Vol. 27: 171–173, 2004 | CLIMATE RESEARCH Clim Res | Published October 7

COMMENT

Are temperature trends affected by economic activity? Comment on McKitrick & Michaels (2004)

R. E. Benestad*

The Norwegian Meteorological Institute, PO Box 43, 0313 Oslo, Norway

McKitrick & Michaels (2004) raise an important question regarding whether economic activity and other non-climatic conditions may affect the estimate of true temperature trends derived from a climate station network. They used a number of non-climatic factors and the econometrics program SHAZAM (White 1993) to derive a heteroskedasticity-consistent covariance matrix (White 1980) for modelling the statistical relationship between the trend estimates and non-climatic factors. Ordinary Least Squares (OLS) models may produce biased estimates, and the presence of heteroskedasticity in the residuals may be an indication of model mis-specification such as incorrect functional form. The SHAZAM model therefore ought to give unbiased estimates of the coefficients describing the relationship between a number of factors and the temperature trend. Nevertheless, if the conclusion is that there is a clear and significant contamination from non-climatic factors on temperature trend estimates, their results should not be particular to one specific model. Although OLS and Generalised Linear Models (GLM) may not yield the optimal unbiased values, the same pattern should be found in these more common regression analyses. Here, their results are re-tried in order to check whether the neglect to take inter-station dependencies into account may have influenced their conclusion. The selection of stations by McKitrick & Michaels (2004) implies a significant spatial correlation, as for instance 44 of the stations were from eastern or western Europe.

The analysis used the data from McKitrick & Michaels (2004), available from www.uoguelph.ca/~rmckitri/research/gdptemp.html. The data were provided in Excel DIF-format, read using Excel, and subsequently saved as a pure DOS text file. 'D-' was changed to 'E-' using Emacs in order to read the data into the R-environment (Gentleman & Ihaka 2000, Ellner 2001) (a 'GNU version of Splus' freely available from http://cran.r-project.org/). The linear and generalised linear models in the R-environment (lm and glm) were used to do the regression analysis. These models gave very similar results, and therefore only the results from the linear model are shown here. Table 1 gives a summary of the regression results given by these models. A comparison with Table 4 in McKitrick & Michaels (2004) (their results are reproduced here in Table 2) suggests a good agreement: The R^2 are similar, and all the coeffcients that are considered statistically significant at the 5% level (shown in bold) have similar values.

One concern about the analysis carried out by McKitrick & Michaels (2004) is that the effect of inter-station dependencies will result in spurious results. In order to reduce the possible effect of inter-station dependency, the data were sorted according to latitude. Then, half of the data (latitudes from 75.5° S to 35.2° N) was used to calibrate the statistical models

Table 1. Summary of coefficients given by linear model. Significance (Sig.) codes: 0 < '***' < 0.001 < '**'< 0.01 < '*'< 0.05 < '.'< 0.1 < ' ' < 1; residual standard error: 0.40 on 205 degrees of freedom; multiple R^2: 0.26; adjusted R^2: 0.22; *F*-statistic: 6.03 on 12 and 205 df; p-value: 5.4e−09. Values significant at p < 0.05 are shown in **bold**

	Estimate	SE	*t*-value	Pr(> |*t*|)	Sig.
(Intercept)	−1.567e+01	6.472e+00	−2.421	0.016363	*
PRESS	+1.608e−02	6.435e−03	2.498	0.013260	*
WATER	+9.706e−02	5.856e−02	1.657	0.098978	.
ABSCOSLAT	−3.786e−01	2.069e−01	−1.830	0.068760	.
POP	+1.139e−08	1.927e−08	0.591	0.555085	
SCALE79	−1.218e−09	1.850e−09	−0.658	0.511020	
COAL80	**−4.595e−01**	1.540e−04	−2.983	0.003199	**
COALGROW	−5.131e−03	6.473e−03	−0.793	0.428826	
INC79	**+3.906e−02**	1.035e−02	3.774	0.000211	***
GDPGROW	**+8.711e−02**	2.186e−02	3.985	9.37e−05	***
SOVIET	**+5.046e−01**	1.272e−01	3.968	0.000100	***
SURFMISS	+1.802e−04	1.169e−03	0.154	0.877696	
LIT79	**−5.224e−03**	1.717e−03	−3.043	0.002647	**

*Email: rasmus.benestad@met.no

Table 2. Coefficients after McKitrick & Michaels (2004). Multiple R^2: 0.25; adjusted R^2: 0.20; df = 205. Values significant at $p < 0.05$ are shown in **bold**

	Estimate	*t*-value
(Intercept)	−1.27e+01	1.16
PRESS	+1.3e−02	1.18
WATER	+1.03e−01	1.92
ABSCOSLAT	−8e−03	0.2
POP	+2e−03	0.6
SCALE79	−2e−03	0.51
COAL80	**−4.5e−01**	2.72
COALGROW	−7e−03	1.68
INC79	**+4.6e−02**	4.46
GDPGROW	**+9.1e−02**	4.55
SOVIET	**+5.92e−01**	5.46
SURFMISS	−0	0.09
LIT79	**−5e−03**	2.99

and the remaining data were used for evaluation (latitudes 35.3° to 80.0° N). Fig. 1 shows the results of the regression analysis for a number of models: (1) a model using all the dependent variables in Table 4 of McKitrick & Michaels (2004); (2) a model using only physical geographical information; (3) a model using only geographical information and population data; (4) a model using only non-climatic factors; and (5) a model using only those factors that McKitrick & Michaels found to be significant. None of these models are able to reproduce the independent data. Moreover, none of these models can account for the high positive temperature trends seen in the station data. The models based on non-climatic factors yield near-zero trends for the region of independent data. Additional tests of Pearson correlation give a value of 0.096 between the station-based trends and GDP growth (the probability of no correlation [null-hypothesis], or the p-value, is 0.16, assuming 216 degrees of freedom [df], a number which is much higher than the effective degrees of freedom), and 0.25 between trends and an index based on whether the stations were located within the former USSR (p-value of 0.0002, assuming df = 216). McKitrick & Michaels proposed that the collapse of the Soviet Union and the following neglect to maintain the temperature stations within the former USSR may affect

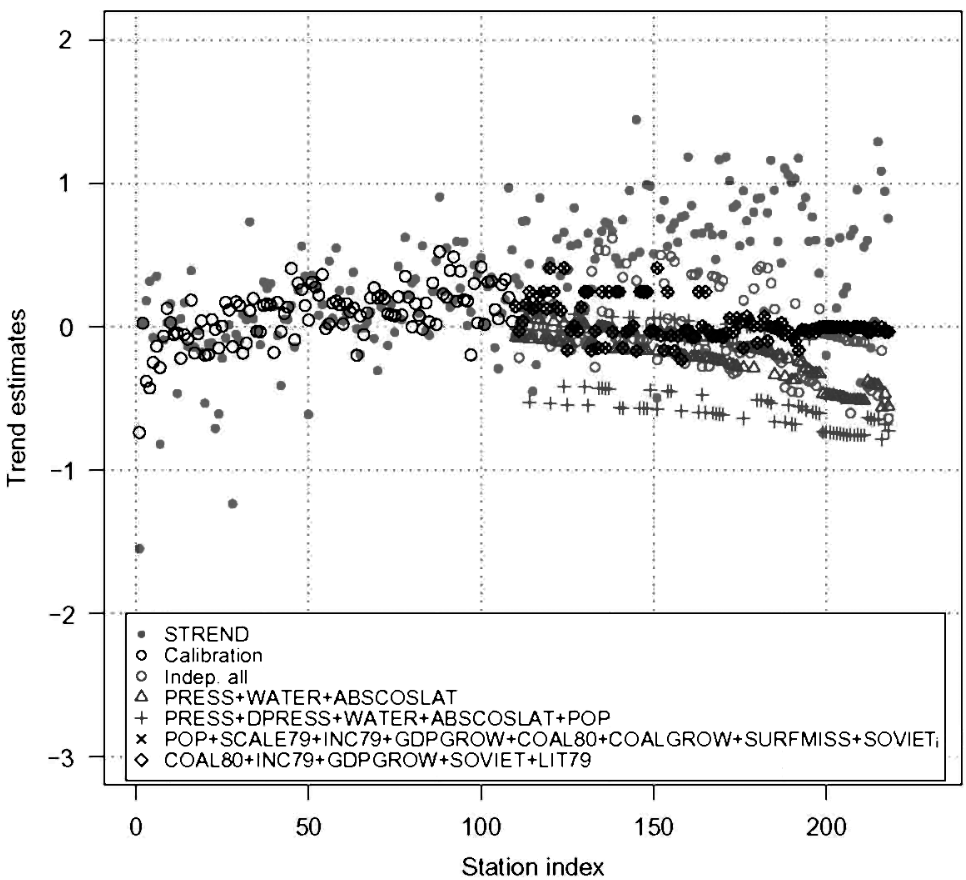

Fig. 1. Results of regression analyses with different models using data from stations within the latitude range 75.5° S to 35.2° N to calibrate the model and stations in the latitude range 35.3° to 80.0° N and corresponding depending variables for prediction and evaluation. The unit along the *y*-axis is °C decade^{-1}

the trend estimates. The temperature change over Russia can be examined at the GISS web site (www.giss.nasa.gov/data/update/gistemp/), from which the climate station temperatures originally were taken. A comparison between the differences in mean winter temperatures over 1978–1989 and 1967–1978 shows a prominent warming pattern over most of the old Soviet Union which also occurred before the collapse of the USSR.

The analysis by McKitrick & Michaels (2004) was repeated using a different statistical modelling technique. If their results were robust, one would expect to find similar patterns with different models. The regression analysis produced similar, although not identical, model coeffcients, *t*-values, and R^2 scores to those reported by McKitrick & Michaels, indicating that the analysis captures similar relationships. However, when a validation was performed on a similar analysis for which the regression model was calibrated with a subset of the data, and the remaining data were used for validation, it became apparent that models based on the factors that McKitrick & Michaels used had no skill (i.e. were not able to reproduce the independent data). The negligence to account for inter-station dependencies in the analysis resulted in spurious results and inflated confidence levels in the analysis of McKitrick & Michaels. There is therefore no evidence suggesting that the temperature trends are systematically influenced by non-climatic factors. Support for this finding can be found in Boehm et al. (1998) over Austria, and Houghton et al. (2001) (Box 2.1 on p. 106) over the northern hemispheric land area where the urban heat island is expected to be greatest, in which the old Soviet states are located, and where coal consumption and GDP growth are greatest.

Acknowledgements. This work was done under the Norwegian Regional Climate Development under Global Warming (RegClim) programme, and was supported by the Norwegian Research Council (Contract NRC-No. 120656/720) and the Norwegian Meteorological Institute. The analysis was carried out using the R data processing and analysis language, which is freely available on the Internet (www.R-project.org/).

LITERATURE CITED

Boehm R, Auer I, Schoener W, Hagen M (1998) Long alpine barometric time series in different altitudes as a measure for 19th/20th century warming. In: 8th Conference on Mountain Meteorology. Am Meteorol Soc, Boston, MA, p 72–76

Ellner SP (2001) Review of R, Version 1.1.1. Bull Ecol Soc Am 82:127–128

Gentleman R, Ihaka R (2000) Lexical scope and statistical computing. J Comput Graph Stat 9:491–508

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (2001) Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of IPCC. International Panel on Climate Change (available from www.ipcc.ch)

McKitrick R, Michaels PJ (2004) A test of corrections for extraneous signals in gridded surface temperature data. Clim Res 26:159–173

White K (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–838

White K (1993) SHAZAM econometrics computer program users manual. McGraw-Hill, Toronto

Editorial responsibility: Chris de Freitas, Auckland, New Zealand

Submitted: August 11, 2004; Accepted: August 30, 2004
Proofs received from author(s): September 6, 2004