

---

# Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends.

Kevin Cowtan<sup>a\*</sup> Robert G. Way<sup>b</sup>

<sup>a</sup>*Department of Chemistry, University of York, UK*

<sup>b</sup>*Department of Geography, University of Ottawa, Canada*

\*Correspondence to: kevin.cowtan@york.ac.uk

---

**Incomplete global coverage is a potential source of bias in global temperature reconstructions if the unsampled regions are not uniformly distributed over the planet's surface. The widely used HadCRUT4 dataset covers on average about 84% of the globe over recent decades, with the unsampled regions being concentrated at the poles and over Africa. Three existing reconstructions with near-global coverage are examined, each suggesting that HadCRUT4 is subject to bias due to its treatment of unobserved regions.**

**Two alternative approaches for reconstructing global temperatures are explored, one based on an optimal interpolation algorithm and the other a hybrid method incorporating additional information from the satellite temperature record. The methods are validated on the basis of their skill at reconstructing omitted sets of observations. Both methods provide superior results than excluding the unsampled regions, with the hybrid method showing particular skill around the regions where no observations are available.**

**Temperature trends are compared for the hybrid global temperature reconstruction and the raw HadCRUT4 data. The widely quoted trend since 1997 in the hybrid global reconstruction is two and a half times greater than the corresponding trend in the coverage-biased HadCRUT4 data. Coverage bias causes a cool bias in recent temperatures relative to the late 1990s which increases from around 1998 to the present. Trends starting in 1997 or 1998 are particularly biased with respect to the global trend. The issue is exacerbated by the strong El Niño event of 1997-1998, which also tends to suppress trends starting during those years. Copyright © 0000 Royal Meteorological Society**

*Key Words:* Instrumental temperature record, coverage bias, temperature trends

## 1. Introduction

The instrumental temperature record, based on land-based weather station readings and sea surface temperature readings from ships and buoys, forms a vital source of information concerning climate over the last century and beyond. Time series of local monthly average temperatures (gridded datasets) and global averages are produced by several organisations, with the HadCRUT product of the Hadley Center and Climatic Research Unit being one of the most widely cited (Brohan *et al.* 2006; Morice *et al.* 2012).

The GISTEMP product from NASA's Goddard Institute for Space Studies (Hansen *et al.* 2010) and the National Climatic Data Center (NCDC) product (Smith *et al.* 2008) are also widely used. A new land-only product from the Berkeley Earth 'BEST' project (Rohde *et al.* 2013a) has introduced a number of statistical improvements in the handling of data homogenisation and coverage.

All these temperature reconstructions are based on *in situ* readings using thermometers. While the production and calibration of reliable thermometers has been well established for several centuries, how the thermometers are

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.2297

used to obtain meteorological data can influence the results. Sources of measurement bias in the land temperature record have been widely studied and include time of observation, instrumentation type and location (Williams *et al.* 2012), and the 'Urban Heat Island' (UHI) effect (Hausfather *et al.* 2013).

Practices for measuring sea surface temperatures (SSTs) have changed over time, leading to biases in the resulting data. Measurements have been performed using uninsulated or insulated buckets, engine room intake temperatures, and anchored or free-floating buoys (Kennedy *et al.* 2011). A pre-WWII bias due to evaporation from uninsulated buckets is dealt with in every version of the SST record. Post-WWII corrections are only present in the Hadley/CRU temperature record (Morice *et al.* 2012), and include an upward adjustment to recent temperatures due to a transition away from warm-biased engine room sensors to measurements by buoys.

A further important source of bias arises from the estimation of a global mean temperature from the incompletely sampled gridded dataset. Weather station coverage is best in the temperate latitudes and particularly in developed nations. Coverage of the polar regions is particularly poor, with no coverage of Antarctica before the 1950's, and limited coverage of the Arctic to this day. Poor sampling of the fastest warming parts of the planet leads to an underestimation of the global trend in the instrumental temperature record (Simmons *et al.* 2010; Folland *et al.* 2013). This problem is exacerbated by the use of equal-angle (5 degree) grids by the Hadley/CRU record: since equal-angle grid cells become smaller at higher latitudes, more stations are required to achieve full coverage when in practice fewer stations are available.

Coverage bias becomes an issue when different parts of the planet are changing temperature at different rates. As a result it is a particular issue over recent decades, owing to the different rates of warming between the tropics and poles, and between land and ocean (Hansen *et al.* 2006). Changes in the Arctic and Antarctic are particularly problematic because the coverage is poor in these regions.

While short term trends are generally treated with a suitable level of caution by specialists in the field, they feature significantly in the public discourse on climate change (e.g. Global Warming Policy Foundation 2012; Daily Mail 2012; The Guardian 2012; The Telegraph 2013), and have also been the subject of scholarly studies into the possible impact of both aerosol emissions (Kaufmann *et al.* 2011) and ocean heat uptake (Meehl *et al.* 2011) on short term trends. A common factor in all these works is an attempt explain or draw conclusions from the comparatively slow warming seen in some versions of the instrumental temperature record over the past decade and a half. However to do so without first addressing the issue of coverage bias is to ignore an important confounding factor in the analysis.

NASA's GISTEMP temperature record (Hansen *et al.* 2010) attempts to address the coverage issue by extrapolating temperatures into unmeasured regions by means of kernel smoothing using a conical kernel of radius 1200km. The Berkeley Earth (BEST) project have adopted an optimal interpolation method (kriging), although only for land temperatures at this stage (Rohde *et al.* 2013a). However each of these approaches assumes that the unobserved high latitude temperature field varies in a similar way to the observed temperatures at lower latitudes.

Other sources of temperature data include reanalysis data, which are determined by using a weather model to reconstruct a global temperature field on the basis of multiple sources of data (e.g. Kalnay *et al.* 1996), and the satellite temperature record (Spencer 1990; Mears and Wentz 2009). The satellite record is of particular interest because it provides a uniform sampling with near-global coverage. However the satellite microwave sounding units measure lower troposphere rather than surface temperatures and so are not directly comparable to the *in situ* temperature record. Furthermore there are temporal uncertainties in the satellite record arising from satellite failure and replacement and the numerous corrections required to construct a homogeneous record (Karl *et al.* 2006). Contamination of the microwave signal from different surface types is also an issue, particularly over ice and at high altitude (Mears *et al.* 2003).

The purpose of this work is to address the issue of coverage bias through the development of new methods for global temperature reconstruction building on the HadCRUT4 data.

## 2. Preliminary analysis

The potential impact of coverage bias may be estimated by use of three (near) global temperature reconstructions: The extrapolated GISTEMP data, the University of Alabama in Huntsville (UAH) satellite data (Spencer 1990; Christy *et al.* 2007), and the NCEP/NCAR reanalysis data (Kalnay *et al.* 1996). Figure (1) shows temperature trend maps for the period 1997/01-2012/12 for HadCRUT4 and each of these three series (the significance of the start date will become clear shortly).

Note that GISTEMP, UAH and NCEP/NCAR all show faster warming in the Arctic than over the planet as a whole, and GISTEMP and NCEP/NCAR also show faster warming in the Antarctic. Both of these regions are largely missing in the HadCRUT4 data. If the other datasets are right, this should lead to a cool bias due to coverage in the HadCRUT4 temperature series.

A preliminary estimate of the size of the bias may be calculated from the three global temperature series. For each series the coverage of the temperature field for each month is reduced to match that of the HadCRUT4 data for that month. Global mean temperature estimates are calculated for both the full- and reduced-coverage temperature fields. The difference between these values gives an estimate of the coverage bias (subject to a constant offset determined by the bias over the baseline period). The bias estimate can vary dramatically from month to month as weather systems move in and out of the omitted regions, however a 60 month moving average, shown in Figure (2), shows long term variations. All the global series show a rapidly increasing cool bias over the past two decades, with a sharp decline starting around 1998. The GISTEMP and UAH estimates for HadCRUT4 coverage bias are very similar, providing some support for the GISTEMP extrapolation approach. The NCEP/NCAR data show a much faster transition to a cool bias followed by a plateau, raising a question over the GISTEMP assumption that temperatures at high latitudes vary similarly to those at lower latitudes.

The ERA-interim reanalysis (Dee *et al.* 2011) and the 20th Century reanalysis (Compo *et al.* 2011) also show a rapidly increasing cool bias after 1998. While it is possible that these results could arise from limitations in all of the

reanalyses and differences between lower troposphere and surface air temperatures, the commonality of this feature is certainly suggestive.

### 3. Global temperature reconstruction

In order to construct a global surface temperature series, either surface temperature observations or proxies which allow their estimation are required. No static weather station observations are available for the central Arctic, and thus a proxy is the only option. Hansen *et al.* (2006) used satellite radiometer data to argue that a warm Arctic anomaly in the GISTEMP temperature record in 2005 was genuine. The UAH satellite data has near global coverage, and the global distribution of satellite temperatures for any given month is correlated with the surface temperatures, with a mean (Pearson) correlation of 0.61 between GISTEMP and UAH when using the same baseline period. The correlation is greater over land (0.75) and lower over oceans (0.32). This raises the possibility of using the satellite data as a proxy for surface temperature to construct a geographically complete hybrid temperature record.

The UAH satellite data is used in preference to data from Remote Sensing Systems (Mears and Wentz 2009) because RSS omits the critical high latitude temperature data, which is most impacted by the surface contamination issue (Mears *et al.* 2003). The UAH data suffers from the same issues, however this work aims to mitigate surface contamination bias by use of the *in situ* data. The UAH data is also interpolated at latitudes above 85°, however the interpolated region is small compared to the unobserved region of the *in situ* data.

The use of the satellite temperatures as a proxy rather than a direct measurement of surface temperature brings additional requirements. Firstly, the satellite and *in situ* observations must be on a common baseline. Secondly, an appropriate method for mapping satellite observations of lower troposphere temperatures onto surface temperatures is required. Thirdly, the method must be validated to ensure that it has skill in predicting unobserved temperature values. The validation step will also serve as a check on the possible issues identified so far such as the mismatch between surface and lower troposphere temperatures and surface contamination of the MSU signal.

The mapping step makes use of the optimal interpolation algorithm known as kriging, which will be used to estimate a slowly varying function of grid coordinates corresponding to the offset between the satellite and surface temperatures.

#### 3.1. The impact of the baseline period

The surface temperature calculation is usually performed using temperature anomalies, which represent the deviation of the current temperature from the mean over a chosen baseline period. For the HadCRUT4 data, the station data are normalised so that the mean over the period 1961-1990 for a given month of the year is zero for each station with sufficient records. The UAH map data uses a similar approach with the mean for each map cell normalised to zero over the period 1981-2010 for a given month of the year.

A problem arises when coverage changes over time. Because the Arctic has warmed significantly since the end of the HadCRUT4 baseline period, a drop in Arctic

coverage leads to a cool bias in the mean of the observed cells. This effect increases as conditions diverge from the baseline period. To obtain realistic short term trends, the baseline period should be as similar to the trend period as possible. For similar reasons, when constructing a hybrid temperature series, the two source map series should have the same baseline period.

The HadCRUT4 map series was therefore renormalised to match the UAH baseline period of 1981-2010. For each map cell and each month of the year, the mean value of that cell during the baseline period was determined. If at least 15 of the 30 possible observations were present, an offset was applied to every value for that cell and month to bring the mean to zero; otherwise the cell was marked as unobserved for the whole period.

Renormalization is not a neutral step - coverage is very slightly reduced, however the impact of changes in coverage over recent periods is also reduced. Coverage of the renormalized HadCRUT4 map series is reduced by about 2%.

#### 3.2. Kriging

Kriging (Cressie 1990) is a linear approach to interpolation/extrapolation in which the values of the field are determined in accordance with a given covariance structure, where the covariance structure is determined from those observations which are present. It has been employed in temperature reconstructions, for example by Rigor *et al.* (2000) and Rohde *et al.* (2013b).

Kriging offers several benefits. The reconstructed values vary smoothly and match the observed values at the coordinates of the observations. The reconstructed values approach the global mean as the distance from the nearest observation increases, i.e. the method is conservative with respect to poor coverage. Clustered observations are downweighted in accordance with the amount of independent information they contribute to the reconstructed value; thus area weighting is an emergent property of the method, with observations being weighted by density in densely sampled regions and by the region over which the observation is informative in sparse regions.

The steps in a Kriging calculation are as follows:

- Determine the radial covariance function from the observations which are available.
- Construct a covariance matrix for the observed coordinates.
- Construct a vector of covariances between the observed coordinates and some coordinate at which an estimate of the field is desired.
- Solve the resulting system of equations to obtain a vector of weights. The estimated value of the field at the target coordinate is given by the dot product of vectors of weights and the observations.

In this work, the expectation of the field (i.e. the global mean temperature) is to be determined, and so ordinary kriging, which does not assume a value for the expectation, is required (Cressie 1990). The kriging calculation will be applied to the reconstruction of gridded temperature values using the grid cells for which observations are available rather than the more normal approach of using individual stations as the observations. This will enable a correspondence between the gridded satellite data and the

surface data, and also makes it computationally realistic to construct a full matrix of cell covariances allowing a global reconstruction from any starting data. In a 5 degree sampled grid there are 2592 cells, and so the correlation matrix can contain up to 8.7m (2592<sup>2</sup>) elements. The covariance function is modeled by a simple exponential function of distance, ensuring a strongly diagonal covariance matrix and a numerically stable calculation.

Kriging the gridded data also has some significant disadvantages: Information about station position within a cell is lost, cells with a single station receive the same weight as cells with many, and (equivalently) no account is taken of the uncertainty in a cell value. The acceptability of these compromises will become apparent in the validation step.

Grid cells for which observations are available are assumed to be exact, and are therefore unmodified by the calculation. This is not a necessary assumption of kriging, but for the purposes of estimating global mean temperatures it is a convenient simplification and means that the resulting temperature fields will preserve the features of the source data.

The covariance function is determined from a variogram, determined by calculating the square of the difference between every pair of observed temperatures in every month of the data from 1979-2012 and averaging the results in 300km radial bins. The bin values are converted to covariances and fitted using a function of the form  $\alpha \exp(-d'/d)$ , where  $d'$  is the great circle distance between any two observations,  $d$  is the 'range' of the variogram which controls the effective extrapolation distance, and  $\alpha$  is the 'sill' value, which is equal to the variance of the map.

### 3.3. The hybrid calculation

The hybrid surface-satellite temperature calculation is extremely simple in form:

$$T_x^{hybrid} = \text{krig}(T_x^{surf} - sT_x^{sat}) + sT_x^{sat} \quad (1)$$

where  $T_x^{surf}$  is the surface temperature of the grid cell at coordinate  $x$ ,  $T_x^{sat}$  is the satellite temperature, and  $s$  is a scale factor applied to the satellite temperatures. The difference between the surface and satellite temperatures may only be calculated for cells where surface temperatures are available, so the result has incomplete coverage, which is completed by the kriging operator. The resulting global coverage field is added to the scaled satellite temperatures to produce the hybrid temperature field. The scale factor  $s$  serves to allow for a difference in scale between surface and lower troposphere (LT) temperatures.

The hybrid field has the following properties:

1. Where a surface temperature observation is present, the hybrid field is equal to the surface temperature.
2. Near a surface temperature observation the value of the field will be similar to the nearest observed value.
3. Far from any surface temperature observation the value of the field will approach the value of the satellite field, adjusted by the difference in global mean between the surface and satellite fields.

The hybrid field around an isolated surface observation will match the satellite data in gradient and curvature, with a constant offset to fit the surface observation

at the grid centre. The distance over which a surface observation can dictate local temperatures is determined by the autocorrelation of the difference field itself.

Any temporal inhomogeneity in the satellite data is eliminated, because the satellite data is tied to the surface temperature observations on a month-by-month basis. Constant or time varying offsets between the surface and satellite data not already dealt with by the baselining step are also eliminated by equation 1. Spatial inhomogeneity, for example due to surface contamination of the tropospheric temperature observations, remains an issue only if inhomogeneity varies over distances shorter than the range of the kriging calculation.

## 4. Validation

The satellite data does not provide direct surface temperature observations, and thus must be treated as a proxy dataset. Also the kriging calculation implemented here does not account for uncertainties in the observations, or even the number of observations in a grid cell. Each of these factors mean that the method can not be assumed to be valid, and therefore the skill of the method must be proven by reconstructing temperature observations which have been hidden from the calculation. For the following tests, the HadCRUT4 ensemble median and UAH data were used over the period 1979/01-2012/12. The UAH data were upscaled to a 5x5 degree grid. Two methods are employed; the first being a 36-fold cross validation approach modified to deal with the spatial autocorrelation of the data.

In the cross validation calculation, a contiguous group of either 1, 3, or 5 latitude bands are omitted from the temperature map data for a given month. The central latitude band of the omitted region is then reconstructed by one of the following methods:

- Null reconstruction - the target cells are set to the (area weighted) mean of the rest of the map.
- Kriging - the target cells are reconstructed using the kriging calculation.
- Hybrid - the target cells are reconstructed using the hybrid method. This approach is repeated for different values of the satellite scale factor  $s$ .

The calculation is repeated 36 times, omitting the latitude bands bracketing each of the 36 latitude bands in turn. The 36 reconstructed bands are then combined to create a composite map in which every cell has been reconstructed using only cells more than a certain distance from that cell.

The skill of the temperature reconstruction methods may be measured in terms of the RMS difference between the reconstructed and original map over all the months in the reconstruction - the results are shown in Figure (3). Maps are shown for the for the null reconstruction, kriging, and the hybrid method with  $s = 1.0$  for the three extrapolation ranges. (For the null reconstruction the range makes no difference to the results.)

The null reconstruction provides surprisingly good predictions over the oceans (apart from the El Nino region), which reflects the fact that sea surface temperatures vary much less from the global mean than land temperatures. The agreement over land however is poor, as expected.

The kriging reconstruction is better over both land and oceans for ranges of 1 and 2 cells. At longer ranges the

North Pacific SST reconstructions begin to be distorted by land temperatures from Siberia and Alaska, however the reconstructed land temperatures remain superior to the null reconstruction.

The hybrid reconstructions shows different behaviour. For a range of 1 cell, the results are very similar to kriging, but as the range increases the land and SST reconstructions deteriorate at high latitudes at a uniform rate over land and ocean. Thus the land reconstruction is better, while the SST reconstruction is worse than either the kriging or null reconstructions.

The reconstructions are compared numerically in Table I, which shows the RMS difference between the original and reconstructed maps averaged over all the months in the temperature series. Results are given for the null and kriging reconstructions, and for the hybrid calculation with  $s$  varying from 0.2 to 1.4. Best results are obtained for  $s \approx 0.6$ , representing a compromise between the improving land reconstruction and deteriorating SST reconstruction. The value of the RMS difference is bounded by the noise level in the grid values.

Of more interest in this work are the accuracy of the global temperature reconstructions. The previous results do not directly address this question because the regions of the globe affected by poor coverage are not uniformly distributed. Since no observations are available for these regions, an estimate must be made on the basis of the error in reconstructing temperatures at the boundaries of the unobserved regions.

Three test cases are therefore constructed by further reducing the coverage of the HadCRUT4 data at the edges of the unobserved regions. A mask is applied to remove the values from cells whose centres are within 600, 1150 or 1700km of a cell with no observations. The masked cells are reconstructed by the three methods used earlier, and the reconstructed values compared to the observations. The 600km dataset involves omitting an additional 16% of the globe from the observed data, comparable to the 18% already missing from that data.

A difference map is calculated between the original and reconstructed temperatures for the masked cells. The mean and RMS of this map give a measure of the bias and error in reconstructing cells in the geographical regions of interest over a range of distances from the nearest observation. For each test the bias (measured by the area weighted mean of the difference between the original and reconstructed values), and the error (measured by the area weighted RMS difference between the original and reconstructed values) are presented for the period from 2005 on when bias is expected to be critical.

The results of reconstructing the omitted cells are shown in Table (II). Kriging gives a lower RMS error over the masked region than the null reconstruction in every case, although its performance degrades as the extrapolation range increases, and the bias is variable. The hybrid method gives both a better mean bias and RMS error than kriging, and the results degrade more slowly with increasing extrapolation range. The optimal scale factor  $s$  for the satellite data is in the range 0.8 – 1.2 in each case.

These results indicate that reconstruction of the regions of the globe where coverage is poor is best achieved with the hybrid method, using a scale for the satellite data in the region of 1.0. This is in contrast with the result for the globe in general, where optimal results are achieved by a

combination of the kriging and hybrid methods to cover land and ocean, or failing that a hybrid calculation with the satellite data down-weighted.

## 5. Reconstruction issues by domain

It is clear from Figure (3) that land and sea surface temperatures show different behaviours, and should ideally be modeled separately (as in Smith *et al.* 2008). The cross-validation calculation was therefore repeated for the land and SST data separately. The implications are considered for land, sea and sea ice domains.

### 5.1. Land temperatures

The surface air temperature over land is well modeled by the hybrid method. When the hybrid calculation is performed with land temperatures only, the range of the variogram is 560km (compared to 680km with land-ocean data), and the optimum cross validation results are obtained for a scale factor  $s = 1.0$ .

As has been noted the MSU lower troposphere data is potentially subject to biases over ice and at high altitudes. The use of anomalies for both surface and satellite data eliminates the effect of a constant offset, however differences in temperature variation over ice or high altitude surfaces could bias the results. The fact that the hybrid method gives better cross validation results than kriging for the Antarctic and high Arctic land stations suggests that this is not a major issue. Furthermore there is no obvious evidence of worse performance at high altitude in Figure (3).

### 5.2. Sea surface temperatures

Sea surface temperatures are better modeled by ordinary kriging. When the calculation is performed with sea surface temperatures only, the range of the variogram is 915km (compared to 830km with land-ocean data), and the optimum cross validation results are obtained for  $s = 0.0$  (i.e. ordinary kriging) or  $s = 0.2$ ; this is consistent with the poor correlation between SST and satellite temperatures (section 3). However the kriging results are only marginally better than the null reconstruction.

Given the difference in the optimum value of  $s$ , is it reasonable to use a single approach for the land and ocean data? For short extrapolation ranges (e.g. one cell in latitude or 550km) the difference between kriging and the hybrid method is small, and at mid latitudes the unobserved regions in the SST data tend to be small and isolated, and thus the choice of infilling method make little difference.

The only large contiguous unobserved regions in the SST data are in the Arctic and Southern oceans. These regions are also characterised by seasonal or perennial sea ice, which must be considered separately.

### 5.3. Air temperatures over sea ice

Sea ice represents a special case. Sea ice, especially when covered in snow, can effectively insulate the air from the sea below (Kurtz *et al.* 2011), and therefore the water temperature cannot be considered to be a surface temperature. One alternative is to treat sea ice as land, since to the atmosphere snow covered ice is similar to snow covered land. This happens implicitly for the Arctic

when using the blended land-ocean data because the nearest observations are from land stations.

The International Arctic Buoy Program (IABP/POLES, Rigor *et al.* 2000) provides gridded data from ice buoys and land stations from 1979-1998. Rigor *et al.* found significant correlation between land stations and air temperatures over ice at ranges of up to 1000km in winter, when most of the bias occurs. The land and ocean reconstructions were therefore tested against both this data and the NCEP/NCAR and ERA reanalyses.

Temperatures were compared for the group of 26 cells centered around  $83^{\circ}N172.5^{\circ}E$ , which are most isolated from the nearest weather station with an average distance of  $\approx 850$ km. Three years of temperature estimates for this region from buoys, reanalysis, and the land-only and SST-only reconstructions are shown in Figure (4a). The land-only reconstruction for this region is very similar to the hybrid reconstruction from the blended data because the nearest observations are from land stations.

The land-only reconstruction is more realistic than the SST-only reconstruction in terms of month-on-month variation. Figure (4b) shows a 60 month moving average over the whole period, which reveals that the IABP and NCEP/NCAR reanalysis data show significantly more long term variation than even the land-only reconstruction. This might suggest that the hybrid reconstruction is too conservative, however the ERA-interim reanalysis (Dee *et al.* 2011) shows less long term variation and is broadly consistent with the land-only reconstruction, so the discrepancies may lie in the IABP homogenisations and NCEP/NCAR model.

On this basis a hybrid reconstruction based on either land-only or blended data appears to be the most realistic reconstruction for the central Arctic. This is consistent with the earlier cross-validation results for the high Arctic stations in Figure (3).

The unobserved region of the Southern ocean includes both sea ice and open water. The open water will be well modeled close to the edges of the observed region, and the air temperatures over ice are expected to be well modeled from the land stations by analogy with the Arctic. Some ocean cells are likely to be poorly modeled, however the slower rate of change in the Antarctic climate means that this is unlikely to significantly bias the results.

These results support the use of separate land and ocean reconstructions, with sea ice treated as land. However this would introduce a significant additional problem: as Arctic ice cover declines, some cells will change status from land to ocean. Given that anomalies cannot be calculated against the same baseline period for such cells, the change in status introduces an unknown temperature bias. Future work will attempt to place bounds on this uncertainty, however in this work the problem is avoided by use of the blended data.

## 6. Global reconstruction results

The kriging and hybrid methods have been applied to the full HadCRUT4 ensemble median data to obtain global temperature reconstructions. A global mean temperature estimate is then calculated using an area weighted mean of the map cells. The results are compared over the period of the satellite data in Figure (5) using 12 and 60 month moving averages. The kriging and hybrid reconstructions are compared to the null reconstruction, which corresponds to the HadCRUT4 data except in that the baseline period has

been adjusted and a global mean is calculated instead of the Hadley practice of calculating the mean of the hemispheric means (the latter change having minimal effect).

The kriging and hybrid reconstructions give very consistent results over most of the satellite era, but show divergence from the null series over parts of the record. Of particular interest are the periods 2005-present when the new reconstructions show warmer temperatures than the null series, and 1997-2000 when the new reconstructions are cooler than the null series, with the hybrid results showing cooler temperatures than kriging. The 60 month moving average shows the impact of coverage, both on the 1998 peak and more strongly on recent temperatures. Note that all temperatures are relative to the baseline period which includes the recent period of cool bias, and so the apparent warm bias in the late 1990s is relative.

How do different regions of the planet contribute to coverage bias? The impact of coverage is calculated by masking unobserved cells in the hybrid reconstruction in three latitude bands. The result provides a measure of coverage bias due to limited coverage in each band in turn. The results are shown in Figure (6) using a 60 month moving average. Incomplete coverage of the rapidly warming Arctic is the principal cause of coverage bias since 2005, despite the comparatively limited area affected. The Antarctic shows much more variability on short time scales owing to the larger area affected, however there is less trend. Notable however is a relative warm bias in the late 1990s. The rest of the world (of which central Africa is the primary region of poor coverage) contributes comparatively little temperature bias.

Trends from 1997 to the present are particularly impacted by coverage bias (and have also been the subject of significant media coverage, see for example The Telegraph 2013). The trends over this period have therefore been calculated for the three series and are given in Table (III), along with corresponding trends for the HadCRUT4, NCDC, GISTEMP and NCEP/NCAR temperature data. Both the kriging and hybrid series yield similar trends and both show faster warming than GISTEMP. The difference in trend between the original HadCRUT4 data and the null reconstruction is apparent, and arises from the reduction in bias due to changes in coverage over the trend period as discussed in section (3.1).

The NCEP/NCAR trend is higher than for the observational records. Most of this difference comes from mid-latitudes and from ocean rather than land data (see supporting information), however the high latitude trends are in good agreement with the hybrid reconstruction. The increased trend in the kriging reconstruction in comparison to GISTEMP is probably related to the additional corrections in the HadSST3 data.

The coverage bias in the HadCRUT4 trends estimated using the hybrid data is shown in Table (IV). The trend bias is maximised for starting dates around 1998, and then again around 2003. Also of interest is the effect on the statistical significance of the trend, which increases as the number of months raised to the power  $3/2$ . A rough estimate of this impact may be calculated from the ratio of the bias trend to the uncertainty in the temperature trend. By this metric temperature trends starting in 1997 are most likely to give a misleading result with respect to the statistical significance of the resulting trend.

The impact of coverage bias on the 1997-2012 trend is greatest in the winter and smallest in the summer: The

estimated bias in the trend is  $-0.086^{\circ}\text{C decade}^{-1}$  (DJF),  $-0.051^{\circ}\text{C decade}^{-1}$  (MAM),  $-0.026^{\circ}\text{C decade}^{-1}$  (JJA) and  $-0.055^{\circ}\text{C decade}^{-1}$  (SON).

### 6.1. Uncertainty in the global temperature reconstruction

The uncertainty in the HadCRUT4 global temperature reconstruction arises from a number of sources and is discussed in detail by Morice *et al.* (2012). Since the reconstructions presented here preserve the map cell values for cells where HadCRUT4 has data, most of those uncertainties are unaffected by this analysis.

The principal difference comes in the coverage uncertainty term. Morice *et al.* estimate this by reducing the coverage of the NCEP/NCAR data to match the HadCRUT4 data for every available month from the reanalysis and determining the error in the resulting global temperature estimate. A similar approach may be applied for the kriging method, however the hybrid method is problematic in that it is dependent on the satellite temperature observations. Further the HadCRUT4 approach uses reanalysis data from months which can be substantially separated in time, and thus in climatic norms, from the month under consideration.

To address this issue and enable uncertainty estimation for all the reconstructions, a modification of the HadCRUT4 approach was adopted. For each month, the coverage of the NCEP/NCAR data for the corresponding month was reduced in coverage to match the HadCRUT4 data, and then restored by each of the three methods (using the satellite data for that month in the case of the hybrid method). The error in the resulting global mean temperature estimate was determined, and an estimate of the uncertainty obtained from a 60 month moving RMS of the errors centered on the month under consideration. At the start and end of the record the number of samples reduces to 30. A disadvantage of this approach is that it masks the annual cycle in the coverage uncertainty.

The mean of the  $2\sigma$  uncertainties for the period 1979-2012 is  $0.155^{\circ}\text{C}$  for the null reconstruction,  $0.065^{\circ}\text{C}$  for the kriging reconstruction, and  $0.056^{\circ}\text{C}$  for the hybrid reconstruction. The uncertainty in the null reconstruction is very similar to the estimates of Morice *et al.* (2012). The kriging and hybrid reconstructions produce a substantial reduction in the coverage uncertainty.

Uncertainties in the resulting temperature maps are determined by range and sill parameters of the covariance function. For the kriging calculation the range  $d$  is approximately 830km with a sill of  $3.5^{\circ}\text{C}^2$ . For the difference map used in the hybrid calculation,  $d$  is 680km with a sill of  $2.3^{\circ}\text{C}^2$ . Despite the shorter range the lower sill value means that the uncertainty of the hybrid reconstruction is always lower than for kriging, independent of distance from the nearest observation.

#### 6.1.1. Other sources of uncertainty

HadCRUT4 consists of an ensemble of 100 realisations of the temperature record providing a sampling of the correlated uncertainties in processing the data. The potential impact of these uncertainties on the global temperature reconstruction was investigated by calculating the hybrid reconstruction from each member of the ensemble in turn. The mean of the trends for the period 1997-2012 over the 100 realisations is  $0.120^{\circ}\text{C decade}^{-1}$ , with a standard deviation of  $0.007^{\circ}\text{C decade}^{-1}$ . The contribution of the

ensemble spread to the uncertainty in the trend is therefore negligible.

## 7. Conclusions

The existence of bias in recent global mean temperature estimates has been confirmed by multiple means. This bias leads to an underestimation of recent temperature trends. The evidence is as follows:

1. Microwave sounding data from satellites, reanalysis data from weather models, and observations from isolated weather stations all confirm that the unobserved regions of the planet, and in particular the Arctic, have been warming faster than the globe as a whole since the late 1990s.
2. Extrapolation by kriging and a hybrid method guided by the satellite data have been examined. Both provide good temperature reconstructions at short ranges. Over longer ranges the hybrid method performs well over land and kriging performs well for SSTs.
3. Extrapolation over land/ocean boundaries is problematic, however observations and reanalyses confirm that air temperatures over ice are better reconstructed from land-based air temperature readings. Since the highest latitude observations are land based, reconstruction from the blended land/ocean data is realistic.
4. The application of either kriging or the hybrid method to coverage-reduced reanalysis data produces global temperature estimates with substantially reduced uncertainties than simply omitting the unobserved regions.

The main benefit of the hybrid method is to bring observational data to bear on the question of coverage bias. However the method is dependent on the satellite data and so only applicable from 1979. Given that the hybrid results are not very different from the kriging results, simple extrapolation appears to be justified for current levels of global coverage. In practice the choice of extrapolation method makes little difference once the Antarctic stations are available: Kriging, the GISTEMP kernel smoothing method, inverse distance weighting and even basic nearest neighbour give very similar results, especially used in combination with a 1200km cutoff as employed by GISTEMP (see supporting information). For periods prior to the establishment of the Antarctic stations, coverage is less complete and coverage bias remains an issue.

The NCDC temperature series has similar coverage issues around the poles to the HadCRUT4 data, although coverage at low latitudes is better. The mean coverage above 60N since 1979 is 63% for the NCDC data compared to 65% for HadCRUT4. Since most of the bias comes from higher latitudes, recent trends in the NCDC data are expected to be similarly impacted to the HadCRUT4 trends.

The Arctic has experienced a very rapid temperature change over recent years through a combination of polar amplification of greenhouse warming, albedo change due to both black carbon and snow/ice loss and possibly a contribution from multidecadal variability (AMAP 2011; Semenov *et al.* 2010). The pace of this change means that Arctic coverage has dominated bias in the global temperature estimates, despite the unobserved region being

rather smaller than in the Antarctic. On this basis the problem of coverage bias may exist whenever the Arctic has experienced rapid warming (or cooling) in the past. However, given the magnitude of recent Arctic warming and polar amplification relative to global trends, it is expected those previous periods of warming (or cooling) in the Arctic are unlikely to bias the record to a greater extent than the bias recorded in recent decades by this study.

The Arctic is likewise the primary source of uncertainty in this work. Further work is needed to address the differing behaviour of the land and ocean domains. The need for a continuous climatology for cells which transform from ice to open ocean presents a significant impediment, however it may be possible to place bounds on results by assuming constant high and low ice coverage. The Antarctic is the only region in which the kriging and hybrid temperature estimates are noticeably different. Further investigation into the origins of the difference are required.

It is hoped that the preliminary global temperature reconstructions presented here, by highlighting the potential scale of the bias in the short-term temperature trends, will provide an impetus for other groups to look at the problem using more sophisticated tools including climate and reanalysis models.

Data and methods for this paper are available at <http://www-users.york.ac.uk/~kdc3/papers/coverage2013>.

## Acknowledgements

This work was produced without funding in the authors' own time, however KC is grateful to the University of York for providing computing facilities, and to the organisers of the 2013 EarthTemp network meeting (NERC Grant NE/I030127/1) for enabling him to benefit from the expertise of the other attendees. The authors would also like to acknowledge the reviewers for their invaluable comments, the online community of professional and amateur climate scientists who have provided advice over the 18 months of the work, and also John Kennedy at the Hadley Centre who provided useful feedback on some very rudimentary initial results.

## References

- AMAP. 2011. Snow, water, ice and permafrost in the arctic (swipa). *Oslo: Arctic Monitoring and Assessment Programme (AMAP)*.
- Brohan P, Kennedy JJ, Harris I, Tett SF, Jones PD. 2006. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.: Atmospheres* **111**(D12).
- Christy JR, Norris WB, Spencer RW, Hnilo JJ. 2007. Tropospheric temperature change since 1979 from tropical radiosonde and satellite measurements. *J. Geophys. Res.* **112**(D6): D06 102.
- Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan R, Yin X, Gleason B, Vose R, Rutledge G, Bessemoulin P, et al. 2011. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**(654): 1–28.
- Cressie N. 1990. The origins of kriging. *Mathematical Geology* **22**(3): 239–252.
- Daily Mail. 2012. Global warming stopped 16 years ago, reveals met office report quietly released... and here is the chart to prove it. URL <http://www.dailymail.co.uk/sciencetech/article-2217286/Global-warming-stopped-16-years-ago-reveals-Met-Office-report-quietly-released-chart-prove-it.html>. Accessed: 2013-07-31. (Archived by WebCite® at <http://www.webcitation.org/6IWYr0dnc>).
- Dee D, Uppala S, Simmons A, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M, Balsamo G, Bauer P, et al. 2011. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**(656): 553–597.
- Folland CK, Colman AW, Smith DM, Boucher O, Parker DE, Vernier JP. 2013. High predictive skill of global surface temperature a year ahead. *Geophys. Res. Lett.* **40**: 761–767, doi:10.1002/grl.50169.
- Foster G, Rahmstorf S. 2011. Global temperature evolution 1979–2010. *Environmental Research Letters* **6**(4): 044 022.
- Global Warming Policy Foundation. 2012. No underlying global warming in recent years. URL <http://www.thegwpf.org/no-underlying-global-warming-in-recent-years/>. Accessed: 2013-07-31. (Archived by WebCite® at <http://www.webcitation.org/6IWYggUyw>).
- Hansen J, Ruedy R, Sato M, Lo K. 2010. Global surface temperature change. *Reviews of Geophysics* **48**(4): RG4004.
- Hansen J, Sato M, Ruedy R, Lo K, Lea DW, Medina-Elizade M. 2006. Global temperature change. *P. Natl Acad. Sci. USA* **103**(39): 14 288–14 293.
- Hausfather Z, Menne MJ, Williams CN, Masters T, Broberg R, Jones D. 2013. Quantifying the effect of urbanization on us historical climatology network temperature records. *J. Geophys. Res.: Atmospheres* doi:10.1029/2012JD018509. In press.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, et al. 1996. The ncep/ncar 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.* **77**(3): 437–471.
- Karl TR, Hassol SJ, Miller CD, Murray WL. 2006. Temperature trends in the lower atmosphere. steps for understanding and reconciling differences.
- Kaufmann RK, Kauppi H, Mann ML, Stock JH. 2011. Reconciling anthropogenic climate change with observed temperature 1998–2008. *P. Natl Acad. Sci. USA* **108**(29): 11 790–11 793.
- Kennedy J, Rayner N, Smith R, Parker D, Saunby M. 2011. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. biases and homogenization. *J. Geophys. Res.* **116**(D14): D14 104.
- Kurtz N, Markus T, Farrell S, Worthen D, Boisvert L. 2011. Observations of recent arctic sea ice volume loss and its impact on ocean-atmosphere energy exchange and ice production. *J. Geophys. Res.: Oceans* **116**(C4).
- Mears CA, Schabel MC, Wentz FJ. 2003. A reanalysis of the msu channel 2 tropospheric temperature record. *J. Climate* **16**(22): 3650–3664.
- Mears CA, Wentz FJ. 2009. Construction of the remote sensing systems v3. 2 atmospheric temperature records from the msu and amsu microwave sounders. *J. Atmos. Oceanic Technol.* **26**(6): 1040–1056.
- Meehl GA, Arblaster JM, Fasullo JT, Hu A, Trenberth KE. 2011. Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change* **1**(7): 360–364.
- Morice CP, Kennedy JJ, Rayner NA, Jones PD. 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *J. Geophys. Res.* **117**(D8): D08 101.
- Rigor IG, Colony RL, Martin S. 2000. Variations in surface air temperature observations in the arctic, 1979–97. *J. Climate* **13**(5): 896–914.
- Rohde R, Muller RA, Jacobsen R, Muller E, Perlmutter S, Rosenfeld A, Wurtele J, Groom D, Wickham C. 2013a. A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview* doi:10.4172/gigs.1000101.
- Rohde R, Muller RA, Jacobsen R, Perlmutter S, Rosenfeld A, Wurtele J, Curry J, Wickham C, S M. 2013b. Berkeley earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview* doi:10.4172/gigs.1000103.
- Semenov VA, Latif M, Dommengot D, Keenlyside NS, Strehz A, Martin T, Park W. 2010. The impact of north atlantic-arctic multidecadal variability on northern hemisphere surface air temperature. *J. Climate* **23**(21): 5668–5677.
- Simmons A, Willett K, Jones P, Thorne P, Dee D. 2010. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res.* **115**(D1).
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J. 2008. Improvements to noaa's historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate* **21**(10): 2283–2296.
- Spencer RW. 1990. Precise monitoring of global temperature trends. *Science* **247**: 1558–1558.
- The Guardian. 2012. URL <http://www.theguardian.com/environment/2012/oct/16/daily-mail-global-warming-stopped-wrong>. Accessed: 2013-10-18. (Archived

by WebCite® at <http://www.webcitation.org/6KSfOtbyC>).

The Telegraph. 2013. Look at the graph to see the evidence of global warming. URL <http://www.telegraph.co.uk/earth/environment/globalwarming/9919121/Look-at-the-graph-to-see-the-evidence-of-global-warming.html>. Accessed: 2013-07-31. (Archived by WebCite® at <http://www.webcitation.org/6IWyxHHON>).

Williams CN, Menne MJ, Thorne PW. 2012. Benchmarking the performance of pairwise homogenization of surface temperatures in the united states. *J. Geophys. Res.* **117**(D5): D05 116.

Table I. RMS difference between original and reconstructed cell temperatures calculated over all observed cells when omitting one or more rows of data and reconstructing the central row from rows separated by the specified distance.

Method	RMS error in reconstruction ( $^{\circ}\text{C}$ )		
	1 cell (550km)	2 cells (1100km)	3 cells (1650km)
Null	1.07	1.08	1.08
Krig	0.68	0.90	1.03
Hybrid 0.2	0.67	0.86	0.96
Hybrid 0.4	0.67	0.84	0.91
Hybrid 0.6	0.67	0.83	0.89
Hybrid 0.8	0.67	0.84	0.90
Hybrid 1.0	0.68	0.87	0.94
Hybrid 1.2	0.69	0.92	1.02
Hybrid 1.4	0.70	0.97	1.12

Table II. Mean bias and RMS error between original and reconstructed global temperatures calculated over the *omitted cells* using the reconstructed values for the omitted cells. Results are for the period 2005/01-2012/12 and are given for the maps with coverage reduced by 600km, 1150km and 1700km.

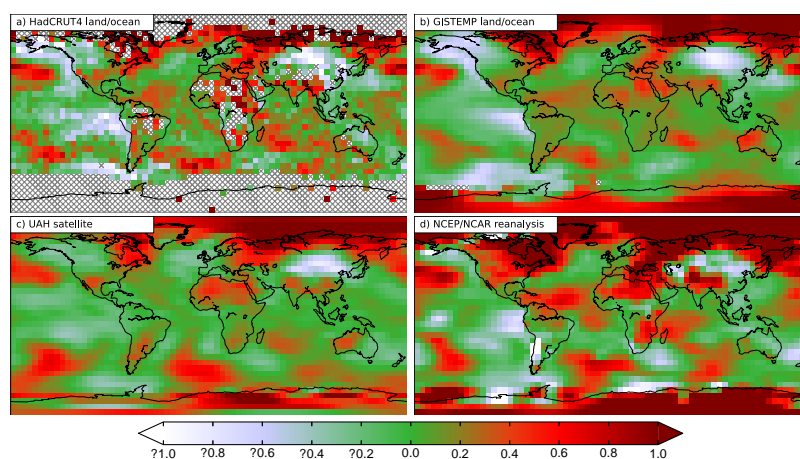
Method	Mean bias ( $^{\circ}\text{C}$ )			RMS error ( $^{\circ}\text{C}$ )		
	600km	1150km	1700km	600km	1150km	1700km
Null	-0.021	-0.027	-0.022	0.043	0.074	0.091
Krig	-0.002	-0.026	-0.023	0.023	0.064	0.085
Hybrid 0.2	-0.002	-0.022	-0.013	0.021	0.055	0.069
Hybrid 0.4	-0.014	-0.017	-0.010	0.023	0.047	0.054
Hybrid 0.6	-0.011	-0.013	-0.012	0.021	0.040	0.054
Hybrid 0.8	-0.002	-0.008	-0.007	0.017	0.035	0.050
Hybrid 1.0	-0.007	-0.003	-0.005	0.018	0.033	0.051
Hybrid 1.2	-0.002	0.001	-0.001	0.017	0.035	0.056
Hybrid 1.4	-0.001	-0.004	0.003	0.017	0.047	0.064

Table III. Temperature trend in  $^{\circ}\text{C decade}^{-1}$  on the 16 year period 1997/1 to 2012/12 for the GISTEMP, NOAA and HadCRUT4 temperature series and in the null, kriging and hybrid reconstructions. The standard error in the trend is calculated assuming an ARMA(1,1) error model using the method in the appendix of Foster and Rahmstorf (2011) (however in contrast to that work the temperature series have not been adjusted for exogenous influences).

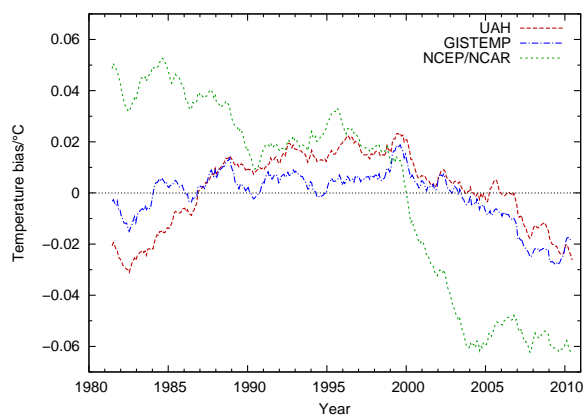
Dataset	Trend $\pm \sigma$
NCEP/NCAR	$0.178 \pm 0.107$
GISTEMP	$0.080 \pm 0.067$
NOAA	$0.043 \pm 0.062$
HadCRUT4	$0.046 \pm 0.063$
Null reconstruction	$0.064 \pm 0.078$
Kriging	$0.108 \pm 0.073$
Hybrid $s = 1.0$	$0.119 \pm 0.076$

Table IV. Bias in HadCRUT4 temperature trends running from various dates to the present, estimated using the hybrid data ( $s = 1.0$ ), in units of  $^{\circ}\text{C decade}^{-1}$ . The impact of the bias on the significance of the trend is given in the third column; this is the trend bias in units of uncertainty ( $\sigma$ ) of the trend, and measures how much significance test on the trend will be impacted by the bias.

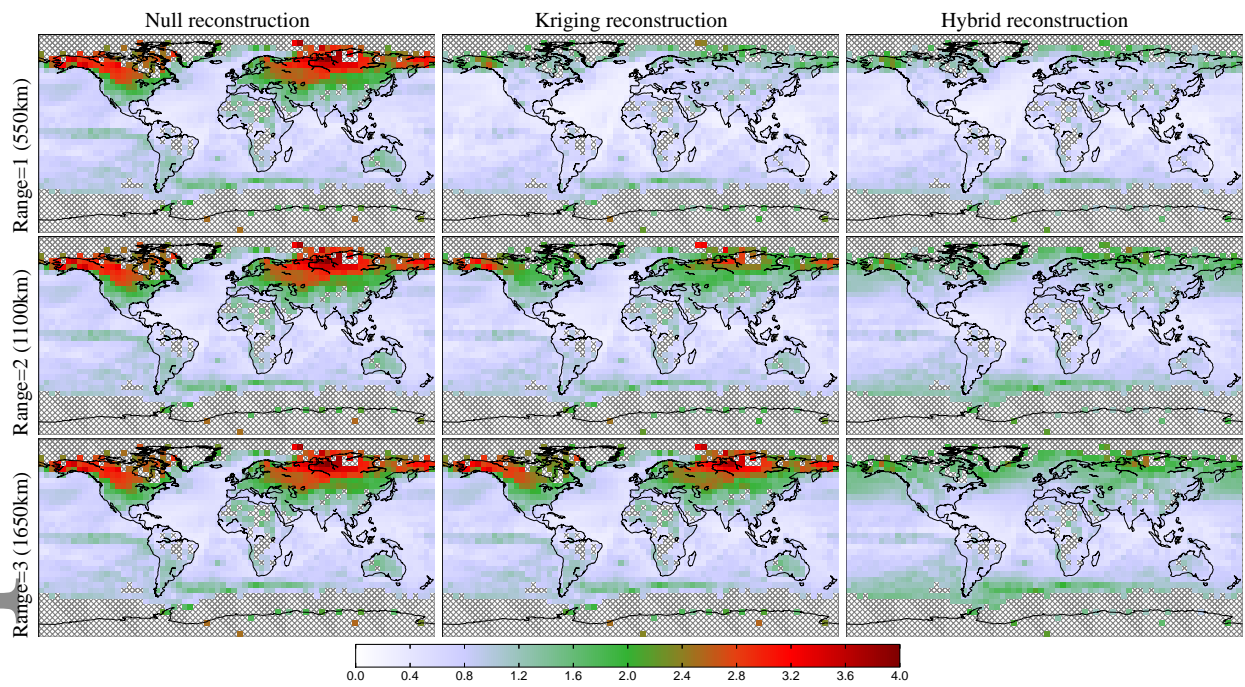
Start year	Trend bias	Trend bias in $\sigma s$
1990	-0.020	-0.39
1991	-0.020	-0.38
1992	-0.027	-0.47
1993	-0.030	-0.50
1994	-0.034	-0.54
1995	-0.036	-0.52
1996	-0.039	-0.51
1997	-0.055	-0.70
1998	-0.058	-0.67
1999	-0.056	-0.60
2000	-0.055	-0.54
2001	-0.057	-0.53
2002	-0.056	-0.46
2003	-0.083	-0.58
2004	-0.081	-0.48
2005	-0.045	-0.22



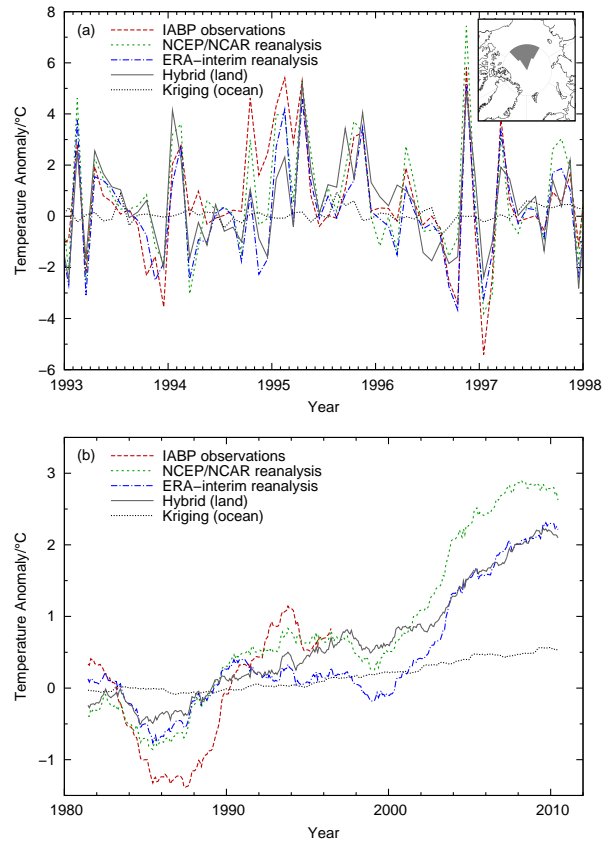
**Figure 1.** Temperature trends over the 16 year period 1997/1-2012/12 in  $^{\circ}\text{C decade}^{-1}$  for HadCRUT4 and three near global reconstructions: GISTEMP extrapolated surface temperatures, UAH satellite data and NCEP/NCAR reanalysis data. Areas with no coverage are shown with hatching. Note that the cylindrical projection exaggerates the missing area at high latitudes.



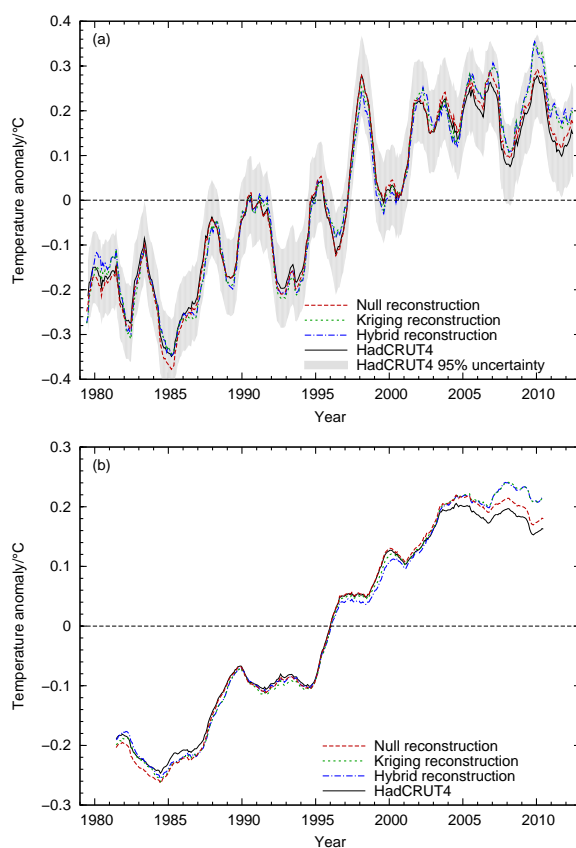
**Figure 2.** Potential coverage bias in the HadCRUT4 data estimated using three global temperature series. A 60 month moving average has been applied to the data.



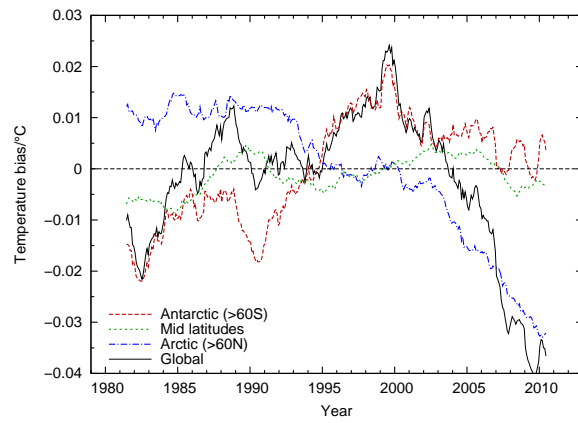
**Figure 3.** RMS difference in  $^{\circ}\text{C}$  between observed temperatures and 36-fold cross validated reconstruction, omitting different numbers of latitude bands to control the minimum extrapolation range.



**Figure 4.** Mean temperature anomaly for an isolated region of the central Arctic (see right inset) determined from observational data (IABP/POLES), NCEP/NCAR reanalysis, ERA-interim reanalysis, hybrid reconstruction from land-only data and kriging from SST-only data. (a) Monthly data on the period 1994-1998. (b) 60 month moving average on the period 1979-2012. (For additional monthly data see supporting information.)



**Figure 5.** Comparison of null, kriging and hybrid reconstructions of the HadCRUT4 data over the period 1979/1-2012/12. The data are shown with (a) a 12 month moving average and (b) a 60 month moving average. The original HadCRUT4 record before rebaselining of the map series (but offset to match the other series) is also shown, including uncertainty bounds in the upper figure.



**Figure 6.** Coverage bias in the HadCRUT4 global mean (rather than mean of the hemispheric means) estimated using the hybrid ( $s = 1.0$ ) reconstruction. Contributions are shown for three latitude bands and for the whole globe. The data are shown with a 60 month moving average.