Socioeconomic Patterns in Climate Data

Ross McKitrick
(corresponding author)
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

December 2010

Summary

To generate a climate data set, temperature data collected at the Earth's surface must be adjusted to remove non-climatic effects such as urbanization and measurement discontinuities. Some studies have shown that the post-1980 spatial pattern of temperature trends over land in prominent climate data sets is strongly correlated with the spatial pattern of socioeconomic development, implying that the adjustments are inadequate, leaving a residual warm bias. This evidence has been disputed on three grounds: spatial autocorrelation of the temperature field undermines significance of test results; counterfactual experiments using model-generated data suggest such correlations have an innocuous interpretation; and different satellite covariates yield unstable results. Somewhat surprisingly, these claims have not been put into a coherent framework for the purpose of statistical testing. We combine economic and climatological data sets from various teams with trend estimates from global climate models and we use spatial regressions to test the competing hypotheses. Overall we find that the evidence for contamination of climatic data is robust across numerous data sets, it is not undermined by controlling for spatial autocorrelation, and the patterns are not explained by climate models. Consequently we conclude that important data products used for the analysis of climate change over global land surfaces may be contaminated with socioeconomic patterns related to urbanization and other socioeconomic processes.

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

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

Socioeconomic Patterns in Climate Data

1 Introduction

1.1 Background

Temperatures are monitored at thousands of sites around the world, for a variety of purposes: weather analysis and forecasting, tracking aviation conditions at airports, assessing the risk of smog formation in urban areas, etc. In most cases, the actual temperature recorded at a specific site, such as an airport, is the measure sought. But for the purpose of measuring long term climatic trends, such as for use in studies of global warming, raw temperature data are not appropriate. This is because, over time, local mean temperatures can change for reasons unrelated to the climate. If we consider a monitoring site in a region which is known to have experienced no climatic warming, there might still be a trend in the temperature record if a nearby city expanded and encompassed the site, or the type of thermometer changed and the equipment was not carefully calibrated, or the time of day at which readings are taken was changed, etc. These local changes (called "inhomogeneities") must be filtered out to reveal, if possible, the underlying trends attributable solely to long-term climatic changes.

The ideal monitoring site would be one which has undergone no changes to the local land surface since the start of modern temperature monitoring (roughly, since the late 1800s), and which has not experienced any local air pollution, which can cause small but measurable local cooling or warming, and which has had the same equipment and recording methodology across time. Because there are few such sites in the world, long temperature records are typically taken from sites that fail one or more of these conditions. Hence the records must be modified in some way to reveal the climatic trend. The dangers of using raw temperature data for climate analysis has long been noted. In 1953, commenting on the growing use of weather records for measuring climate change, J. Mitchell Jr. [19] pointed to various sources of bias and cautioned: "The problem remains one of determining what part of a given temperature trend is climatically real and what part the result of observational difficulties and of artificial modification of the local environment."

This paper is concerned with the problem Mitchell, and many others since then, have discussed. We will examine the claims made by providers and users of climate data to the effect that the problems of non-climatic influences on climate data sets have been taken into account and rectified. We will show that this assumption has been tested and found unlikely to be true. We will then focus on some recent counter-arguments that defend the standard interpretation of climatic data, and we will show that those arguments are not supported in formal testing. Our conclusion will be that climate data products likely do not measure what climate scientists think they measure; that they are likely contaminated with patterns related to socioeconomic development and industrialization, and that these influences render them biased for the purpose of measuring global warming and interpreting its causes.

1.2 Climate Data Adjustments

The Climatic Research Unit (CRU) at the University of East Anglia disseminates the world's most widely-used surface temperature data sets, including those used in reports of the influential Intergovernmental Panel on Climate Change (IPCC, [6]). The CRU web page (http://www.cru.uea.ac.uk/cru/data/hrg/) presents two data compilations: CRU TS and CRUTEM. The TS series are not subject to adjustments for non-climatic influences, and users are thus cautioned not to use them for climate analysis (see http://www.cru.uea.ac.uk/cru/data/hrg/timm/grid/ts-advice.html). The TS website notes that, for climate analysis, it is necessary 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 anthropogenic trends." Readers are referred to Brohan et al. [3], Jones and Moberg [8] and Jones et al. [9] for explanations of the specific adjustments.

Brohan et al. [3] does not explain the adjustments, it mainly focuses on the claim that any biases are very small. [3] Section 2.3 says of their temperature archive that adjustments were made to raw temperature series but the originals were not retained, so it is now impossible to say how large the adjustments were. [3] assumes any inhomogeneity uncertainties are symmetric around zero (p. 6). [3] Section 2.3.3 states that to properly adjust the data for urbanization bias 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 instead invoke the assumption that the bias is no larger than 0.006 degrees per decade. [8] likewise offers little information about the data adjustments. They discuss combining multiple site records into a single series, but do not discuss removing non-climatic contamination. Moreover, the article points out (page 208) that it is difficult to say what homogeneity adjustments have been applied to the raw data since the original sources do not always include this information. [9] emphasizes that non-climatic influences must be corrected (Section 2, p. 37) for the data to be useful for climatic research. But the part of the paper that outlines the adjustments consists of only three paragraphs in Section 2.1, none of which explains the procedures. The only explanatory statement is the following (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])."

The two reports cited ("Jones et al. 1985, 1986c") are technical reports submitted to the US Department of Energy some 25 years ago. They only cover data sets ending in the early 1980s, whereas the data now under dispute is the post-1979 interval. Even if the adjustments were adequate in the pre-1980 interval it is likely impossible to have estimated 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) are inappropriate for climatic analysis, and refers users to the CRUTEM products. Yet the accompanying documentation does not appear to explain the adjustments made or the grounds for claiming the CRUTEM products are reliable for climate research purposes.

Nevertheless, the assumption that the adjustments are adequate is widely held. For example, Jun et al. [10] used surface climate data to test some properties of climate models. While they are aware of the many faults of the underlying data, they dispensed with them as follows (p. 935):

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

In its 4th Assessment Report, as in the previous three, the IPCC also claimed that their data are adjusted to remove non-climatic contamination. This forms an essential assumption behind all its key conclusions. Global temperature trends were presented in Table 3.2 on page 243 of [7] (Working Group I). The accompanying text (page 242) states that the data uncertainties "take into account" biases due to urbanization. The Executive Summary to the chapter (page 237) asserts that "…the very real but local [urbanization] effects are avoided or accounted for in the data sets used." The influential Summary for Policymakers stated:

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

The 0.006°C is referenced back to [3], where it is merely an assumption about the standard error, not the size of the trend bias itself. IPCC Chapter 9 provides the summary of evidence attributing warming to greenhouse gases. The problem of surface data contamination is set aside as follows ([7] p. 693):

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

The citation to IPCC [6] Chapter 3 is uninformative. That chapter does not describe the data adjustments and only briefly mentions two studies (McKitrick and Michaels [15] and de Laat and Maurellis [5]) that provided evidence of strong correlations between the spatial pattern of warming trends and the spatial pattern of industrialization, a correlation not predicted by climate models and an indication of data contamination. However the IPCC ([7], p. 244) dismisses these findings on the basis of an unsupported claim that the correlations results were spurious artifacts of atmospheric circulation patterns. McKitrick [14] refuted that claim by showing that controlling for atmospheric circulation effects does not affect the correlations.

1.3 Evidence of Data Problems

[15] and [16], collectively herein MM04, and [17], herein MM07, tested the adequacy of the data adjustments 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 [22] (herein S09) marshaled four arguments to defend the view that, notwithstanding the evidence in MM04 and MM07, as well as the de Laat and Maurellis [5] findings, CRUTEM data are unbiased. First, he noted that an overall warming trend is observed in numerous post-1980 data sets. 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 [18], denoted RSS) rather than from the University of Alabama-Huntsville ([25], 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.

Regarding the first point, it is true that numerous different data sets show an upward trend after 1980. But the surface data show a relatively large trend compared to data collected from satellites [4], which is consistent with the view that there is an upward bias in the land-based data. Turning to the latter three points, it is noteworthy that they are all statistical in nature, yet S09 did not conduct any statistical testing. In this paper we develop a regression-based framework that allows direct hypothesis testing to shed some light on 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. [14] shows that SAC was not detected in the residuals of the MM07 regression model, although that estimation was not maximum likelihood, and a revised version shown in Section 3 below shows significant SAC in the MM07 residuals. However on a larger group of updated data sets we find the dependent variable is spatially autocorrelated but the regression residuals are not, providing evidence that the explanatory model is well-specified. By contrast, when using data generated by GISS-E and other climate models (General Circulation Models, or GCM's), spatial autocorrelation is observed in both the dependent variable and the residuals, and the effect is stronger in the residuals than in the dependent variable, 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 largely eliminates this contrast. 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.

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, are of opposite signs and different magnitudes compared to 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 does not affect the mean surface trend in the model-generated data. 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 an ensemble mean of a large suite of GCM's. 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 influence appear to be an essential component of a well-specified model of the spatial trend pattern in climate data over land. The coefficient pattern on observational data differs in both sign and magnitude 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 to be 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.

1.3 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.

MM07 estimated the regression equation

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

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 ^{o}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 by MM07 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 UAH lower tropospheric record of Spencer and Christy [25].

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 RSS [18]. S09 provides trends derived from the CRUTEM2v [8] and CRUTEM3v data sets [3]. 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).

---

[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. The combined data sets for this paper are available in the SI.

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%.

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 ([4] Sct. 2.5.3) and [23]. 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 [7] report (see Appendix). These data are referred to herein as the all-GCM mean, and the trend vectors are denoted as MSM (surface) and MTM (troposphere).

The average GISS-E land surface trend is 0.20 ^{o}C/decade, well below the reported trend of 0.30 ^{o}C/decade in the CRU3v compilation. The range of trends over land in the all-GCM mean is 0.07 to 0.56 ^{o}C/decade with a mean of 0.23 ^{o}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 results can be run with no spatial autocorrelation terms, or with corrections on either the error or lagged dependent variables, 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 UAH4 and RSS4 rather than UAH and RSS.

2 Robustness across multiple data sets: the non-SAC case

2.1 Observational data sets

We begin this section by looking at whether the MM07 results were unique to the particular data configuration used therein. Table 2 presents the regression coefficients for Equation (1) using 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 final two columns use the GISS-E and all-GCM ensemble means.

Estimations throughout this paper were done using STATA version 8.2, running on a Dell Studio laptop with an Intel i7 quad core 64-bit processor. All data and code are available in the supplementary information.

Variable	MM07	C2/UAH	C2/RSS	C3/UAH	C3/RSS	C3/UAH4	C3/RSS4	C3/RSS4x	GISS	GCM
trop	**0.8631**	**0.8323**	**0.9872**	**0.8146**	**0.9627**	**0.9330**	**0.9538**	**0.9263**	**1.2212**	**1.4649**
	(8.62)	(8.58)	(14.36)	(8.80)	(13.84)	(11.03)	(8.77)	(7.49)	(14.73)	(10.87)
slp	0.0044	0.0043	0.0024	0.0058	0.0039	0.0071	0.0059	0.0043	-0.0017	-0.0011
	(1.02)	(1.03)	(0.66)	(1.43)	(1.11)	(1.92)	(1.67)	(1.33)	(1.06)	(1.26)
dry	0.5704	1.6771	0.1986	1.5010	0.2125	4.5837	4.3175	2.8723	1.1352	**2.5261**
	(0.10)	(0.33)	(0.04)	(0.30)	(0.04)	(1.01)	(1.01)	(0.52)	(0.54)	(2.04)
dslp	-0.0005	-0.0016	-0.0001	-0.0014	-0.0001	-0.0044	-0.0042	-0.0027	-0.0011	**-0.0024**
	(0.09)	(0.31)	(0.03)	(0.28)	(0.03)	(0.99)	(0.99)	(0.51)	(0.55)	(2.01)
water	-0.0289	-0.0293	-0.0169	-0.0240	-0.0117	-0.0278	-0.0161	-0.0155	**-0.0501**	**-0.0229**
	(1.37)	(1.37)	(0.68)	(1.06)	(0.46)	(1.29)	(0.66)	(0.56)	(5.33)	(3.30)
abslat	0.0006	0.0010	**0.0028**	0.0014	**0.0033**	0.0005	**0.0041**	**0.0040**	**0.0010**	**0.0022**
	(0.51)	(0.82)	(2.28)	(1.22)	(2.65)	(0.54)	(3.00)	(2.54)	(1.97)	(4.90)
g	**0.0432**	**0.0450**	**0.0444**	**0.0449**	**0.0446**	**0.0383**	**0.0401**	0.0383	$\boxed{-0.0107}$	$\boxed{-0.0076}$
	(3.36)	(3.66)	(3.03)	(4.01)	(3.27)	(3.44)	(2.58)	(1.78)	(1.53)	(1.76)
e	**-0.0027**	**-0.0026**	**-0.0025**	**-0.0029**	**-0.0028**	**-0.0028**	**-0.0025**	**-0.0031**	$\boxed{\mathbf{0.0005}}$	$\boxed{0.0000}$
	(5.14)	(5.32)	(4.73)	(4.41)	(4.10)	(4.32)	(3.73)	(4.75)	(2.66)	(0.23)
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**	$\boxed{-0.0919}$	$\boxed{-0.0118}$
	(2.72)	(2.67)	(1.04)	(2.52)	(0.97)	(2.18)	(1.55)	(2.71)	(1.54)	(0.38)
m	**0.4093**	**0.3844**	0.2663	**0.3732**	0.2578	0.2810	0.2595	**0.5553**	$\boxed{-0.1191}$	$\boxed{-0.0444}$
	(2.39)	(2.33)	(1.52)	(2.19)	(1.42)	(1.68)	(1.50)	(3.00)	(1.55)	(1.18)
y	**-0.3047**	**-0.2839**	-0.2160	**-0.2804**	-0.2139	-0.2178	-0.1944	**-0.4300**	$\boxed{0.0791}$	$\boxed{0.0320}$
	(2.22)	(2.15)	(1.55)	(2.06)	(1.48)	(1.63)	(1.42)	(2.91)	(1.33)	(1.07)
c	**0.0062**	**0.0060**	**0.0076**	**0.0063**	**0.0079**	**0.0057**	**0.0071**	0.0084	0.0004	0.0010
	(3.45)	(3.46)	(3.77)	(3.38)	(3.62)	(2.93)	(3.58)	(1.92)	(0.75)	(2.13)
Constant	-4.2081	-4.1522	-2.2291	-5.5546	-3.7704	-6.9317	-5.8808	-5.3457	$\boxed{1.5622}$	$\boxed{0.9162}$
	(0.96)	(0.97)	(0.60)	(1.36)	(1.05)	(1.84)	(1.63)	(1.23)	(0.99)	(1.03)
***P*(H:g—c=0)**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.001**
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	0.68	0.77
LLF	139.22	151.75	145.50	142.64	135.83	157.10	147.54	135.35	584.97	819.27

Table 2: Regression results from a group of data configurations. MM07 shows results from original data set (McKitrick and Michaels 2007). Next seven 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 two columns show regressions using GISS-E and all-GCM-generated surface and tropospheric data. Numbers in parentheses are absolute *t*-statistics. All columns reflect corrections for clustered errors and heteroskedasticity. **Bold** denotes significant at 95% confidence. A box around a coefficient in the final two columns indicates that the sign on model-generated data was opposite to that on estimates from all observational data sets. *P*(H:g—c=0) is prob value of test that coefficients g–c are jointly zero. Variable *x* is dropped in GISS-E and GCM regressions since there are no missing surface values.

The socioeconomic coefficient estimates are quite similar across the observational columns (2 though 9). Use of CRU3v data does not yield much of a change in socioeconomic coefficients compared to CRU2v and CRU. Use of RSS data rather than UAH data yields smaller and less significant coefficients, 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. 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 largest drop in the coefficient magnitudes is associated with using the CRU3v-RSS4 pair, yet the effect is apparently due to a relatively small number of outliers. The C3/RSS4x column repeats the CRU3v/RSS4 results with outliers removed. The Ordinary Least Squares "hat matrix" was evaluated, and an observation was flagged as an outlier if it exceeded twice the mean diagonal element of the hat matrix (see [12] 424—426). In this case 26 observations were removed. This had the effect not only of bringing the CRU3v/RSS4 results into line with the results computed with other data configurations, but indeed yielding the largest socioeconomic coefficients of any data combination. Figure 1 shows the spatial locations of the outlier observations. There is some clustering in the vicinity of the UK and Thailand, but the most telling feature is that about half are from small oceanic islands. For these it is difficult to impute accurate socioeconomic data (MM07 used the relevant data for the governing country) so these locations are most likely to be subject to measurement error. Hence their removal from the data set is a reasonable robustness check. Overall, Table 2 shows that the socioeconomic effects seem to be a robust feature of the data and are not unduly dependent on selecting one particular data product.

2.2 Model-generated data sets

S09 hypothesized that these effects arise from a lucky 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 GISSES and GISSET 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 2, 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. Coefficients in Table 2 with a box drawn around them indicate that the estimates on both model-generated data sets take the opposite sign to that estimated on all configurations of observed data. A box is drawn around all socioeconomic coefficients except coal (*c*). Additionally, the parameter values are much smaller and in almost all cases are insignificant. Consequently we find there is no similarity between the socioeconomic coefficients estimated on model-generated data and those estimated on observed data, or in other words, the models predict the opposite pattern to that observed in the data.

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 2. In the next section this will be shown after re-estimating the model using a correction for spatial

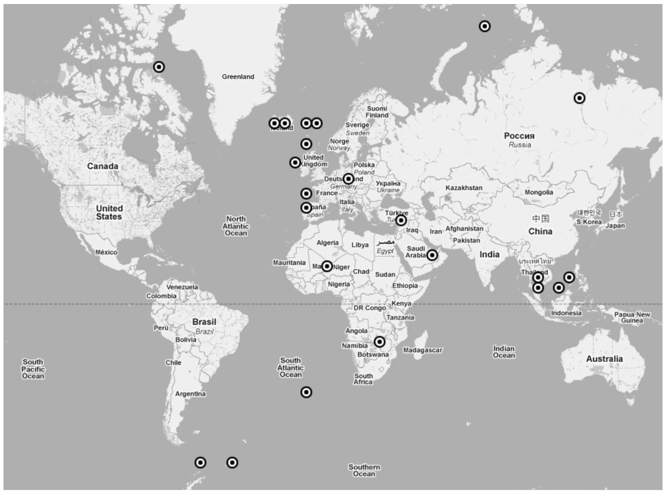

Figure 1: Locations of outliers referenced in Table 2 column 9.

autocorrelation. Anticipating the findings, for none of the socioeconomic coefficients does the 95% Confidence Interval estimated on model-generated data encompass the coefficients estimated on observed data. Consequently the null hypothesis as stated by Schmidt (that there is no contamination) *is* rejected.

3 Spatial autocorrelation of the trend field

3.1 Testing framework

The analysis in the previous section ignores the issue of spatial dependence, so we must now reconsider the estimating equations taking it into account. Both S09 and Benestadt [2] 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. 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, in other words 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

$$\mathbf{T} = \mathbf{Xb} + u \qquad (2)$$

where **T** is a 440x1 vector of temperature trends in each of 440 surface grid cells, **X** is a 440x*k* matrix of climatic and socioeconomic covariates, **b** is a *k*x1 vector of least-squares slope coefficients and *u* is a 440x1 residual vector. Spatial autocorrelation in the residual vector can be modeled using

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

where λ is the autocorrelation coefficient, **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 [20]). The rows of **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. [1] point out that if the alternative model allows for possible spatial dependence of **T**, i.e.

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

where **Z** is a matrix of spatial weights for **T** which may not be identical to **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. In this case ϕ is a nuisance parameter. The covariance between the score of ϕ and the score of λ conditional on **b** must be zero for the conventional Lagrange Multiplier (LM) test of H_0: $\lambda = 0$ to follow a central χ_1^2 distribution. If the conditional covariance is not zero, the test follows a non-central $\chi^2(\delta)$ distribution, where the non-centrality parameter δ is proportional to the conditional covariance between the scores. [1] derives an LM test of $\lambda = 0$ robust to nonzero value of the nuisance parameter ϕ by applying a correction to the score terms. It 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.

Autocorrelation can affect either the dependent variable or the residuals or both. Florax et al. [6] discuss a sequential testing and estimation regime for models of the form combining equations (3) and (4), namely

$$\mathbf{T} = \phi \mathbf{ZT} + \mathbf{Xb} + \lambda \mathbf{W} u + e \qquad (5).$$

Here ϕ is the spatial lag and λ is the spatial error term. If neither one is significant then a least-squares model with no spatial autocorrelation term can be used. If only one of them is significant, then (5) should be used retaining only the significant lag term. If both are significant then the one that has a higher significance level should be chosen, implying either the spatial lag model (4) or the spatial error model (2, 3).

Hypothesis tests, and any subsequent parameter estimations, are conditional on the assumed form of the spatial weights matrix **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}. The weighting function is $g_{ij}^{-\mu}$ where μ determines the rate at which the relative influence of one cell on adjacent cells declines. The function was estimated using a grid search to find the maximum-likelihood value. For observational data groupings the likelihood function was maximized at values of μ between 2.5 and 2.7, whereas for the GISS-E and all-GCM model-generated data, the likelihood function was maximized at μ values of 3.2 or 3.0 respectively. The optimal exponent values are listed in Table 3.

3.2 Spatial Autocorrelation Testing Results

The Stata command 'spatwmat' was used to generate the row-standardized weights and eigenvalues, then the command 'spatdiag' was used to generate the test scores. Table 3 presents the results of SAC hypothesis tests on both the dependent variable and residuals for seven different model configurations. The second column (exponent) reports the maximum likelihood value of μ. The next column reports the robust LM score for a null hypothesis of no spatial dependence on the dependent variable, while the final column reports the robust LM score for the test on the residuals. In each column the corresponding p value is shown in parentheses.

The first row 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). In this case the robust LM score is significant for both the dependent variable and the residuals, but much more so for the dependent variable, indicating that a spatial lag model is appropriate. The second and third rows show the test scores using CRU3v surface data and either UAH4 or RSS4 tropospheric series. In both cases the dependent variable lag is significant while the residual lag term is insignificant. Again this indicates that the spatial lag model is appropriate, and also indicates that the regression model is well-specified in the sense that the SAC is removed from the error terms.

The next two rows report the results after substituting in model-generated data on both sides of Equation (1). The results change in an interesting way. The no-SAC test in the residuals is now strongly rejected, and the ranking of the significance also changes, so that the appropriate estimation model is now the spatial error form rather than the spatial lag. This is suggestive of a deficiency in the specification of (1), whereby the terms on the right hand side fail to explain the spatial lag pattern for model-generated data the way they do when observational data are used.

The final two rows refer to a test applied in the next section and will be discussed below.

To summarize, in regressions using observational temperature data, the right-hand side terms in equation (1) appear to remove the SAC from the regression residuals, leaving an independent error term. But the same variables do not remove the SAC from the residuals when data generated by a climate model are used. This is evidence that socioeconomic data are a necessary component of a well-specified explanatory model of surface temperature trends in the CRU data sets, and this effect arises from factors not accounted for by the anthropogenic and natural forcings coded into the GCMs. It also indicates that correction for SAC on either the dependent variable or the residuals is required in the regressions, hence we will re-do the above regressions applying the appropriate lag models.

Estimation Model	Weighting Function	Dependent Variable	Residuals
CRU: UAH + MM07	$g_{ij}^{-2.6}$	**22.231** (0.000)	**4.288** (0.038)
CRU3v: UAH4 + MM07	$g_{ij}^{-2.7}$	**9.970** (0.002)	1.974 (0.160)
CRU3v: RSS4 + MM07	$g_{ij}^{-2.5}$	**13.429** (0.000)	1.297 (0.255)
GISSES: GISSET + MM07	$g_{ij}^{-3.2}$	**41.546** (0.000)	**116.170** (0.000)
MSM: MTM + MM07	$g_{ij}^{-3.0}$	**4.376** (0.036)	**56.991** (0.000)
CRU3v–MSM: UAH4 + MM07	$g_{ij}^{-2.7}$	**7.665** (0.006)	1.945 (0.163)
CRU3v–MSM: RSS4 + MM07	$g_{ij}^{-2.7}$	**11.828** (0.001)	2.503 (0.114)

TABLE 3. Spatial autocorrelation tests for regression models. Estimation model described as [surface measure]: [tropospheric measure] + MM07. Dependent variable is 1979-2002 trend in gridded surface data, using CRU series as in McKitrick and Michaels (2007), or CRU3v update from Brohan et al [3], 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 McKitrick and Michaels (2007) model, namely *SLP* through *c* in Equation (1). Weighting Function refers to form of spatial dependence, where g_{ij} denotes great circle distance between grid cells *i* and *j* and the exponent is the result of the gridsearch routine described in the text.. Third and fourth columns: each entry shows robust LM autocorrelation parameter and associated *P* value of hypothesis that it equals zero. **Bold** denotes hypothesis rejected at 5% significance.

3.3 Regression results with SAC controls

Table 4 presents results in the same format as Table 2. Five new sets of parameter estimates are shown, along with the MM07 results in column 2 for comparison. The "+S" in the column heading denotes the addition of either a spatial lag or spatial error term, as indicated. Column 3 augments the MM07 result with a spatial lag term. Columns 4 and 5 show the results using CRU3v surface data and UAH4 or RSS4 tropospheric data, respectively, and the last two columns show the results using the GISS-E and the all-GCM average results, respectively.

The pattern observed in Table 2, whereby the observations yield one type of result while the model-generated data yield another, shows up again. The socioeconomic coefficients are relatively large and significant for all observational data configurations, but vanish in the model data columns. The row denoted *P*(H:g—c=0) reports the *P* value of the joint significance test on the socioeconomic effects. They are jointly highly significant in the observational data columns, albeit less so than in the MM07 case where no spatial lag term was included. In the model data columns the socioeconomic effects are jointly insignificant.

Variable	MM07	MM07+S	C3/U4+S	C3/R4+S	GISS+S	GCM+S
trop	**0.8631**	**0.6058**	**0.7269**	**0.6976**	**1.7856**	**1.5859**
	(8.62)	(7.63)	(9.04)	(8.05)	(10.84)	(11.07)
slp	0.0044	0.0037	**0.0061**	0.0051	**-0.0044**	**-0.0024**
	(1.02)	(1.35)	(2.16)	(1.83)	(2.57)	(2.52)
dry	0.5704	1.7438	4.7378	4.7540	-0.1545	1.1861
	(0.10)	(0.50)	(1.37)	(1.36)	(0.08)	(1.00)
dslp	-0.0005	-0.0017	-0.0046	-0.0046	0.0001	-0.0011
	(0.09)	(0.48)	(1.35)	(1.34)	(0.08)	(0.99)
water	-0.0289	-0.0287	-0.0287	-0.0209	**-0.0179**	**-0.0129**
	(1.37)	(1.53)	(1.56)	(1.12)	(2.58)	(4.18)
abslat	0.0006	0.0004	0.0004	**0.0030**	**0.0016**	**0.0030**
	(0.51)	(0.45)	(0.45)	(3.06)	(2.84)	(7.27)
g	**0.0432**	**0.0294**	**0.0298**	**0.0285**	0.0005	0.0006
	(3.36)	(2.25)	(2.44)	(2.57)	(0.08)	(0.19)
e	**-0.0027**	**-0.0015**	**-0.0019**	**-0.0014**	0.0002	-0.0001
	(5.14)	(2.67)	(3.58)	(2.73)	(0.71)	(0.97)
x	0.0041	0.0023	0.0002	-0.0009		
	(1.66)	(0.78)	(0.07)	(0.29)		
p	**0.3839**	**0.2967**	**0.2403**	0.1682	-0.0319	0.0230
	(2.72)	(2.55)	(2.07)	(1.44)	(0.60)	(1.14)
m	**0.4093**	0.2472	0.1775	0.1331	-0.0386	0.0262
	(2.39)	(1.91)	(1.36)	(1.02)	(0.74)	(1.00)
y	**-0.3047**	-0.1826	-0.1386	-0.0989	0.0203	-0.0239
	(2.22)	(1.79)	(1.35)	(0.97)	(0.54)	(1.25)
c	**0.0062**	**0.0037**	**0.0040**	**0.0045**	0.0002	0.0011
	(3.45)	(2.04)	(2.19)	(2.43)	(0.29)	(2.92)
Constant	-4.2081	-3.6601	**-6.0249**	-5.1528	**4.2465**	**2.2039**
	(0.96)	(1.31)	(2.11)	(1.83)	(2.45)	(2.32)
Spatial lag		**0.4204**	**0.3054**	**0.3996**		
		(5.40)	(3.84)	(5.11)		
Spatial error					**0.8463**	**0.8882**
					(16.74)	(26.34)
***P*(H:g—c=0)**	**0.0000**	**0.0037**	**0.0005**	**0.0050**	0.5937	0.0709
N	440	440	428	428	440	440
R^2	0.53	0.57	0.58	0.57	0.64	0.73
LLF	139.22	154.58	164.98	159.90	660.27	961.52

Table 4: Regression results for five data configurations applying corrections for spatial dependence. Notation as for Table 2. All columns reflect corrections for spatial autocorrelation, clustered errors and heteroskedasticity. Spatial lag: model augmented with spatially lagged dependent variable. Spatial error: model augmented with spatially lagged residual (see text for explanation). **Bold** denotes significant at 95% confidence. *P*(H:g—c=0) is prob value of test that coefficients g–c are jointly zero. R^2 is squared correlation between observed and predicted values. Variable *x* is dropped in GISS-E and GCM regressions since there are no missing surface values.

Note that the coefficients in Table 4 from the spatial lag model are somewhat smaller than the corresponding coefficients in Table 2 due to the respecification. Equation (4) can be rewritten $\mathbf{T} = \mathbf{JXb} + \mathbf{v}$ where $\mathbf{J} = (\mathbf{I} - \phi\mathbf{Z})^{-1}$, $\mathbf{I}$ is an identity matrix and $\mathbf{v} = \mathbf{J}e$. Then the coefficient estimates from (4) are $(\mathbf{X}^T\mathbf{J}^T\mathbf{JX})^{-1}\mathbf{X}^T\mathbf{J}^T\mathbf{T}$, which differ from those of (2), with or without (3), by the extent to which $\mathbf{J}$ differs from $\mathbf{I}$. The distance weighting matrix $\mathbf{Z}$ has ones along the diagonal and off-diagonal terms that go rapidly to zero, hence the influence on **b** will be, to a simple approximation, $1/(1-\hat{\phi})$. For the lag estimates shown in Table 4, this implies they should be increased by a factor of approximately 1.4 to 1.6 to compare against magnitudes with Table 2.

Another interesting contrast concerns the tropospheric coefficient (first row). The usual expectation from climate modeling studies is that the trend aloft will be slightly stronger than that at the surface, but as shown in Table 1, a univariate comparison shows the mean trend in the lower troposphere (~0.23) is less than that at the surface (~0.3). This issue has been widely noted: see [4] and [11] for instance. However, once the socioeconomic and other effects are controlled in a multivariate regression, the coefficient relating the tropospheric trend to the surface goes below unity, indicating a slightly amplified trend aloft compared to the surface, consistent with expectations. The mean tropospheric trend in the models is slightly above that at the surface (see Table 2) but in the regression model the relationship reverses, so that the tropospheric coefficient goes above unity, which is not consistent with expectations.

3.4 Do GCM's predict the observed temperature-industrialization correlation pattern?

Returning to Schmidt's [22] argument that the distribution around coefficients estimated on model-generated needs to encompass those estimated on observed data, Table 5 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. An indicator takes a value of 1 if the coefficient estimated on observed data is within the comparison CI or 0 if it is not. As is shown, all indicator variables take a zero value, indicating that in no case do the coefficients estimated on observed data fall within the distribution of coefficients predicted by the models. Table 5 reports results using SAC-corrected estimates, but the results would be the same using non SAC-corrected results. Hence the null hypothesis in the form stated by S09, namely parameter overlap implying an absence of contamination, is rejected.

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-E and all-GCM data. The MM07 method sets the contaminating influences to zero and assumes each country has equivalent measurement resources as the United States. Since the climate model does not contain any contaminating processes or quality-control variations 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 6 shows the results. The first row repeats the findings in MM07 for comparison. The second column shows the mean surface trend, the third column shows the troposphere trend (using either UAH or RSS data, as indicated for that row), the fourth column shows the mean filtered surface trend and the final column shows the mean filtered surface trend weighting each gridcell by the cosine of latitude to adjust for declining gridcell areas in higher latitudes. In MM07 the filtering step reduced the surface trend by about one-third. The next two rows apply the method to the CRU3v and UAH4/RSS4 data, with similar results, albeit slightly less using RSS data. The final two rows show that when applied to model-generated data the adjustments are, as expected, quite small, and actually raise the GISS-E trend slightly rather than reducing it. 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.

Variable	GISS 95% CI	CRU3v/UAH4 Compared to GISS	GCM	CRU3v/RSS4 Compared to GISS	GCM	GCM 95% CI
g	-0.0135 to 0.0146	0	0	0	0	-0.0059 to 0.0071
e	-0.0004 to 0.0008	0	0	0	0	-0.0004 to 0.0001
p	-0.1389 to 0.0752	0	0	0	0	-0.0174 to 0.0634
m	-0.1435 to 0.0663	0	0	0	0	-0.0260 to 0.0785
y	-0.0542 to 0.0949	0	0	0	0	-0.0621 to 0.0143
c	-0.0011 to 0.0015	0	0	0	0	0.0003 to 0.0019

Table 5. Comparison of coefficient magnitudes between model runs and observational runs. Middle four columns show results of 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 estimated on observed data 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.

Finally, to investigate the ability of the all-GCM ensemble to explain the surface trend pattern we took the differences between the observed surface trends (CRU3v) and the all-GCM mean surface trends and examined if the differences can be explained by the socioeconomic variables in equation (1). To the extent that the GCM ensemble mean explains 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. Table 7 shows the result of 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. Spatial lag terms are included in all regressions. The second column (C3/U4+S) shows the results using CRU3v as the dependent variable and UAH4 as the tropospheric measure, i.e. the same as Table 4 column 4. Column 3 then replaces the dependent variable with the observation-model differences. The socioeconomic coefficients retain their size and significance levels. The next two columns repeat the experiment using RSS4 data. Again the socioeconomic coefficients retain their size and joint significance. This 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.

MM07 method using:	Surface	Troposphere	Filtered Surface	Weighted Filtered Surface
CRU & UAH	0.302	0.232	0.205	0.173
CRU3v & UAH4	0.303	0.247	0.202	0.166
CRU3v & RSS4	0.303	0.236	0.226	0.192
GISS data	0.196	0.222	0.211	0.209
all-GCM data	0.231	0.234	0.234	0.218

Table 6: Filtering results using various observational and model-generated data sets. Each table entry shows the mean trend in the global sample. The Surface trend is the mean trend in the surface observations for that row, likewise for the Troposphere trend. The fourth column shows the results from applying the filtering method from MM07 without gridcell cosine-weighting, and the final column shows the results with cosine-weighting applied to each gridcell. original MM07 results.

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 lower tropospheric trends. Had the socioeconomic coefficients retained their size and significance it would provide evidence that the surface results might be spurious. Some variables, such as growth in coal consumption (c) and population growth (p) can yield regional changes that affect the lower troposphere. One mechanism is the changing atmospheric aerosol load ([13]); another is induced changes in regional precipitation downwind from major cities ([24]). Table 3 of MM07 shows 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.

Variable	C3/U4+S	Diff-GCM	C3/R4+S	Diff-GCM
trop	**0.7269**	**0.6804**	**0.6976**	**0.6377**
	(9.04)	(8.57)	(8.05)	(7.64)
slp	**0.006**1	0.0051	0.0051	0.0042
	(2.16)	(1.75)	(1.83)	(1.49)
dry	4.7378	2.3197	4.7540	2.1985
	(1.37)	(0.63)	(1.36)	(0.61)
dslp	-0.0046	-0.0023	-0.0046	-0.0022
	(1.35)	(0.63)	(1.34)	(0.61)
water	-0.0287	-0.0087	-0.0209	-0.0008
	(1.56)	(0.48)	(1.12)	(0.04)
abslat	0.0004	-0.0006	**0.0030**	**0.0019**
	(0.45)	(0.73)	(3.06)	(2.03)
g	**0.0298**	**0.0343**	**0.0285**	**0.0337**
	(2.44)	(2.61)	(2.57)	(2.88)
e	**-0.0019**	**-0.0015**	**-0.0014**	**-0.0011**
	(3.58)	(2.97)	(2.73)	(2.24)
x	0.0002		-0.0009	
	(0.07)		(0.29)	
p	**0.2403**	**0.2524**	0.1682	0.1865
	(2.07)	(2.11)	(1.44)	(1.54)
m	0.1775	0.2405	0.1331	0.2037
	(1.36)	(1.78)	(1.02)	(1.50)
y	-0.1386	-0.1749	-0.0989	-0.1425
	(1.35)	(1.65)	(0.97)	(1.35)
c	**0.0040**	**0.0036**	**0.0045**	**0.0042**
	(2.19)	(2.00)	(2.43)	(2.33)
Constant	**-6.0249**	-5.1686	-5.1528	-4.3729
	(2.11)	(1.75)	(1.83)	(1.52)
Spatial lag	**0.3054**	**0.2874**	**0.3996**	**0.3620**
	3.84	3.40	5.11	4.69
***P*(H:g—c=0)**	**0.0005**	**0.0031**	**0.0050**	**0.0044**
N	428	428	428	428
***R*²**	0.58	0.49	0.57	0.48
LLF	164.98	162.90	159.90	155.51

Table 7: Results as for Table 4 using differences between observed and GCM-generate trends in each gridcell. For notation see notes to Table 4. 2nd and 4th columns taken from Table 4 for comparison purposes. 3rd and 5th columns: dependent variable is CRU3v trends minus trends generated by all-GCM average. Absolute *t* statistics in parentheses. Bold denotes significant at 5%.

However, the tropospheric trends vector is spatially autocorrelated, so the regression equation needs to be augmented with an SAC correction, which was not done in MM07. Table 8 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 regression results from MM07 for comparison, and the next four columns show the results after applying the SAC correction using UAH4 and RSS4, each shown with spatial lag and spatial error terms. The preferred model is somewhat ambiguous. The robust LM score indicates the spatial lag correction is more appropriate, but when both are estimated, the spatial error coefficient is more significant. However, nothing much depends on the choice in this case.

As in MM07, x changes sign and in one case (RSS4 with a spatial lag model) 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 never significant (see Tables 2 and 4), so its spurious effect does not drive any results in the rest of the analysis. Regarding the other coefficients, there is a plausible attenuation of the surface effects pattern, more so using UAH4 data than RSS4. In three cases the population growth measure falls to between one-quarter and one-sixth its size in the MM07 case (and one-half to one-third the size compared to the results in Table 7) but retains significance. This, in turn, carries over to the joint significance test in the RSS4 case, though not in the UAH4 case. The joint significance of the RSS4 test could be taken as support for the view that the patterns in our data are spurious. However, the effect appears to be mainly related to population growth, and in relation to this variable a regional effect in the lower troposphere is not implausible. Also, all the other specification tests up to this point, including the model-observation contrast, consistently support the view that the socioeconomic effects are actual features of the data, rather than fluke correlations.

The results in Table 8 are close enough to our expectations to allow us to maintain our interpretations of the rest of the analysis with respect to the complete model estimations and the comparison of modeled versus observational data. However, to provide a decisive test against spurious results 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 research.

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. We have shown that a coefficient pattern connecting temperature trends to indicators of industrialization is robust across a wide range of data configurations in the surface and lower troposphere, but is absent in climate model-generated data. The failure to reproduce this pattern in models indicates that it is not a natural feature of the climate system nor a response to greenhouse gas-induced forcing.

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, SAC is not found in the residuals when updated observations are used, but is strongly present when model-generated data are used. This again points to a qualitative difference between observations and models that may be explained by the patterns of socioeconomic development. After re-estimating our results with the appropriate corrections for spatial dependence we find the socioeconomic coefficients remain significant in observations, but, again, disappear in model-generated data. Also, we find no overlap between the implied distribution of coefficients from estimations on model-generated data and coefficients estimated on observed data, refuting a key conjecture in Schmidt [22].

Variable	MM07	UAH4	UAH4	RSS4	RSS4
uah	**0.8631**				
	(8.62)				
slp	0.0044	**0.0114**	0.0004	**0.0098**	0.0004
	(1.02)	(3.97)	(0.41)	(3.75)	(0.51)
dry	0.5704	1.3590	-1.9207	-1.6802	-1.9530
	(0.10)	(0.44)	(1.61)	(0.64)	(1.84)
dslp	-0.0005	-0.0013	0.0019	0.0017	0.0019
	(0.09)	(0.44)	(1.62)	(0.65)	(1.85)
water	-0.0289	0.0030	0.0091	0.0049	0.0095
	(1.37)	(0.45)	(1.80)	(0.72)	(1.77)
abslat	0.0006	0.0019	**0.0007**	0.0008	**0.0005**
	(0.51)	(1.53)	(2.45)	(0.75)	(2.20)
g	**0.0432**	-0.0091	0.0034	-0.0016	0.0049
	(3.36)	(0.67)	(0.68)	(0.12)	(0.91)
e	**-0.0027**	0.0001	0.0001	-0.0004	0.0000
	(5.14)	(0.42)	(0.79)	(1.69)	(0.20)
x	0.0041	0.0006	-0.0015	-0.0003	**-0.0022**
	(1.66)	(0.84)	(1.48)	(0.35)	(2.50)
p	**0.3839**	0.0451	**0.0599**	**0.0755**	**0.0948**
	(2.72)	(1.57)	(2.32)	(3.00)	(3.95)
m	**0.4093**	0.0513	0.0558	0.0336	0.0522
	(2.39)	(1.12)	(1.58)	(0.78)	(1.62)
y	**-0.3047**	-0.0370	-0.0416	-0.0265	-0.0418
	(2.22)	(1.11)	(1.53)	(0.84)	(1.76)
c	**0.0062**	-0.0005	0.0006	-0.0003	0.0005
	(3.45)	(0.46)	(1.16)	(0.38)	(1.01)
Constant	-4.2081	**-11.4593**	-0.4255	**-9.7904**	-0.4333
	(0.96)	(3.97)	(0.45)	(3.70)	(0.54)
Spatial error		**0.9474**		**0.9185**	
		(77.04)		(51.68)	
Spatial lag			**0.9196**		**0.8949**
			(61.85)		(45.02)
***P*(H:g—c=0)**	**0.0000**	0.6581	0.1287	**0.0116**	**0.0002**
N	440	440	440	440	440
***R*²**	0.53	0.23	0.94	0.12	0.879
LLF	139.22	647.65	639.52	617.32	611.62

Table 8: 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%.

Therefore, our overall finding is that the strong explanatory influence of socioeconomic effects on the pattern of climatic trends over land cannot be explained away as spurious effects due to spatial autocorrelation, data selection or fluke correspondence with known atmospheric circulation patterns. In the absence of any alternative explanation we conclude with some confidence that the temperature data being used for most modern analysis of climate change is inadequately filtered to remove known contamination patterns related to urbanization and other socioeconomic influences. The counterfactual experiments in Table 6 indicate that the contamination yields an overall warm bias 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 spatial trends for all climate models in the PCMDI archive trends. Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

References

1. 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.
2. Benestad RE (2004) Are temperature trends affected by economic activity? Comment on McKitrick & Michaels (2004). *Clim. Res.* 27:171–173
3. 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
4. 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.
5. De Laat, A.T.J., and A.N. Maurellis (2006), Evidence for influence of anthropogenic surface processes on lower tropospheric and surface temperature trends, *Int. J. Climatol.* 26:897—913.
6. Florax, Raymond J.G.M., Hendrik Folmer and Sergio J. Rey (2003). “Specification searches in spatial econometrics: the relevance of Hendry’s methodology.” *Regional Science and Urban Economics* 33 557–579.
7. 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.
8. 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.
9. 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.
10. 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.
11. Klotzbach, P.J., R.A. Pielke Sr., R.A. Pielke Jr., J.R. Christy, and R.T. McNider, (2009): An alternative explanation for differential temperature trends at the surface and in the lower

troposphere. *Journal of Geophysical Research-Atmospheres*, 114, D21102, doi:10.1029/2009JD011841, 2009
12. Kmenta, Jan (1986) *Elements of Econometrics* New York: MacMillan.
13. Li, Zhanqing Li, H. Chen, M. Cribb, R. Dickerson, B. Holben, C. Li, D. Lu, Y. Luo, H. Maring, G. Shi, S.-C. Tsay, P. Wang, Y. Wang, X. Xia, Y. Zheng, T. Yuan, and F. Zhao. "Preface to special section on East Asian Studies of Tropospheric Aerosols: An International Regional Experiment (EAST-AIRE). *Journal of Geophysical Research-Atmospheres* (112, D22S00, doi:10.1029/2007JD008853).
14. McKitrick, R.R. (2010), Atmospheric Oscillations do not Explain the Temperature-Industrialization Correlation. *Statistics, Politics and Policy* Vol 1, July 2010.
15. McKitrick, 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.
16. McKitrick, R.R. and P. J. Michaels (2004b), Erratum, *Clim. Res.* 27(3) 265—268.
17. McKitrick, 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.
18. 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 .
19. Mitchell Jr., J.M. (1953), On the causes of instrumentally-observed secular temperature trends, *Journal of Meteorology* 10, pp. 244-261.
20. Pisati, Maurizio (2001) Tools for spatial data analysis. *Stata Technical Bulletin* STB-60, March 2001, 21—37.
21. 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.
22. Schmidt, GA (2009) Spurious correlation between recent warming and indices of local economic activity. *International Journal of Climatology* 10.1002/joc.1831
23. 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.
24. Shepherd, J. Marshall, Harold Pierce and Andrew J. Negri (2002). "Rainfall modification by major urban areas: observations from spaceborne rain radar on the TRMM satellite." *Journal of Applied Meteorology* 41: 689-701.
25. Spencer, R.W. and J.C. Christy (1990), Precise monitoring of global temperature trends from satellites, *Science* 247:1558—1562.

APPENDIX:

The all-GCM data was constructed using 55 runs from 22 GCMs used in the IPCC [7] 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. ([21]). 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.