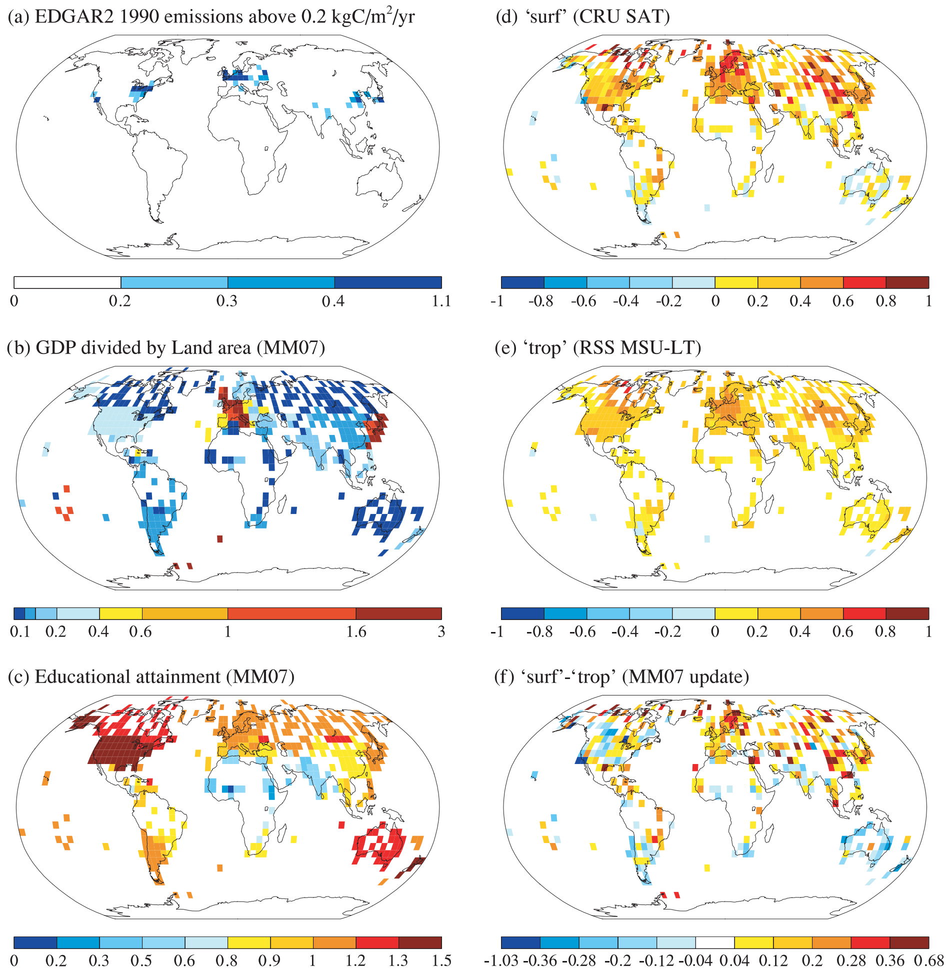


Figure 3. The spatial patterns of (a) the high $CO_2$ emission grid boxes, (b) the GDP divided by land area ('g') and (c) Educational attainment ('e'), (d) the 'surf' temperature trends from CRU, (e) the 'trop' MSU-LT trends from RSS, and (f) 'surf'-'trop' temperature differences. This figure is available in colour online at www.interscience.wiley.com/ijoc

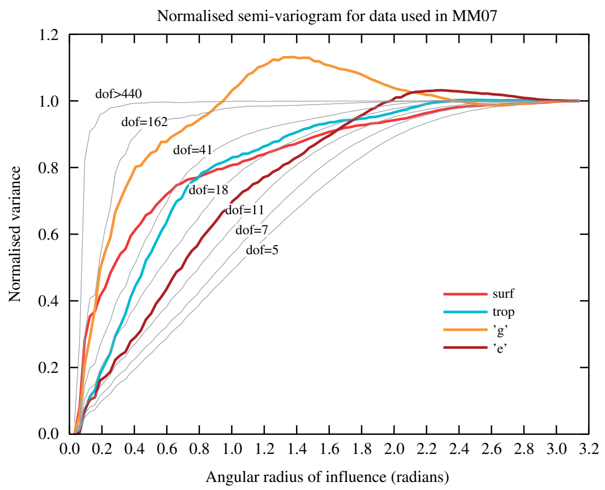


Figure 4. The semi-variogram for the data used in MM07 compared to a example of random synthetic data on the same grid (grey lines). The synthetic data range from a completely uncorrelated field, to fields that have been progressively smoothed. As the smoothing increases, the number of effective degrees of freedom decreases, calculated as the number of independent areas that will fit over the globe at that level of smoothing. The effective spatial degrees of freedom in the MM07 data can be estimated from the correspondence with the synthetic data. This figure is available in colour online at www.interscience.wiley.com/ijoc

stressed is that there is nothing like as many as 440 dof as assumed in the nominal significance calculations used in MM07. If we sub-sample the data so that the number of samples is closer to the dof indicated above, the significance of correlations to 'g' and 'e' become marginal or disappear depending on the sub-sample. Notably, the correlation to 'g' is very fragile disappearing completely even with a sub-sampling of 1 in 4 points.

Thus in both cases examined here, it is apparent that the true significance of the analyses is much less than previously claimed, and that the previous conclusions that there must be a large bias to the surface temperature record, are unsupported. It is possible that a more rigorous analysis could reveal such a bias, but at minimum, the robustness to alternate sources of data must be assessed and a much more conservative approach taken to the problem of spatial correlation. The availability of dozens of 20th Century climate model simulations from the CMIP3 database (http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php) is an invaluable resource for testing any such corrections and should be used in assessing any new claims.

## Acknowledgments

Thanks to Rasmus Benestad, Mike Mann and two reviewers for their comments, and to Jos de Laat for clearing up some ambiguities. Climate modelling at GISS is funded by NASA MAP and output is available at data.giss.nasa.gov/modelE. Remote Sensing Systems MSU data are sponsored by the NOAA Climate and Global Change Program and are available at www.remss.com. Data associated with MM07 was downloaded from www.uoguelph.ca/~rmckitri/research/jgr07/jgr07.archive.zip. Temperature data were downloaded from www.cru.uea.ac.uk/cru/data/temperature.


## References


Benestad R. 2004. Are temperature trends affected by economic activity? Comment on McKitrick and Michaels. *Climate Research* **27**: 171–173.

Bindoff NL, Willebrand J, Artale V, Cazenave A, Gregory J, Gulev S, Hanawa K, Le Quéré C, Levitus S., Nojiri Y, Shum CK, Talley LD, Unnikrishnan A. 2007. Observations: Oceanic climate change and sea level. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: Cambridge, New York.

Brohan P, Kennedy J, Harris I, Tett S, Jones P. 2006. Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research* **111**: D12106.

Christy JR, Spencer RW, Norris WB, Braswell W. 2003. Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperatures. *Journal of Atmospheric and Oceanic Technology* **20**: 613–629.

de Laat ATJ, Maurellis AN. 2004. Industrial $CO_2$ emissions as a proxy for anthropogenic influence on lower tropospheric temperature trends. *Geophysical Research Letters* **31**: L05204.

de Laat ATJ, Maurellis AN. 2006. Evidence for influence of anthropogenic surface processes on lower tropospheric and surface temperature trends. *International Journal of Climatology* **26**: 897–913.

Gates WL, Boyle JS, Covey C, Dease CG, Doutriaux CM, Drach RS, Fiorino M, Gleckler PJ, Hnilo JJ, Marlais SM, Phillips TJ, Potter GL, Santer BD, Sperber KR, Taylor KE, Williams DN. 1999. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP1). *Bulletin of the American Meteorological Society* **80**: 29–55.

Hansen J, Sato M, Ruedy R, Kharecha P, Lacis A, Miller RL, Nazarenko L, Lo K, Schmidt GA, Russell G, Aleinov I, Bauer S, Baum E, Cairns B, Canuto V, Chandler M, Cheng Y, Cohen A, Del Genio A, Faluvegi G, Fleming E, Friend A, Hall T, Jackman C, Jonas J, Kelley M, Kiang NY, Koch D, Labow G, Lerner J, Menon S, Novakov T, Oinas V, Perlwitz Ja, Perlwitz Ju, Rind D, Romanou A, Schmunk R, Shindell D, Stone P, Sun S, Streets D, Tausnev N, Thresher D, Unger N, Yao M, Zhang S. 2007. Climate simulations for 1880–2003 with GISS modelE. *Climate Dynamics* **29**: 661–696.

IDAG (International ad hoc Detection and Attribution Group). 2005. Detecting and attributing external influences on the climate system: A review of recent advances. *Journal of Climate* **18**: 1291–1314.

Jones PD, Moberg A. 2003. Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *Journal of Climate* **16**: 206–223.

Jones PD, Osborn T, Briffa K. 1997. Estimating sampling errors in large-scale temperature averages. *Journal of Climate* **10**: 2548–2568.

Legendre P. 2002. Program for multiple linear regression with permutation test. Technical Report, Département de sciences biologiques, Université de Montréal: http://www.bio.umontreal.ca/casgrain/en/labo/regression.html.

Lemke P, Ren J, Alley RB, Allison I, Carrasco J, Flato G, Fujii Y, Kaser G, Mote P, Thomas RH, Zhang T. 2007. Observations: Changes in snow, ice and frozen ground. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: Cambridge, New York.

Livezey RE, Chen WY. 1983. Statistical field significance and its determination by Monte Carlo techniques. *Monthly Weather Review* **111**: 46–59.

McKitrick R, Michaels P. 2004. A test of corrections for extraneous signals in gridded surface temperature data. *Climate Research* **26**: 159–173.

McKitrick R, Michaels P. 2007. Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data. *Journal of Geophysical Research* **112**: D24S09.

Mears CA, Wentz FJ. 2005. The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science* **309**: 1548–1551.

Santer BD, Wigley TML, Boyle JS, Gaffen DJ, Hnilo JJ, Nychka D, Parker DE, Taylor KE. 2000. Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research* **105**: 7337–7356.

Schmidt GA, Ruedy R, Hansen JE, Aleinov I, Bell N, Bauer M, Bauer S, Cairns B, Canuto V, Cheng Y, Del Genio A, Faluvegi G, Friend AD, Hall TM, Hu Y, Kelley M, Kiang NY, Koch D, Lacis AA, Lerner J, Lo KK, Miller RL, Nazarenko L, Oinas V, Perlwitz Ja, Perlwitz Ju, Rind D, Romanou A, Russell GL, Sato M, Shindell DT, Stone PH, Sun S, Tausnev N, Thresher D, Yao MS. 2006. Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data. *Journal of Climate* **19**: 153–192, http://www.giss.nasa.gov/tools/modelE.

Spencer RW, Christy JR. 1990. Precise monitoring of global temperature trends from satellites. *Science* **247**: 1558–1562.

Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Klein Tank A, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, Soden B, Zhai P. 2007. Observations: Surface and atmospheric climate change. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: Cambridge, New York.

van Aardenne JA, Dentener FJ, Olivier JGJ, Klein Goldewijk CGM, Lelieveld J. 2001. A $1° \times 1°$ resolution data set of historical anthropogenic trace gas emissions for the period 1890–1990. *Global Biogeochemical Cycles* **15**: 909–928.

Wang X, Shen S. 1999. Estimation of spatial degrees of freedom of a climate field. *Journal of Climate* **12**: 1280–1291.

Zhang X, Zwiers F, Stott P. 2006. Multi-model multi-signal climate change detection at regional scale. *Journal of Climate* **19**: 4294–4307.

# Errata for "Spurious correlations between recent warming and indices of local economic activity" by G.A. Schmidt

A number of minor errata are required for a recent paper, Schmidt (IJoc, 2009).

- S09 mistakenly described the temperature trends for McKitrick and Michaels (2007) section as being from 1979-2001, when actually the trends are taken from 1979-2002 (inclusive) as in MM07.

- Calculations of the trends in CRU data in the paper were affected by an error in how missing data was dealt with. The results affected are Table 1 (the update to CRUv3+RSS). No substantial change in the text is required.

- The model SAT and MSU-LT trends used were too small by a constant factor of 0.7188 due to an error in how missing data was dealt with (the constant factor was particular to the number of years read in vs. the number of years used in the calculation). Since the relationship between the new corrected values and the originals is constant, only the regression coefficients for variables other than 'trop' in Tables 2 and 3 are affected (by roughly 1/0.7188, with some variation due to round off). None of the significance calculations, nor the discussion is affected.

| Variable | CRUv3+RSS MSU SURF | G3 |
|---|---|---|
| Intercept | -4.19849 | -2.76839 |
| trop | **0.96992** | **0.97963** |
| slp | 0.00437 | 0.00287 |
| dry | 0.39609 | -0.42244 |
| dslp | -0.00032 | 0.00048 |
| water | -0.00857 | -0.00412 |
| abslat | **0.00331** | **0.00334** |
| g | **0.04176** | **0.05273** |
| e | **-0.00296** | **-0.00246** |
| x | 0.00117 | 0.00144 |
| p | 0.15243 | |
| m | 0.27626 | |
| y | **-0.23124** | |
| c | **0.00825** | |
| N | 422 | 422 |
| $r^2$ | 0.51 | 0.50 |

**Table 1.** Partial original Table 1 (corrected).

| | SURF test (GISS-ER) | | | | | |
|---|---|---|---|---|---|---|
| Variable | runA | runB | runC | runD | runE | Ensemble |
| Intercept | 3.14215 | 2.88634 | 2.95166 | 9.53691 | -6.00788 | 1.56216 |
| trop | **1.44038** | **1.43259** | **1.32441** | **1.48327** | **1.12639** | **1.22117** |
| slp | -0.00307 | -0.00311 | -0.00307 | **-0.00965** | **0.00577** | -0.00165 |
| dry | -2.18790 | 0.03039 | 0.25836 | 1.28293 | 5.25923 | 1.13518 |
| dslp | 0.00212 | -0.00004 | -0.00023 | -0.00136 | -0.00520 | -0.00114 |
| Water | -0.02822 | **-0.04908** | **-0.06361** | **-0.04267** | **-0.05863** | **-0.05006** |
| abslat | -0.00007 | **0.00178** | **-0.00109** | **0.00272** | **0.00160** | **0.00103** |
| g | -0.01621 | -0.00268 | -0.00465 | **-0.02505** | -0.01548 | -0.01066 |
| e | -0.00042 | **0.00079** | **0.00121** | **0.00105** | **0.00084** | **0.00049** |
| p | **-0.24432** | **-0.12847** | -0.08108 | -0.08785 | 0.05528 | **-0.09191** |
| m | -0.21386 | **-0.23692** | -0.07981 | -0.04503 | -0.03921 | **-0.11909** |
| y | 0.12405 | **0.19291** | 0.07154 | 0.04976 | -0.00792 | 0.07907 |
| c | 0.00082 | -0.00159 | -0.00164 | 0.00094 | 0.00260 | 0.00044 |
| $r^2$ | 0.69 | 0.79 | 0.69 | 0.73 | 0.65 | 0.68 |

**Table 2.** Original Table 2 (corrected).

| | G3 test (GISS-ER) | | | | | |
|---|---|---|---|---|---|---|
| Variable | runA | runB | runC | runD | runE | Ensemble |
| Intercept | 4.63495 | 3.14662 | 2.17974 | 8.81204 | -3.63595 | 2.59554 |
| trop | **1.37773** | **1.41792** | **1.31052** | **1.46841** | **1.08877** | **1.22810** |
| slp | -0.00467 | -0.00328 | -0.00230 | **-0.00894** | 0.00344 | **-0.00269** |
| dry | -4.23896 | -0.44838 | 0.59262 | 1.50947 | 3.46035 | 0.04195 |
| dslp | 0.00414 | 0.00043 | **-0.00056** | -0.00158 | -0.00343 | -0.00006 |
| Water | **-0.03852** | **-0.04850** | **-0.06291** | **-0.03780** | **-0.06491** | **-0.05209** |
| abslat | 0.00075 | **0.00134** | -0.00103 | **0.00267** | **0.00138** | **0.00112** |
| g | -0.01857 | -0.01446 | -0.00411 | -0.02044 | **-0.02527** | **-0.01559** |
| e | 0.00013 | 0.00042 | **0.00108** | **0.00100** | **0.00100** | **0.00060** |
| $r^2$ | 0.67 | 0.79 | 0.69 | 0.72 | 0.62 | 0.67 |

**Table 3.** Original Table 3 (corrected)