Referee Report:

Influence of Urban Heating on the Global Temperature Land Average Using Rural Sites Identified from MODIS Classifications
by Wickham et al.

Overall Comment

This paper tries to do 2 things in a single, short paper; namely introduce a new global temperature data product with a much larger number of stations than are available in GHCN and related products, and provide a quantification of non-climatic biases in surface temperature records. While the authors have developed an impressive new data base, the paper fails unfortunately to do a satisfactory job of either task. First, it omits many of the technical details readers need to assess the new data base construction methodology. Second, the analysis of the urban-rural split is simplistic in light of where the current literature stands, and is not able to support the conclusions drawn. Specifically, the authors' empirical results are consistent either with the stated conclusion or its opposite, and therefore they are in no position to say anything decisive.

I will recommend that the paper be rejected in its current form. I have no doubt that presentation of an important new surface data base is a publishable contribution, as long as some major improvements to the manuscript are made, as detailed below. But with regard to the analysis of surface disruptions and the spatial distribution of temperature trends, the analysis presented herein has serious inadequacies that make it unpublishable in its current form.

Introduction of new data base

The paper referred to as Rhode et al. (2010) on page 6, lines 106-107, and elsewhere, does not appear to be a publication. It should not be listed in the references. Yet almost all the necessary technical details that should be in this manuscript are apparently in it instead. This is a disservice to the reader. The material relegated to an unpublished source would appear to include just about everything that readers need to know about the new data set to decide on its validity.

A partial listing of technical material that needs to be incorporated into the present paper includes the following.

- List the source data sets and metadata.
- Explain the averaging methodology in detail, using sufficient math to permit independent replication.
- p. 7 lines 124-131: show the effect of varying the definition of "very rural" to something other than the assumed tenth of a degree separation. How important is this parameter?
- p. 10 lines 174-183: Explain how many missing months are permitted in a continuous series before the series is split, or discarded.
- p. 12 lines 210-212: Explain the rationale behind this apparently ad hoc statistical procedure for determining standard errors. Is this some sort of block bootstrap method? There needs to be reference to standard, mainstream statistical literature explaining why

this resampling procedure is used and why the authors believe it yields asymptotically valid standard errors. If no theoretical guidance is available the authors could perhaps use Chebychev's inequality to provide an upper bound on the variance.

- p. 14 lines 252-266: Provide a discussion of how the tradeoff between continuity and fragmentation affects the data quality. That is, the rule for terminating a series will determine whether there are a few long but intermittent series, or many short but continuous series. Under what circumstances is the latter a better measurement, and how is the choice optimized?
- p. 14 lines 252-266: The authors claim to have "taken into account spatial correlation" yet there isn't a word anywhere in the paper about how this is done. Since the authors cite McKitrick 2010 and McKitrick and Nierenberg 2010 they presumably have read both papers, which contain (especially the latter) detailed explanations about how spatial autocorrelation is tested for and corrected in models of surface temperature trends. The elements of the discussion required for a proper treatment of this topic include reporting a robust LM statistic, a parametric model of the spatial weights, a description of the estimation method for computing the SAC terms and the optimal distance weighting parameter, and test results on the residuals to indicate whether the SAC model was adequate.
- p. 14 lines 252-266: Again in this section there is reference to a resampling method to compute standard deviations, but no explanation is given, nor is there any reference to statistical literature. Is this some sort of bootstrap method? An explanation is needed.
- p. 14 lines 268-272: With respect to the iterative weighting procedure, how do you know that this converges to a unique solution? It is possible the weights are path-dependent. We need to be shown some details about the convergence rule and the way the results are tested by trying different starting values.

In the absence of so much elementary material it is difficult even to review this paper. I understand that a great deal of work has gone into the project, and the release of a new data set with improved sampling characteristics is a valuable contribution. In rejecting the present manuscript I hope the authors will revisit the task of explaining their work with some alacrity and will resubmit a much expanded paper so that the new data base can be published.

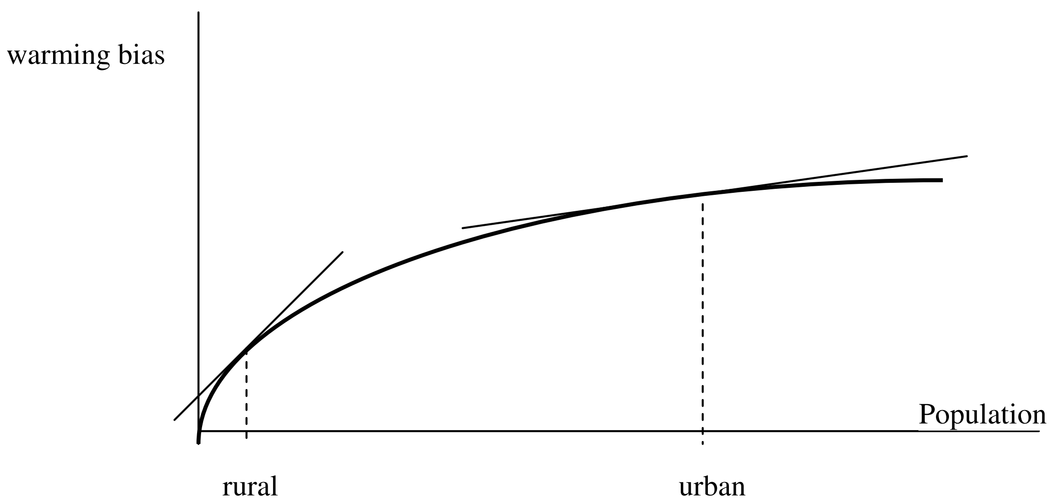
Quantifying the effect of nonclimatic contamination of the data

Judging by the paper's title this appears to be the topic the authors want to focus on. It is clear that, if published, this will be a very prominent paper and its findings will be wielded to considerable polemical effect: indeed one of the authors has already taken the liberty of announcing partial findings in Congressional testimony. Great care must be taken to ensure that findings are accurate and are fully supported by the empirical analysis. In this regard I note two problems: the paper reads as if the authors have been careless in reviewing the existing debate, and the empirical work does not imply the conclusions.

The authors cite, in passing, papers by de Laat and Maurellis and McKitrick and coauthors (pp. 5-6) that present evidence of significant surface data contamination. They also cite papers that argue for the absence of such contamination. Despite the fact that Wickham et al. purport to adjudicate between these different literatures they do not summarise or explain the very different methodologies involved nor how their analysis relates to them, if at all. On page 13 lines 234-235 the authors conclude that their result "agrees with the conclusions in the literature that we cited previously" which is a baffling statement given that they cite papers that directly contradict one another. My overall impression is that the authors have not actually read all the papers they cite, and have not come to terms with the technical issues involved in the current debate. If it is their purpose to draw conclusions about the surface data contamination question they need to position their own analysis properly in the existing literature, which will require a detailed explanation of what has been done hitherto, and the use of an empirical framework capable of encompassing existing methodologies.

With regard to their own empirical work, a basic problem is that they are relating a change term (temperature trend) to a level variable (in this case MODIS classification) rather than to a corresponding change variable (such as the change in surface conditions). I will give a simple example of why this is a flawed method, then I will demonstrate it empirically.

Suppose there are only two weather stations in the world, one rural and one urban. Suppose also that there is zero climatic warming over some interval, but there is a false warming due to local population growth, the effect of which is logarithmic, as is commonly assumed. Then the measured trends would be proportional to the respective tangent lines:



A sample split according to the rural/urban distinction would apparently show that the rural station has a higher trend than the urban one. Far from proving that there is no urban bias in the overall average, it is precisely the result we expect if there *is* such a bias! And the contrast would be larger, the wider the difference between "urban" and "very rural". Consequently the authors' univariate analysis cannot, in principle, be the basis of their assertion that there is little or no urbanization bias, since the results are consistent with such a bias being present.

To provide an empirical demonstration, I obtained the GEcon data base from Yale University (http://gecon.yale.edu/) which provides gridded population, GDP, climatic and other indicators over the 1990-2005 interval for 27,500 terrestrial grid cells at 1 degree resolution. I then interpolated CRU grid cell trends over 1990-2010 for the same grid cells. After removing cells with missing socioeconomic data, or in which more than 25% of the years are missing 4 or months of temperature data, I was left with just under 18,000 grid cells with observations on the linear temperature trend, latitude, minimum temperature, standard deviation of precipitation, distance to coast, number of missing months in temperature record, 2005 population per square km, change in population (1990 to 2005), 2005 GDP (U$, PPP-based) per square km, change in GDP per sq km 1990 to 2005, 2005 GDP per capita and change in GDP per capita over 1990 to 2005.

To replicate the results in Wickham et al, I regressed the vector of trends on a static measure of surface disruption, namely 2005 grid cell population/km^2, using White's corrected residuals.

```
regress trend90 pop2005 , robust

Regression with robust standard errors                 Number of obs =   17962
                                                       F(  1, 17960) =   11.59
                                                       Prob > F      =  0.0007
                                                       R-squared     =  0.0003
                                                       Root MSE      =  .37354

------------------------------------------------------------------------------
             |               Robust
     trend90 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     pop2005 |  -.0270307   .0079415    -3.40   0.001    -.0425969   -.0114645
       _cons |   .3275282     .00293   111.78   0.000     .3217851    .3332713
------------------------------------------------------------------------------
```

The results mirror those of Wickham et al. The coefficient on POP2005 is negative and significant, apparently indicating that regions with higher population per square km have slightly (but significantly) lower trends. I then re-did the same analysis using 2005 GDP/km^2 as the measure of surface temperature disruption.

```
regress trend90 gdp2005 , robust

Regression with robust standard errors                 Number of obs =   17962
                                                       F(  1, 17960) =   14.77
                                                       Prob > F      =  0.0001
                                                       R-squared     =  0.0007
                                                       Root MSE      =  .37347

------------------------------------------------------------------------------
             |               Robust
     trend90 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     gdp2005 |  -1.915341   .4984296    -3.84   0.000    -2.892311   -.9383709
       _cons |   .3272088   .0028301   115.62   0.000     .3216615    .3327561
------------------------------------------------------------------------------
```

Again the results mirror those of Wickham et al. The coefficient on GDP2005 is negative and significant, thus "confirming" that relatively undisturbed regions apparently have higher warming trends, a result they deem anomalous in light of prior expectations.

But it is not anomalous at all, it just reflects the fact that this class of empirical model cannot measure what the authors have tried to measure. The problem can be remedied by adding in fixed climatic covariates and socioeconomic *change* terms. Ideally I would also put in the lower tropospheric trend terms on the right hand side, but I don't have them handy and they are not needed for the illustration. Here are the results of the multivariate model:

```
regress trend90 lat tempmin precsd dist2cst miss_months pop2005 gdp2005 inc2005 chg_gdp chg_pop
chg_inc
>   , robust

Regression with robust standard errors                Number of obs =   17265
                                                      F( 11, 17253) =  305.16
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.1417
                                                      Root MSE      =  .33096

------------------------------------------------------------------------------
             |               Robust
     trend90 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lat |    .004216   .0000899    46.88   0.000      .0040397    .0043923
     tempmin |   .0056988   .0002292    24.86   0.000      .0052496    .0061481
      precsd |   -.001461    .000077   -18.96   0.000     -.001612     -.00131
    dist2cst |  -7.42e-06   5.85e-06    -1.27   0.204     -.0000189    4.04e-06

 miss_months |   .0096775    .000364    26.59   0.000      .0089641     .010391

     pop2005 |  -.0433431   .0350422    -1.24   0.216     -.1120293    .0253432
     gdp2005 |  -.9251954   1.961361    -0.47   0.637     -4.769663    2.919272
     inc2005 |  -30.58095   32.19992    -0.95   0.342     -93.69606    32.53415

     chg_gdp |  -6.464729    4.49109    -1.44   0.150     -15.26772    2.338263
     chg_pop |   .4839018   .1352258     3.58   0.000      .2188455    .7489581
     chg_inc |    .066803   .0045241    14.77   0.000      .0579353    .0756707

       _cons |   .2056122   .0053144    38.69   0.000      .1951955     .216029
------------------------------------------------------------------------------
```

Latitude, MinTemp and SD of precipitation are all significant. "Miss_months" indicates the number of missing months in the data series after 1990. It is significant, and indicates that the more missing months in a series, the higher the estimated trend.

Look carefully at GDP2005 and POP2005: they are still negative but they have become small and insignificant. 2005 per capita income (INC2005) is also insignificant.

Meanwhile the change term CHG_POP (population growth) is positive and significant, as is CHG_INC (income growth). In other words it is the *change* in socioeconomic measures that correlates to the *change* in temperature over an interval of time, and once these effects are controlled the apparent contrast in trends based on a static measure of surface disruption such as GDP or Population (or, likely, MODIS land classification) becomes insignificant and irrelevant.

The joint test on the socioeconomic variables has an F statistic of 132.47, which is extremely significant, indicating that we would reject the hypothesis that surface trends are unaffected by socioeconomic factors at the surface. Using the method outlined in McKitrick and Michaels 2007 to filter the trend vector, the mean trend falls from about 0.33 to 0.26, indicating the socioeconomic effects add up to a net warm bias of about 0.07 C/decade, which is comparable to the results in Table 6 of McKitrick and Nierenberg 2010, even though this is a different data set using different covariates on a different time period; but this part of the analysis is difficult to do without the full set of covariates including the satellite-based trends.

To emphasize the contrast: on a large global data set, if I use a naïve analysis comparable to Wickham et al., namely relying on 2005 population as the only regression covariate, I get the same, "anomalous" result that they do, namely that higher-population regions apparently have slightly lower trends than low-population regions. But when I remedy the conceptual weakness in their model by introducing change terms on the right hand side, the population *level* turns out to be insignificant, and instead the population *change* term has a positive and significant effect on the trend, implying that population growth biases the surface trends upwards. Likewise per capita income *growth*, but not the *level*, is positively correlated with the size of the trend.

Conclusion

The simple univariate analysis in Wickham et al. does not establish a sound basis for their assertion that surface temperature data are unaffected by urbanization and related socioeconomic disruption of the surface. To draw such a conclusion would require setting up a model capable of measuring these effects. At least three improvements to the modeling framework are needed to bring the analysis up to the level of the current debate.

- Use of a suite of covariates that can identify the contrasting effects of different sources of bias such as anthropogenic surface processes, data inhomogeneities and regional atmospheric pollution;

- Comparison of the observed spatial trend pattern to those predicted in climate models so that a null hypothesis is clearly identified and spurious results can be ruled out;

- Examination of spatial autocorrelation of the model residuals to permit identification of the explanatory variables needed to yield iid residuals, in support of making asymptotically accurate inferences.

A very simple way to proceed would be to compute post-1979 gridded trends in the BEST archive and merge them with the McKitrick and Nierenberg data set, then run the code available online. I conjecture that the results will look a lot like those reported in McKitrick and Nierenberg (2010), but whatever is the case I encourage the authors to use their new data set for such an analysis and see what emerges. Meanwhile I cannot recommend this draft for publication.

- Minor point: M&N should be cited

McKitrick, Ross R. and Nicolas Nierenberg (2010) Socioeconomic Patterns in Climate Data. *Journal of Economic and Social Measurement*, 35(3,4) pp. 149-175. DOI 10.3233/JEM-2010-0336.

Signed review: Ross McKitrick.