

Spatial Autocorrelation and the Detection of Non-Climatic Signals in Climate Data

Short title:

Autocorrelation and socioeconomic influence on temperatures

Ross McKittrick

Department of Economics

University of Guelph

Guelph ON Canada N1G 2W1

Tel. 519-824-4120 x52532

Fax. 519-763-8497

rmckitri@uoguelph.ca

and

Nicolas Nierenberg

Nierenberg Foundation

9494 La Jolla Farms Rd.

La Jolla, CA 92037

Summary

Surface climate data undergo processing to remove non-climatic effects such as urbanization and measurement irregularities. Some studies have shown that the processing is inadequate, leaving a residual warm bias. This has been disputed on three grounds: spatial autocorrelation of the

temperature field undermines significance of test results; counterfactual experiments using model-generated data suggest earlier results were spurious; and different satellite covariates yield unstable results. Surprisingly, these claims have not been statistically tested. We combine the data sets of various teams with trend estimates from global climate models and test the competing hypotheses. We find that controls for non-climatic, socioeconomic influences are necessary for a well-specified model of surface trends, supporting the view that the climatic data are not adequately filtered.

Research Supported by Social Sciences and Humanities Research Council of Canada Grant
Number 430002

Key words: global warming, data quality, spatial autocorrelation, economic activity

Spatial Autocorrelation and the Detection of Non-Climatic Signals in Climate Data

1 Introduction

1.1 Background

Empirical climatology relies on the assumption that surface temperature data over land have been adjusted to remove effects due to local non-climatic influences, such as population growth, urbanization, equipment changes, data quality problems (especially in developing countries), variations in local air pollution levels, etc. The assumption that the adjustments are sufficient underpins the standard interpretation of climatic data (see., e.g., Jones et al. 1999, Jun et. al. 2008, p. 935). McKittrick and Michaels (2004a,b, herein MM04, and 2007, herein MM07) tested the assumption by regressing the observed 1979-2002 trends in 440 surface grid cells on a vector of climatological variables (lower tropospheric temperature trends and fixed factors such as latitude, mean air pressure and coastal proximity) augmented with a vector of socioeconomic variables, including income and population growth, Gross Domestic Product (GDP) per square km, education levels, etc. If the data have been adjusted to remove all non-climatic influences then the spatial pattern of warming trends should not vary systematically with socioeconomic indicators. MM04 and MM07 both rejected, at very high significance levels, independence of the surface temperature field and the socioeconomic variables, thus concluding that the adjusted surface climatic data likely still contain residual influences of industrialization on local temperature records. They estimated that the non-climatic effects could account for between one-third and one-half of the post-1979 average warming trend over land in the temperature data.

Schmidt (2009, herein S09) defended the standard interpretation of surface climate data on four grounds. First, he noted that an overall warming trend is observed in numerous data sets. This is not under dispute, but since the surface data show a relatively large trend compared to data from satellites, the accuracy of the land-based data is the point at issue. Second, he argued that the surface temperature field exhibits spatial autocorrelation (SAC), which reduces the “effective degrees of freedom” in the sample and biases the test statistics towards over-rejection of the null (no correlation) hypothesis. Third, he argued that use of the lower troposphere satellite series from Remote Sensing Systems (Mears et al. 2003, denoted RSS) rather than the University of Alabama-Huntsville series (Spencer and Christy 1990, denoted UAH), reduces the significance of the coefficients, indicating a lack of robustness of the conclusions. Fourth, he argued that the results were spurious on the basis of a comparison with results obtained by swapping the observed surface and tropospheric trends with model-generated data from NASA’s Goddard Institute of Space Studies (GISS) model E, denoted herein as GISS-E. These model-generated data are by construction uncontaminated by industrialization-induced surface changes. Schmidt’s hypothesis was that if the GISS-E data yield the same regression coefficients as the observational data in MM07, it would indicate that the seeming correlations between patterns of warming and patterns of industrialization were a fluke. This is not what the S09 GISS-E runs showed however (as we explain below), but S09 also proposed a more general argument that if *any* significant correlations appeared, this would imply the results of MM07 were spurious.

Focusing on the latter three points, it is noteworthy that they are all statistical in nature, but formal hypothesis testing was not presented in S09. In this paper we present a regression-based framework that allows direct testing to try and settle the unresolved counterclaims. The regression framework we use is common to all estimations herein. The unit of measurement is a 5 degree x 5 degree

grid cell on the land surface. The dependent variable consists of about 440 observations, each one a linear trend through 1979-2002 monthly temperatures in that grid cell. We refer to this vector as the “trend field.” The independent variables include climatic and geographic data for each grid cell that are expected, under the null hypothesis, to have explanatory power on the temperature trend vector, and socioeconomic data that are not. McKittrick (2010) shows that SAC was not detected in the residuals of the MM07 regression model. We extend those results herein to a large suite of data sets. Our consistent finding is that, when observed climatic data are used, the dependent variable is spatially autocorrelated but the regression residuals are not. This provides evidence that the explanatory model is well-specified and autocorrelation does not bias the inferences in MM04 or MM07. By contrast, when using data generated by a climate model (General Circulation Model, or GCM), spatial autocorrelation is observed in both the dependent variable and the residuals, indicating the explanatory variables do not provide a well-specified model of the GCM-generated trends. This may indicate that the process that generates the observations is structurally different than the processes represented in the climate models.

On the second claim, we do find that use of RSS data rather than UAH data weakens the MM07 coefficients, although removal of a small number of outliers from the data set eliminates this contrast. Also, S09 did not present joint significance tests on which the core conclusions were based, and using RSS data these still uphold the MM07 findings, albeit at reduced significance. For instance the null hypothesis of no-socio-economic effects rejects at $P = 2.04 \times 10^{-6}$ using RSS data, rather than $P = 9.92 \times 10^{-13}$ using UAH data; also some individual coefficients lose size and significance.

On the third point, S09 reported significant socioeconomic coefficients in a regression using GISS-E data. However, we show that the significance of individual coefficients disappears when the residuals are treated for SAC, something not done in the S09 analysis. In addition, the coefficients

estimated on GISS-E data, as well as those estimated on the ensemble means of a much larger suite of climate models, do not match the signs and magnitudes of those estimated on observations. This provides further evidence against the view that the socioeconomic correlations are spurious, since the coefficient pattern on observed data is significantly different from that on data generated by climate models operating on the assumption that local socioeconomic process do not influence surface trends. An additional piece of evidence comes from applying the filtering methodology of MM07 to the GISS-E data. The methodology uses the regression coefficients from the socioeconomic variables to estimate the trend distribution after removing the estimated non-climatic biases in the temperature data. On observational data this reduces the mean warming trend by between one-third and one-half, but it should not affect the mean surface trend in the model-generated data, something we observe to be the case. Again this is consistent with the view that the observations contain a spatial contamination pattern not present in, or predicted by, the climate models.

Finally, we look at the differences between observed surface trends and the predicted values from the GCM ensemble mean. If the models explain the observations, and if the observations have been filtered to remove socioeconomic influences, these trend differences should be independent of the socioeconomic variables. But we find that the differences are highly correlated with the socioeconomic indicators, and the coefficients are very close to those estimated on the observed trends themselves. This strengthens the argument that the socioeconomic pattern in the data is not accounted for by the processes in the climate models.

Taken together we find significant evidence against the view that surface climate data are free of biases due to socioeconomic development and other inhomogeneities. Instead, measures of socioeconomic contamination appear to be an essential component of a well-specified model of the

measured spatial temperature trend pattern over land. The coefficient pattern on observational data differs from that predicted by climate models as a response to natural oscillations and anthropogenic (greenhouse) forcing. Hence we consider the standard interpretation of climatic data untenable.

In the next section we explain the data sets used throughout this paper. In Section 2 we model spatial autocorrelation and give detailed results for the data configurations of interest. In Section 3 we explore the mismatch between the regression results from model-generated and observed data. Section 4 presents further specification tests and Section 5 concludes.

1.2 Data sets

Most data sets used herein are taken from MM07 and S09.¹ Readers should consult both these papers for detailed explanations; only a brief summary will be presented herein.

Contamination problems with raw temperature data are widely-acknowledged. The Climatic Research Unit (CRU) at the University of East Anglia disseminates some widely-used surface temperature data sets, including the ones used herein. Their web page (<http://www.cru.uea.ac.uk/cru/data/hrg/>) references data compilations called CRU TS 1.x, 2.x and 3.x which are not subject to adjustments for non-climatic influences. Users are explicitly cautioned (see <http://www.cru.uea.ac.uk/cru/data/hrg/timm/grid/ts-advice.html>) not to use these data for measuring climate change or attributing its origins because the data have not been adjusted to remove, *inter alia*, “all influences of urban development or land use change on the station data.” Users are directed instead to the CRUTEM data products, which, it is claimed, have been adjusted “for the reliable detection of

¹ The data are available respectively at <http://www.uoguelph.ca/~rmckitri/research/jgr07/jgr07.html> and http://www.giss.nasa.gov/staff/gschmidt/supp_data_Schmidt09.zip.

anthropogenic trends.” Readers are referred to Brohan et al. (2006), Jones and Moberg (2002) and Jones et al. (1999) for explanations of the adjustments.

It is not immediately apparent what the adjustments are, or how their adequacy has been tested. Brohan et al. (2006) is the paper that introduced Version 3 of the CRUTEM data base. It does not itself explain how the data are adjusted, instead it focuses on defending the claim that the potential biases are very small. Brohan et al. cautions (Section 2.3.3) 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” (p. 11). The authors work on the assumption that the problems add no more than 0.006 degrees per decade to the trend standard error. Jones and Moberg (2002) introduced the CRUTEM version 2 data product. This paper also has little information about the data adjustments. Reference is made to combining multiple site records into a single series, but not to removing non-climatic contamination. Moreover, the article points out (page 208) that it is difficult to say what homogeneity adjustments have been applied since the original data sources do not always hold this information.

Jones et. al (1999) emphasizes that non-climatic influences (therein referred to as “inhomogeneities”) must be corrected (Section 2, p. 37) for the data to be useful for climatic research. The only part of the paper that provides information on the adjustments is the following statement in Section 2.1 (page 174):

“All 2000+ station time series used have been assessed for homogeneity by subjective interstation comparisons performed on a local basis. Many stations were adjusted and some

omitted because of anomalous warming trends and/or numerous nonclimatic jumps (complete details are given by Jones et al. [1985, 1986c]).”

Jones et al. (1985, 1986c) are technical reports that were submitted to the US Department of Energy, and are online at <http://www.cru.uea.ac.uk/st/>. They only cover data sets ending in the early 1980s, whereas the data currently under dispute is the post-1979 interval. Those documents caution (page 3) that even with station-by-station examination, correction of all the problems is not possible due to insufficient detail in the site records to calculate correction factors. Even if the adjustments were adequate in the pre-1980 interval it is likely impossible to have estimated empirical adjustments in the early 1980s that would apply to changes in socioeconomic patterns that did not occur until the 1990s and after.

In sum, the CRU cautions that its unadjusted temperature data products (TS 2.x etc.) are contaminated with nonclimatic influences that make them unusable for the measurement and attribution of climatic changes. The CRU refers users instead to the CRUTEM products. Yet the accompanying documentation does not appear to explain the adjustments made or the grounds for claiming the latter data products are reliable for climate research purposes.

In order to test whether the adjustments were adequate, MM07 estimated the regression equation

$$\begin{aligned} \theta_i = & \beta_0 + \beta_1 TROP_i + \beta_2 PRESS_i + \beta_3 DRY_i + \beta_4 DSLP_i + \beta_5 WATER_i + \beta_6 ABSLAT_i \\ & + \beta_7 p_i + \beta_8 m_i + \beta_9 y_i + \beta_{10} c_i + \beta_{11} e_i + \beta_{12} g_i + \beta_{13} x_i + u_i \end{aligned} \quad (1)$$

where θ_i is the 1979-2002 trend in CRUTEM gridded surface climate data in grid cell i , $TROP_i$ is the time trend of Microwave Sounding Unit (MSU)-derived temperatures in the lower troposphere in the

same grid cell as θ_i over the same time interval, $PRESS_i$ is the mean sea level air pressure, DRY_i is a dummy variable denoting when a grid cell is characterized by predominantly dry conditions (which is indicated by the mean dewpoint being below 0 °C), $DSLP_i$ is $DRY_i \times PRESS_i$, $WATER_i$ is a dummy variable indicating the grid cell contains a major coastline, $ABSLAT_i$ denotes the absolute latitude of the grid cell, p_i is local population change from 1979 to 2002, m_i is per capita income change from 1979 to 2002, y_i is total Gross Domestic Product (GDP) change from 1979 to 2002, c_i is coal consumption change from 1979 to 2002, g_i is GDP density (national Gross Domestic Product per square kilometer) as of 1979, e_i is the average level of educational attainment, and x_i is the number of missing months in the observed temperature series and u_i is the regression residual. Equation (1) was estimated using the generalized least squares routine in Stata 8.0 with corrections for error clustering and heteroskedasticity.

For ease of notation we will drop the gridcell subscript i when doing so does not create ambiguity. Equation (1) explains the spatial pattern of temperature trends using three main variable groups: temperature trends in the lower tropospheric layer about 5 km above the surface, fixed geographical factors and socioeconomic variables. The geographical variables include latitude, coastal proximity, mean air pressure, etc. The socioeconomic variables measure factors that influence data quality, land use change, etc. The standard interpretation of climate data is that their effects have been filtered out of climatic data products like CRUTEM.

Summary statistics for the data are in Table 1. The MM07 data set has 440 records, one for every 5x5 degree grid cell over land for which adequate observations were available in the CRUTEM data archive to identify a trend over the 1979-2002 interval. Each record contains the linear surface trend

expressed as degrees C per decade, and the corresponding linear trend from the University of Alabama-Huntsville lower tropospheric record of Spencer and Christy (1990), denoted UAH.

The S09 data set comprises surface and tropospheric grid cell trends like those in MM07, except the surface trends are from later CRU compilations and the tropospheric trends are from Remote Sensing Systems (RSS) (Mears et al. 2003). S09 provides trends derived from the CRUTEM2v (Jones and Moberg 2002) and CRUTEM3v data sets (Brohan et al. 2006). For brevity the version used in MM07 is denoted CRU and the updated versions used in S09 are denoted CRU2v and CRU3v. As is clear in Table 1 these data sets are very similar to one another. CRU3v is the most recent but has slightly less spatial coverage compared to CRU and CRU2v (428 cells).

The tropospheric data used in MM07 and S09 were at a 2.5x2.5 degree level, one-fourth of the 5x5 CRU surface grid size, so the top-right tropospheric cell was used. For some of our calculations herein we retain the 2.5 degree scale aloft where our intent is to replicate earlier results. Otherwise, in order to reconcile the spatial scales between surface and tropospheric gridcells we develop matched 5x5 grid cells. We denote the data series in which four tropospheric cells have been combined to yield a 5x5 grid cell as UAH4 and RSS4.

S09 also provided synthetic trends from GISS-E. For a description of this model see CCSP (2008 Sect. 2.5.3) and Schmidt et al. (2006). The climate model was run five times, and the mean over the five runs was taken as the ensemble average. The mean trends in the GISS-E surface data are denoted herein as GISSES, and the mean trends in the GISS-E lower troposphere data are denoted GISSET. We also obtained the ensemble mean trends for 55 model runs used in the IPCC (2007) report (see Appendix), which are denoted as MSM (surface) and MTM (troposphere).

The average GISS-E land surface trend is 0.14 °C/decade, well below the reported trend of 0.30 °C/decade in the CRU3v compilation. The range of trends over land across all GCMs is 0.07 to 0.57 °C/decade with a mean of 0.225 °C/decade, putting the CRU3v data in the upper half of the model spread. The standard deviations of the ensemble mean modeled trends are much smaller (one-third or less) than those in the observational data. This is not because the trends are averaged across multiple model runs, instead the trends in individual model runs have very small standard deviations to begin with.

With a vector of trend terms on both the left- (surface) and right-hand (troposphere) side there are 24 possible data combinations: CRU, CRU2v, CRU3v at the surface, UAH, UAH4, RSS and RSS4 aloft, and GISS and the all-GCM averages. Additionally we will be examining 3 weighting schemes for the spatial autocorrelation terms, making 72 possible model configurations. Since there are many common results across different specifications we will only report those central to our argument, but other results are available on request. For instance, since CRU2v was not used in MM07 and has been superseded by CRU3v we will not report CRU2v results, and we will generally use RSS4 rather than RSS.

2 Spatial autocorrelation of the trend field

2.1 Testing framework

Both S09 and Benestad (2004) point out, correctly, that the surface temperature field is spatially autocorrelated, and argue that this can, in principle, bias the inferences from regressions on the spatial trend field. They both concluded on this basis that the results in MM07 and MM04 were unreliable. However neither one formulated the argument as a testable hypothesis, though S09 presented variograms of the dependent variable and some independent variables from MM07. It is insufficient to observe

autocorrelation in a dependent variable and conclude that the inferences from a regression model are therefore biased. An additional step in the argument is required, namely a test showing that the regression residuals also exhibit SAC. As we will show, they do when model-generated data are used, as in S09, but they do not when observational data are used, as in MM07. The contrast is important. Inferences concerning the coefficients in a regression model are based on the statistical properties of the residuals, not the dependent variable: the dependent variable does not even appear in the expression for the variance-covariance matrix estimator. Thus, the absence of SAC in the residuals of a regression model in which the dependent variable is spatially autocorrelated is evidence in support of the specification, i.e. that the right hand side variables do have explanatory power.

We test for residual spatial dependence as follows. The regression model (1) can be rewritten in matrix notation as

$$\mathbf{T} = \mathbf{X}\mathbf{b} + u \quad (2)$$

where \mathbf{T} is a 440×1 vector of temperature trends in each of 440 surface grid cells, \mathbf{X} is a $440 \times k$ matrix of climatic and socioeconomic covariates, \mathbf{b} is a $k \times 1$ vector of least-squares slope coefficients and u is a 440×1 residual vector. Spatial autocorrelation in the residual vector can be modeled using

$$u = \lambda \mathbf{W}u + e \quad (3)$$

where λ is the autocorrelation coefficient, \mathbf{W} is a symmetric $n \times n$ matrix of weights that measure the influence of each location on the other, and e is a vector of homoskedastic Gaussian disturbances, (Pisati

2001). The rows of \mathbf{W} are standardized to sum to one. n equals 440 except in some regressions where grid cells are missing, as noted below.

A test of $H_0 : \lambda = 0$ measures whether the error term in (1) is spatially independent. As argued in S09, it is likely the dependent variable is spatially autocorrelated. Anselin et al. (1996) point out that if the alternative model allows for possible spatial dependence of \mathbf{T} , i.e.

$$\mathbf{T} = \phi \mathbf{ZT} + \mathbf{Xb} + e \quad (4),$$

where \mathbf{Z} is a matrix of spatial weights for \mathbf{T} and may not be identical to \mathbf{W} , then conventional tests of $\lambda = 0$ assuming an alternative model of the form $\mathbf{y} = \mathbf{X}\beta + e$ will be biased towards over-rejection of the null. They derive a $\chi^2(1)$ Lagrange Multiplier (LM) test of $\lambda = 0$ robust to possibly nonzero ϕ in (4), which has substantially superior performance in Monte Carlo evaluations compared to the non-robust LM test. All results quoted herein use the robust form of the LM test.

Hypothesis tests, and any subsequent parameter estimations, are conditional on the assumed form of the spatial weights matrix \mathbf{W} in (3). Denote the great circle distance between the grid cell centers from which observation i and observation j are drawn as g_{ij} . We examine three weighting functions: inverse square root weights are $1/\sqrt{g_{ij}}$, inverse linear weights are $1/g_{ij}$ and inverse square weights are $1/g_{ij}^2$. These weights allow the relative influence of one cell on adjacent cells to decline by relatively slower to relatively faster rates, respectively.

2.2 Spatial Autocorrelation Testing Results

Table 2 presents the results of SAC hypothesis tests on both the dependent variable and residuals for eight different model configurations, reporting for each of them the results based on between one and three different spatial weighting schemes. A result common to all blocks is that the log-likelihood values are lowest for the inverse-root weighting scheme $1/\sqrt{g_{ij}}$ and highest for the inverse-square weighting scheme $1/g_{ij}^2$. This means that the data give progressively more support to the spatial weighting schemes that decay most quickly with distance. Since this was observed in every estimation we do not always report the inverse-root and inverse-linear results.

The first block of results refers to the original configuration in MM07: the CRU gridded trends regressed on the UAH tropospheric trends and the rest of the MM07 model variables in Equation (1), and the second block shows the results with the dependent variable updated to CRU3v and the tropospheric series revised to be UAH4. Under the three spatial weighting schemes, in both data configurations, there is no evidence of SAC in the residuals. In four of the six configurations we reject the null hypothesis of spatial independence in the dependent variable. In other words, there is reasonably consistent evidence of SAC in the trend field, but the regression model explains it and we find no evidence of spatial dependence in the residuals. Hence there is no evidence SAC biases the regression inferences reported in MM07, nor would it do so if updated CRU3v data were used.

The next two blocks report the same results for the two surface trend groups (CRU and CRU3v), with the RSS4 data substituted in for UAH. The pattern is the same in all six cases: SAC is found in the dependent variable but not in the residuals. Consequently the inferences in MM07 would not have been undermined by SAC if these data sets had been used.

Having established that the list of variables on the right hand side of in Equation (1) explains the trend field sufficiently well to leave an uncorrelated residual, we were curious about what would happen

if we made only one change, namely instead of using observed (UAH4 or RSS4) tropospheric trends we used model-generated trends from GISS-E or the all-GCM ensemble mean. If similar results emerged, i.e. SAC in the dependent variable but not in the residuals, this would indicate a good structural match between the spatial structure of the model and that of the actual climate. The next two rows show the results using, respectively, GISSET and MTM for the lower troposphere data, with CRU3v used as the dependent variable, under the inverse-square weighting scheme, which obtains the highest likelihood values. The spatial lag term on the residuals is now significant in both cases, indicating that use of observed tropospheric trends is key for removing SAC in the residuals, and indicating the possible absence of an essential explanatory in climate models relating to the spatial dependence of actual temperature trends.

The final two blocks of Table 2 show the results using model-generated data on both sides of Equation (1). Interestingly the no-residual SAC null is rejected in all cases, indicating that the structural mismatch between surface observations and tropospheric trends also occurs when model-generated surface trends are used on both sides. We do not reject the null of spatial independence on the dependent variable when the all-GCM ensemble mean is used, but it is rejected in the residuals, indicating that while Equation (1) shows evidence of being a well-specified model of observational data, it is a poorly-specified model of GCM-generated data.

As noted above, the explanatory variables in Equation (1) can be put into three groups: the tropospheric trends, the geographic variables (*SLP* to *ABSLAT*) and the socioeconomic variables (*g* to *c*). Using CRU3v and RSS4, we examined the effect on SAC in the residuals of using any two of the three data groups. Under the *W3* weighting rule, RSS4 plus the geographic variables yields a marginally significant score ($P = 0.055$), i.e. rejecting the null of independent residuals. Using RSS plus only the

economic variables yields a robust test statistic of 0.000, $P = 0.986$, thus failing to reject the null of independent residuals. Using the geographic plus economic variables yields $P = 0.02$, rejecting the null. Consequently, for Equation (1) to remove the SAC from the residuals it is necessary for both the tropospheric trends and the socioeconomic variables to be included in the regression model. Inclusion of the socioeconomic variables alone is not sufficient to remove SAC ($P = 0.004$).

To summarize, in regressions using observational temperature data, the inclusion of spatial socioeconomic data is necessary to remove the SAC from the regression residuals, whereas any data combination omitting these variables fails to do so. Hence it appears that the socioeconomic data are a necessary component of a well-specified explanatory model of surface temperature trends in the CRU data sets. When climate model-generated data are used, SAC is not removed from the residuals under the same specifications, providing further evidence that a component of the variability in the CRU data arises from factors not accounted for by the anthropogenic and natural forcings coded into the GCMs. It also indicates that correction for SAC is required in the regressions that use modeled data to test whether the observed correlations are spurious, and we will show in the next section that this makes a difference to earlier results.

3 Do GCMs predict the observed temperature-industrialization correlation pattern?

We begin this section by looking at whether the MM07 results were unique to the particular data configuration used therein. Table 3 presents the regression coefficients for Equation (1) under a sequence of data configurations. The second column (MM07) replicates the results from the MM07 model, namely the CRU-UAH data pair. The next six columns apply, in sequential combinations, CRU2v, CRU3v, UAH, RSS, UAH4 and RSS4. The C3/R4x column repeats the CRU3v/RSS4 results with outliers

removed. The methodology for outlier removal is part of the specification tests described in the next Section. The Ordinary Least Squares “hat matrix” is evaluated, and an observation is flagged as an outlier if it exceeds twice the mean diagonal element of the hat matrix (see Kmenta 1986 424—426). In this case 26 observations were removed. This had the effect of bringing the CRU3v/RSS4 results into line with the results computed with the UAH data. The final two columns use the GISS-E and all-GCM ensemble means.

The coefficient estimates are reasonably similar across the observational columns (2 though 9). Use of CRU3v data yields smaller socioeconomic coefficients compared to CRU2v and CRU. This is consistent with the claim that the updates have improved the filtering process, although the effect is lost when outliers are removed from the sample. Use of RSS data yields smaller and less significant coefficients than the UAH data, though for some reason leaves a greater component to be explained by the latitude variable. Use of reconciled gridcell sizes also yields smaller and less significant coefficients compared to the 2.5x2.5 tropospheric grids. However, across all these specifications the coefficient sizes and signs remain comparable and the socioeconomic effects taken as a group $P(g=c=0)$ remain jointly significant. The socioeconomic effects thus seem to be a robust feature of the data.

S09 hypothesized that these effects arise from a fortuitous match between the spatial pattern of socioeconomic activity and the spatial pattern of enhanced natural and greenhouse forcing of the climate. Since the GCM does not contain a socioeconomic component, if, upon using GISS-E and GISS-E in place of observations in the MM07 regression model, significant coefficients of the same approximate size and sign emerge on the socioeconomic variables, then correlations such as those in MM07 could be dismissed as coincidental. It is worth quoting the argument in S09 directly to make this point clear.

“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.*

(S09, p. 2, emphasis added)

However, as is shown in columns 10 and 11 of Table 3, the regression coefficients estimated on the data generated by the GISS-E ensemble and the all-GCM ensemble are quite different from those estimated on observational data. A superscript *a* in Table 3 denotes cases in which the coefficient estimate on model-generated data takes the opposite sign to one estimated on observed data. Individual socioeconomic coefficients in columns 10 and 11 are insignificant. S09 reported that some were significant, however, that estimation did not take the residual SAC into account, which we saw in Table 2 is necessary when doing the estimation on model-generated data. Columns 10 and 11 in Table 3 include a correction for SAC.

With regard to the quoted paragraph, the distributions of the coefficients estimated on GCM data do not encompass the coefficients from either the MM07 data set or any other observational grouping in Table 3. Table 4 lists the 95% Confidence Intervals for the socioeconomic coefficients estimated on either GISS or all-GCM ensemble mean data, along with four columns of indicator variables. The indicators take the value of 1 if the coefficient estimated on observed data is within the comparison CI or not. As is shown, all indicator variables take a zero value. We can also conduct a test that takes into

account the uncertainty bounds on the observation-based coefficients. At the bottom of Table 4 is a group of P -values reporting a chi-square test of parameter equivalence between the indicated models. The P -values are all extremely small, indicating clear rejection of the hypothesis that the coefficient distributions overlap. Hence the null hypothesis in the form stated by S09, namely parameter overlap implying an absence of contamination, is rejected.

With the SAC correction applied, using GISS-E data the socioeconomic variables taken as a group are only marginally significant ($P = 0.075$). Using the all-GCM ensemble the socioeconomic variables are jointly significant ($P = 0.001$). However, in this case, since the distributions do not overlap and the coefficients of interest take opposite signs, significance of either set of coefficients increases, rather than decreases, the evidence that the observations contain a pattern at odds with that in the models.

The spatial autocorrelation results in the previous section, and the mismatch between coefficients estimated on observed and model-generated data, point to the likely existence of a non-climatic contamination pattern in the observed surface trend data. Further evidence on this point is obtained by repeating the filtering experiment of MM07 on the GISS and GCM data. Since the model does not contain any contaminating processes, we should not expect much difference between the raw data from the models and that obtained by applying the MM07 method for removing socioeconomic effects. Table 5 shows that the adjustments are, indeed small, and actually raise the GISS-E trend slightly rather than reducing it. In the MM07 case, using observational data, the filtering removed about one-half the surface trend. When RSS4 data (with CRU3v) are used, the filtering removes about one-third of the mean surface trend. But when GISS or all-GCM data are used, the filtering step does not affect the mean surface trend. This is consistent with the view that the models do not contain processes that are, by contrast, present in the observational data, and which are explained by the socioeconomic variables.

Finally, we generated the vector of differences between the observed surface trends (CRU3v) and the all-GCM mean trends. To the extent the GCM ensemble mean summarizes the projected surface effects of greenhouse gases and other known climate forcings, the differences with observations should be unsystematic, as long as the observational data are truly generated by processes that are fully represented by the current suite of climate models. Regressing the trend differences on the right-hand side variables in Equation (1) (except x since there are no missing cells in the GCM grid), we find that there are strong systematic patterns that resemble the patterns in the surface data itself. Table 6 presents the coefficient estimates of the trend differences and the coefficients from the same regression using the CRU3v trends, in both cases using the RSS4 tropospheric trends. The similarities in the coefficients is striking, and is even stronger if the UAH4 data are used (not shown). It implies that if we take the observational trends in gridded surface data and subtract that portion explained by climate models, the remaining portion exhibits the same pattern of correlations to socioeconomic variables as did the original observations. Consequently the original coefficients cannot be attributed to the forcings and processes as represented in climate models, and can instead be considered as likely due to the non-climatic, contaminating influences measured by the socioeconomic variables.

4 Further specification tests

MM07 presented a series of specification tests using the UAH data. We repeated all of them using CRU3v and RSS4. The results closely follow those reported in MM07. Details are available on request, and can be summarized as follows.

- The RESET test does not reject a null hypothesis of no un-modeled residual nonlinearity ($P = 0.241$).

- The Hausman test does not reject a null hypothesis of no endogeneity bias ($P = 0.999$).
- The outlier test (described in Section 3) flags 26 observations as influential. When these are removed the individual and joint socioeconomic coefficient tests become more significant, yet we do not reject a null hypothesis that the coefficient vectors with and without the outliers are equivalent ($P = 0.278$).
- Coefficient results are individually and jointly significant in rich countries but not poor countries, and in economies with growing but not declining incomes.
- After removing a randomly-selected third of the data set and re-estimating the model, the prediction of the withheld sample scatters along a 45-degree line with the observed values. In 500 repetitions, a regression of the predicted and observed values has a constant of 0.011 and a slope of 0.961, and a test of a perfect fit (constant = 0, slope = 1) obtains an average P value of 0.407, i.e. does not reject on average.

The SAC tests reveal that one of the specification tests in MM07 was done incorrectly. In Section 4.6 of MM07, an alternative estimation is presented in which the surface trends were replaced by the UAH-derived tropospheric trends. Had the socioeconomic coefficients retained their size and significance it would provide evidence that the surface results are spurious. As shown in Table 3 of MM07, the socioeconomic coefficients generally lose size and significance, as expected, although an anomalous result emerges whereby the missing variable count in surface data (denoted x) becomes significant. Only 5% of the cells in the sample have at least one missing month. The analysis in MM07 suggests that the occurrence of missing data is possibly acting as a proxy for relatively moist, storm-prone regions, but in any case x is small and insignificant in the main models so its role is unlikely to bias the conclusions.

However, when the tropospheric trends vector is used as the dependent variable, SAC does not disappear in the model residuals, so the regression equation needs to be augmented with a spatial error-correction process. This was not done in MM07. Table 7 shows the results of replacing the dependent variable with the tropospheric trends and regressing on the remaining right-hand side variables. The second column reproduces the original (incorrect) results from MM07, and the next two columns show the results after applying the SAC correction with inverse-square distance weights using, respectively, UAH4 and RSS4.

As in MM07, x changes sign and acquires significance, indicating that in this regression it is likely acting as a proxy for some unrelated spatial pattern. In the full regression x is not significant so this effect is spurious, as in MM07. Regarding the other coefficients, there is a plausible attenuation of the surface effects pattern, though one conspicuous anomaly is the e variable when the UAH4 trends are in the model. The coefficient falls to one quarter its original size and changes sign, but becomes significant. This does not occur when RSS4 data are used. The other coefficients diminish in size and in several cases lose significance. The population growth measure is smaller but still significant using the RSS4 data, but is insignificant using the UAH4 record. The joint significance tests, however, do represent evidence against the MM07 interpretation of this model, since they yield joint significance for the socioeconomic effects taken together. If we are truly measuring socioeconomic influences on surface temperatures, except for some land use-related measures that might be detectable in the troposphere these should be sufficiently attenuated in the lower troposphere as to leave insignificant coefficients.

Since the joint significance arises from the influence of coefficients that change sign compared to the original model, it is reasonable to suppose that the effects in this regression are not the same as in the full model specifications, supporting the view that e and x are here acting as weak proxies for

something else. Also the lambda coefficients (representing the lag-1 spatial autocorrelation coefficients) indicate that our treatment of the residual dependency is likely inadequate, so judgments about significance in these results should be considered tentative. Each lambda coefficient is greater than 0.99 and has a very large t -statistic. In time-series applications this typically suggests that a lag-1 process is inadequate, and in the present context the results suggest that a spatial analogue to a higher-order autocorrelation process may be needed to obtain unbiased variances.

The results in Table 7 are close enough to our expectations that we maintain our interpretations of the rest of our analysis with respect to the complete model estimations and the comparison of modeled versus observational data. However they do raise the caution that we cannot decisively rule out the presence of some kind of spurious spatial match between climatic variations and some of our measures of socioeconomic development. To provide a decisive test will require development of an updated data base that combines both time series and cross-sectional data on a gridcell basis. This is planned as a subsequent development of this line of inquiry.

5 Conclusions

We have examined the question of whether spatial trend patterns in surface temperature data can be explained in part by non-climatic, socioeconomic processes of the kind that are supposed to have been filtered out of the gridded data products. One strand of argument against earlier findings on this issue was that spatial autocorrelation of the temperature field reduces the effective number of degrees of freedom, biasing significance calculations. We have estimated robust SAC test statistics and have shown that while the trend field is spatially autocorrelated, under a range of specification and spatial weighting rules SAC is not found in the residuals as long as both the tropospheric trends and the socioeconomic

covariates are included in the model. Since inclusion of socioeconomic covariates is necessary for a well-specified model, our findings support the interpretation of the regression results as evidence of non-climatic biases on surface temperature data.

We have also looked at the effects of swapping in model-generated data for observations to see if similar results emerge, which would indicate that they are spurious in both types of data sets. When model-generated data are used in place of observations, SAC emerges in the regression residuals. When corrected, the significance of the coefficients largely disappears, especially in the GISS-E results, overturning a specific S09 argument. Additionally, the regression coefficients calculated using the model-generated data do not overlap with those generated using observations, and we reject at high significance the hypothesis that the coefficient vectors are identical. Hence, instead of the GCM counterfactuals showing the results on observational data are spurious, they strengthen the argument that they are not.

The original regression results in MM07 are mirrored across a wide range of data sets at the surface and tropospheric levels. Joint significance tests remain very significant and in each case uphold the MM07 inference that the surface trend field is affected by industrialization and other forms of socioeconomic changes. Depending on whether the RSS or UAH satellite data are used the filtered data yields a surface trend between one-third and one-half smaller than the reported average over land.

The data set presented in MM07 includes trends up to the end of 2002, and includes coarse resolution of some socioeconomic variables at the national level. Further investigation of the potential surface climatic data problems we have identified herein could involve a reconstruction of the MM07 data base using updated socioeconomic and climatic variables, use of cross-sectional time series (panel)

regression rather than trend fields, and use of regional, rather than national, socioeconomic data where available.

Acknowledgments

We thank Chad Herman for assistance in obtaining the PCMDI archive trends. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

References

- Anselin, L., Anil K. Bera, Raymond Florax and Mann J. Yoon (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26: 77—104.
- Benestad RE (2004) Are temperature trends affected by economic activity? Comment on McKittrick & Michaels (2004). *Clim. Res.* 27:171–173
- Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548
- CCSP (Climate Change Science Program). 2006. *Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences*. Thomas R. Karl, Susan J. Hassol, Christopher D. Miller, and William L. Murray, editors, 2006. A Report by the Climate Change Science Program and the Subcommittee on Global Change Research, Washington, DC.
- Davidson, R. and J.G. MacKinnon (2004), *Econometric Theory and Methods* Toronto: Oxford.
- IPCC (2007) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* S. Solomon, D. Qin,

- M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Jones, P.D. and A. Moberg (2003). “Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001.” *Journal of Climate* 16 203—223.
- Jones, P.D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, (1999) Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics.*, **37**, 173–199.
- Jun, Mikyoung, Reto Knutti and Douglas W. Nychka (2008) Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *Journal of the American Statistical Association* 108 No. 483 934—947 DOI 10.1198/016214507000001265.
- Kmenta, Jan (1986) *Elements of Econometrics* New York: MacMillan.
- McKittrick, R.R. (2010), Atmospheric Oscillations do not Explain the Temperature-Industrialization Correlation. *Statistics, Politics and Policy* Vol 1, July 2010.
- McKittrick, R.R. and P. J. Michaels (2004a), A test of corrections for extraneous signals in gridded surface temperature data, *Climate Research* 26(2) pp. 159-173, Erratum, *Clim. Res.* 27(3) 265—268.
- McKittrick, R.R. and P. J. Michaels (2004b), Erratum, *Clim. Res.* 27(3) 265—268.
- McKittrick, R.R. and P.J. Michaels (2007), Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data, *J. Geophys. Res.*, 112, D24S09, doi:10.1029/2007JD008465.
- Mears, Carl A., Matthias C. Schabel, and Frank J. Wentz. 2003. A Reanalysis of the MSU Channel 2 Tropospheric Temperature Record. *Journal of Climate* 16, no. 22 (November 1): 3650-3664 .
- Pisati, Maurizio (2001) Tools for spatial data analysis. *Stata Technical Bulletin* STB-60, March 2001, 21—37.

Santer, B. D., P. W. Thorne, L. Haimberger, K. E. Taylor, T. M. L. Wigley, J. R. Lanzante, S. Solomon, M. Free, P. J. Gleckler, and P. D. Jones (2008) Consistency of Modelled and Observed Temperature Trends in the Tropical Troposphere. *International Journal of Climatology* DOI: 10.1002/joc.1756.

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

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. (2008). Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data. *Journal of Climate* **19**: 153–192.

Spencer, R.W. and J.C. Christy (1990), Precise monitoring of global temperature trends from satellites, *Science* 247:1558—1562.

Var	Definition	Obs	Mean	Std. Dev.	Min	Max
CRU	Surface temperature trend from MM07	440	0.302	0.257	-0.700	1.020
CRU2v	CRUTEM version 2 trends from S09	440	0.296	0.250	-0.699	1.015
CRU3v	CRUTEM version 3 trends from S09	428	0.303	0.253	-0.717	1.042
GISS-ES	Surface gridcell trend from GISS model (S09)	440	0.196	0.113	-0.127	0.958
MSM	Mean surface gridcell trend from all GCM's	440	0.231	0.072	0.071	0.564
RSS	Lower tropospheric gridcell trend from RSS	434	0.237	0.134	-0.085	0.684
UAH	Lower tropospheric gridcell trend from UAH	440	0.232	0.184	-0.197	0.683
GISS-ET	Lower tropospheric gridcell trend from GISS model (S09)	440	0.222	0.077	-0.022	0.458
MTM	Lower tropospheric gridcell trend from all GCM's	440	0.234	0.030	0.131	0.314
Water	Grid cell contains a coast line	440	0.6045	0.4895	0	1
Abslat	Absolute latitude	440	40.602	17.953	2.5	82.5
g	1979 Real National GDP per sq km in millions	440	0.297	0.600	0.001	3.002
e	Literacy +Post-secondary education rates	440	106.5	26.20	11.6	144.2
x	# missing months in grid cell temperature record	440	0.764	2.552	0	24
p	% growth in population*	440	0.279	0.209	-0.069	1.235
m	% growth in real average income*	440	0.380	0.614	-0.790	2.147
y	% growth in real national GDP**	440	0.771	0.839	-0.669	3.003
c	% growth in coal consumption*	440	1.016	4.056	-1	39.333
Rich	1999 real income > median	440	0.493	0.501	0	1
Grow	1999 real income > 1979 real income	440	0.761	0.427	0	1

Table 1: Model Variables. Definitions discussed further in MM07 and S09. *over the interval 1979 to 1999. **Over the interval 1980 to 2000. % Changes should be multiplied by 100, e.g. mean population growth is 27.9%.

Estimation Model	Weighting Scheme	Dependent Variable	Residuals
CRU: UAH + MM07	$1/\sqrt{g_{ij}}$	5.599 (0.018)	2.564 (0.109)
	$1/g_{ij}$	7.369 (0.007)	0.032 (0.858)
	$1/g_{ij}^2$	15.374 (0.000)	0.094 (0.759)
CRU3v: UAH4 + MM07	$1/\sqrt{g_{ij}}$	1.941 (0.164)	1.045 (0.307)
	$1/g_{ij}$	2.325 (0.127)	0.132 (0.717)
	$1/g_{ij}^2$	5.677 (0.017)	0.016 (0.898)
CRU: RSS4 + MM07	$1/\sqrt{g_{ij}}$	4.351 (0.037)	1.325 (0.250)
	$1/g_{ij}$	5.477 (0.019)	0.894 (0.344)
	$1/g_{ij}^2$	11.625 (0.001)	0.189 (0.664)
CRU3v: RSS4 + MM07	$1/\sqrt{g_{ij}}$	4.145 (0.042)	1.597 (0.206)
	$1/g_{ij}$	4.839 (0.028)	0.384 (0.536)
	$1/g_{ij}^2$	10.130 (0.001)	0.018 (0.894)
CRU3v: GISSET+MM07	$1/g_{ij}^2$	12.953 (0.000)	6.823 (0.009)
CRU3v: MTM+MM07	$1/g_{ij}^2$	19.975 (0.000)	4.730 (0.030)
GISSES: GISSET + MM07	$1/\sqrt{g_{ij}}$	43.108 (0.000)	39.232 (0.000)
	$1/g_{ij}$	48.528 (0.000)	84.522 (0.000)
	$1/g_{ij}^2$	51.478 (0.000)	139.484 (0.007)
MSM: MTM + MM07	$1/\sqrt{g_{ij}}$	1.082 (0.298)	10.971 (0.001)
	$1/g_{ij}$	0.828 (0.363)	125.469 (0.000)
	$1/g_{ij}^2$	1.176 (0.278)	127.484 (0.040)

TABLE 2. Spatial autocorrelation tests for regression models. Estimation model described as [surface measure]: [tropospheric measure] + MM07. Surface measure is CRU gridded surface trend vector from McKittrick and Michaels (2007), showing 1979-2002 trend per grid cell, or CRU3v update from Brohan et al (2007), or model-generated (GISSES, MSM). Tropospheric measure is either UAH or RSS

observational trends ('4' denotes expanded to 5x5 gridcell), or model-generated (GISSET, MTM). MM07 denotes other dependent variables from McKittrick and Michaels (2007) model, namely *SLP* through *c* in Equation (1). Weighting scheme refers to assumed form of spatial dependence. Third and fourth columns: each entry shows estimated autocorrelation parameter and associated *P* value of hypothesis that it equals zero. **Bold** denotes hypothesis rejected at 5% significance.

Variable	MM07	C2/UAH	C2/RSS	C3/UAH	C3/RSS	C3/UAH4	C3/RSS4	C3/R4x	GISS-E	GCM
Trop	0.8631	0.8323	0.9872	0.8146	0.9627	0.9330	0.9538	0.9263	1.5622	1.5542
	(8.62)	(8.58)	(14.36)	(8.80)	(13.84)	(11.03)	(8.77)	(7.49)	(13.43)	(13.17)
slp	0.0044	0.0043	0.0024	0.0058	0.0039	0.0071	0.0059	0.0043	-0.0042^a	-0.0020^a
	(1.02)	(1.03)	(0.66)	(1.43)	(1.11)	(1.92)	(1.67)	(1.33)	(2.55)	(2.33)
dry	0.5704	1.6771	0.1986	1.5010	0.2125	4.584	4.3175	2.8723	-0.5264 ^a	1.6915
	(0.10)	(0.33)	(0.04)	(0.30)	(0.04)	(1.01)	(1.01)	(0.52)	(0.31)	(1.68)
dslp	-0.0005	-0.0016	-0.0001	-0.0014	-0.0001	-0.0044	-0.0042	-0.0027	0.0005 ^a	-0.0016
	(0.09)	(0.31)	(0.03)	(0.28)	(0.03)	(0.99)	(0.99)	(0.51)	(0.30)	(1.65)
water	-0.0289	-0.0293	-0.0169	-0.0240	-0.0117	-0.0278	-0.0161	-0.0155	-0.0288	-0.0168
	(1.37)	(1.37)	(0.68)	(1.06)	(0.46)	(1.29)	(0.66)	(0.56)	(3.98)	(5.28)
abslat	0.0006	0.0010	0.0028	0.0014	0.0033	0.0005	0.0041	0.0040	0.0016	0.0028
	(0.51)	(0.82)	(2.28)	(1.22)	(2.65)	(0.54)	(3.00)	(2.54)	(3.64)	(9.42)
g	0.0432	0.0450	0.0444	0.0449	0.0446	0.0383	0.0401	0.0383	-0.0009 ^a	0.0002
	(3.36)	(3.66)	(3.03)	(4.01)	(3.27)	(3.44)	(2.58)	(1.78)	(0.13)	(0.06)
e	-0.0027	-0.0026	-0.0025	-0.0029	-0.0028	-0.0028	-0.0025	-0.0031	-0.0000	-0.0003
	(5.14)	(5.32)	(4.73)	(4.41)	(4.10)	(4.32)	(3.73)	(4.75)	(0.01)	(1.82)
x	0.0041	0.0019	-0.0003	0.0011	-0.0021	0.0016	0.0005	0.0006		
	(1.66)	(0.73)	(0.12)	(0.42)	(0.95)	(0.71)	(0.20)	(0.12)		
p	0.3839	0.3665	0.1513	0.3524	0.1450	0.2916	0.2085	0.3969	-0.0795 ^a	0.0117
	(2.72)	(2.67)	(1.04)	(2.52)	(0.97)	(2.18)	(1.55)	(2.71)	(1.55)	(0.57)
m	0.4093	0.3844	0.2663	0.3732	0.2578	0.2810	0.2595	0.5553	-0.0798 ^a	0.0057
	(2.39)	(2.33)	(1.52)	(2.19)	(1.42)	(1.68)	(1.50)	(3.00)	(1.45)	(0.20)
y	-0.3047	-0.2839	-0.2160	-0.2804	-0.2139	-0.2178	-0.1944	-0.4300	0.0452 ^a	-0.0094
	(2.22)	(2.15)	(1.55)	(2.06)	(1.48)	(1.63)	(1.42)	(2.91)	(1.13)	(0.46)
c	0.0062	0.0060	0.0076	0.0063	0.0079	0.0057	0.0071	0.0084	0.0006	0.0012
	(3.45)	(3.46)	(3.77)	(3.38)	(3.62)	(2.93)	(3.58)	(1.92)	(1.07)	(3.35)
Constant	-4.2081	-4.1522	-2.2291	-5.5546	-3.7704	-6.9317	-5.8808	-5.3457	4.1237^a	1.9070^a
	(0.96)	(0.97)	(0.60)	(1.36)	(1.05)	(1.84)	(1.63)	(1.23)	(2.45)	(2.14)
P(H:g—c=0)	0.000	0.075	0.003							
N	440	440	434	428	422	428	428	402	440	440

R^2	0.53	0.53	0.52	0.53	0.52	0.56	0.54	0.55	0.65	0.71
Log-likelihood	139.22	151.75	145.50	142.64	135.83	157.10	147.54	135.35	640.84	904.28

Table 3 Regression results from a group of data configurations. MM07 shows results from original data set (McKittrick and Michaels 2007). Next six columns show combinations of CRUv2 and CRUv3 surface data sets (denoted C2 and C3 respectively) and the UAH and RSS lower tropospheric data series on the original gridcell basis or the reconciled 5x5 basis (UAH4, RSS4). C3/R4x is CRU3v/RSS4 configuration with outliers removed. Final columns show regressions using GISS-E- and all-GCM-generated surface and tropospheric data. Numbers in parentheses are absolute t -statistics. For Columns 2-8 these reflect corrections for clustered errors and heteroskedasticity. For Columns 10 and 11 a correction for spatial autocorrelation (inverse-square weights) is also applied. **Bold** denotes significant at 95% confidence. ^a indicates the sign of the GISS-E coefficient is opposite to that in all the observation-based regression. $P(I)$ is prob value of test that all inhomogeneity factors (g-x) are jointly zero; $P(S)$ = prob value of test that all surface process coefficients (p-c) are jointly zero; $P(all)$, prob value of test that g -c are jointly zero. R^2 in final two columns is squared correlation between observed and predicted regression values. Variable x is dropped in GISS-E and GCM regressions since there are no missing surface values.

Variable	GISS 95% CI	CRU/UAH Compared to		CRU3v/RSS Compared to		GCM 95% CI
		GISS	GCM	GISS	GCM	
g	-0.0139 to 0.0122	0	0	0	0	-0.0061 to 0.0065
e	-0.0005 to 0.0005	0	0	0	0	-0.0005 to 0.0000
p	-0.1824 to 0.0234	0	0	0	0	-0.0291 to 0.0524
m	-0.0804 to -0.0793	0	0	0	0	-0.0501 to 0.0614
y	-0.0347 to 0.1251	0	0	0	0	-0.0506 to 0.0318
c	-0.0005 to 0.0018	0	0	0	0	0.0005 to 0.0018
<i>P</i> (results same)		0.000	0.000	0.000	0.000	

Table 4. Comparison of coefficient magnitudes between model runs and observational runs. Middle four columns show results of MM07 regression (using CRU surface data and UAH satellite data), and CRU3v/RSS pairing, compared against GISS and all-GCM ensembles respectively. Entry is “0” if coefficient falls outside 95% confidence interval (CI) for coefficient estimated using indicated model data. Second column shows 95% CI around socioeconomic coefficients from GISS regression, seventh column shows same for all-GCM ensemble mean. Last row shows *P* value of chi-squared test of parameter equivalence between regressions.

Mean Trend:	MM07 method using CRU & UAH	MM07 method using CRU3v & RSS4	MM07 method using GISS data	MM07 method using all-GCM data
Surface	0.302	0.303	0.195	0.231
Troposphere	0.232	0.236	0.222	0.234
Filtered surface	0.166	0.195	0.222	0.224

Table 5: Filtering results using UAH, RSS4, GISS and all-GCM data. Each table entry shows the mean trend in the global sample. The Surface trend is the mean trend in the surface observations for that column, likewise for the Troposphere trend. The second column shows the original MM07 results. The third column shows the results from swapping in CRU3v and RSS data, the fourth column shows the results from swapping in the GISS ensemble mean simulations from S09 and the final column shows the results using the all-GCM average.

Variable	DIFF	CRU3/RSS4
Trop	0.9402 (8.39)	0.9517 (8.70)
slp	0.0040 (1.04)	0.0059 (1.64)
dry	1.0535 (0.22)	4.2539 (0.99)
dslp	-0.0010 (0.21)	-0.0041 (0.97)
water	0.0120 (0.57)	-0.0164 (0.66)
abslat	0.0026 (2.21)	0.0042 (3.17)
g	0.0487 (3.36)	0.0401 (2.58)
e	-0.0020 (2.52)	-0.0025 (3.78)
p	0.3477 (2.40)	0.2081 (1.55)
m	0.3904 (2.26)	0.2589 (1.51)
y	-0.2830 (2.04)	-0.1941 (1.42)
c	0.0051 (2.61)	0.0071 (3.58)
Constant	-5.1027 (1.26)	-5.8494 (1.60)
P(all)	0.000	0.000
<i>N</i>	428	428
<i>R</i> ²	0.50	0.54

Table 6: Results as for Table 3 using differences between observed and GCM-generate trends in each gridcell. For notation see notes to Table 3. 3rd column equals column 8 from Table 3 for comparison purposes. CRU3 denotes CRUTEMv3 surface trends, RSS4 denotes RSS 5x5 degree tropospheric trends. Absolute *t* statistics in parentheses. Bold denotes significant at 5%.

	ORIG (UAH)	UAH4+SAC	RSS4+SAC
slp	-0.0001 (0.03)	0.0044 1.67	0.0048 (2.53)
dry	-11.3879 (3.01)	-7.0356 (2.22)	-5.7135 (2.21)
dslp	0.0112 (3.02)	0.0069 (2.23)	0.0056 (2.22)
water	0.0307 (1.37)	0.0146 (1.54)	0.0100 (1.09)
abslat	0.0064 (5.39)	0.0046 (6.01)	0.0014 (2.10)
g	0.0424 (1.81)	0.0129 (1.04)	0.0132 (1.11)
e	-0.0004 (0.58)	0.0006 (2.12)	0.0001 (0.43)
x	-0.0114 (3.22)	-0.0035 (2.00)	-0.0034 (2.55)
p	0.1845 (1.42)	0.0782 (1.59)	0.1557 (3.21)
m	0.2596 (1.55)	0.0546 (0.85)	0.0786 (1.29)
y	-0.2069 (1.59)	-0.0279 (0.57)	-0.0541 (1.19)
c	0.0036 (2.11)	0.0015 (1.45)	0.0006 (0.64)
Const	0.0682 (0.02)	-6.3602 (1.84)	-5.8793 (2.49)
Lambda		0.9953 (211.6)	0.9942 (171.58)
N	440	440	440
LL	269.02	507.64	514.79
P(All)	0.010	0.007	0.002

Table 7: Results from replacing the dependent variable with the tropospheric trends and regressing on the remaining variables. First column: from MM07, UAH data, no correction for SAC. 2nd and 3rd columns: RSS and UAH data respectively, SAC correction applied. Absolute t statistics underneath coefficients, **bold** denotes significant at 5%.

APPENDIX:

The all-GCM data was constructed using 55 runs from 22 GCMs used in the IPCC (2007) report. The archive is at <http://www-pcmdi.llnl.gov>. HadCM3 wasn't used because it did not represent its data in the required IPCC pressure levels. MUIB ECHO G wasn't processed because no atmospheric temperature data was available, thus synthetic MSU brightness temperatures couldn't be calculated.

The calculation of brightness temperature was done using the same algorithm and weighting functions implemented in Santer et al. (2008). Trend fields in degrees C/decade for the surface and lower tropospheric temperature were calculated as follows.

1. Extract all data from Jan 1979 - Dec 2002.
2. Compute the climatology for the same period.
3. Subtract the climatology from the original data.
4. Calculate the trend field for each grid point only if all the data points are valid.
5. Collect only the trends that correspond to the MM07 set of lat/lon coordinates.
6. Multiply the resulting annual trends by 10 to obtain decadal trends

There was no missing data for the surface temperature variable, but there was some missing data in some runs for the TLT brightness temperature. This is because the models originally didn't represent the atmospheric temperature on the same set of pressure levels that the IPCC mandated. Interpolation was required and this resulted in some missing data points in the lower atmosphere. To calculate the brightness temperature, the atmospheric temperature profile was multiplied by a set of weights specific to a given atmospheric layer (TLT, TMT, TLS). The weighted temperatures were then added up and divided by the sum of weights that correspond to non-missing temperature values. If this

total weight did not equal or exceed 0.5 or 50%, then the brightness temperature at that grid point was flagged as missing.