NEW SYSTEMATIC ERRORS IN ANOMALIES OF GLOBAL MEAN TEMPERATURE TIME-SERIES

by

Michael Limburg (Germany)

Reprinted from

ENERGY &
ENVIRONMENT

VOLUME 25 No. 1 2014

MULTI-SCIENCE PUBLISHING CO. LTD.
5 Wates Way, Brentwood, Essex CM15 9TB, United Kingdom

NEW SYSTEMATIC ERRORS IN ANOMALIES OF GLOBAL MEAN TEMPERATURE TIME-SERIES

Michael Limburg
email: limburg@grafik-system.de

ABSTRACT
Existing uncertainty assessments and mathematical models used for error estimation of global average temperature anomalies are examined. The error assessment model of Brohan et al 06 [1] was found not to describe the reality comprehensively and precisely enough. This was already shown for some type of errors by Frank [2];[3] hereinafter named F 10 and F 11. In addition to the findings in both papers by Frank a very common but new systematic error was isolated and defined here named “algorithm error” This error was so far regarded as self canceling or corrected by some unknown and unnamed homogenization processes. But this was not the case. It adds therefore a minimum additional systematic uncertainty of + 0.3 °C and – 0.23°C respectively to any global mean anomaly calculation. This result is obtained when comparing only the most used algorithms against a “true” algorithm of measuring the daily temperature continuously. Due to the fact, that the real distribution of all applied algorithms (acc. as Griffith [4] showed is > 100) over time and space is not known, neither for land based temperature data nor for SST (Sea Surface Temperatures) the minimum value of said error is chosen here.

INTRODUCTION

General remarks

For the purpose of clarity the first part of the paper is used to explain the general behavior of anomaly calculation in respect of error propagation. It is shown that the widely assumed error reduction capabilities of an anomaly model is valid in one special case only, but in general may not reduce the final systematic error – especially in time series – but in most cases increases it. Further a great variety of further potential systematic errors are named here, from which only very few had been quantified and could be corrected, and so far only in part. This is shown also. By knowing this the minimum uncertainty for every annual global mean temperature should be expanded not only to the value described here i.e. with 95 % confidence interval to ± 1.084 °C, but should be at least 3 to 5 times wider. Thus, the average global temperature anomaly for the last 150 years is dissolved in a wide noisy uncertainty band, which is much wider than the whole assumed variation of the 20th century. Therefore every attempt to attribute any possible forcing to that variation remains scientific speculation. The only but very important exception may be the influence of a strong driving force which oscillates around a given mean. Its oscillating signal might be much more easily discriminated from the uncertainty band described before, due to its repetitive nature

Determination of mean temperature for local and worldwide time series construction

Local air temperatures had been observed for meteorological purposes for more than 300 years in a couple places of the world (for example at Berlin; Germany since 1701). Daily air temperatures on land are recorded by measuring them at fixed times and locations. In order to prevent disturbing influences by direct solar radiation, wind flow and other perturbances the used thermometer is shielded against its environment and housed in a weather shed or screen. It has been known from the outset that this shielding has its own properties, which influence the measured temperature, thus does not allow measuring the true air temperature in which meteorologists are really interested. But the need for an acceptable compromise leads to this construction by accepting the error, which are created by this compromise. In addition to this known but not corrected errors introduced by the use of screens all over the world and over time, different regimes (for enabling the algorithms for calculation of a mean temperature) of picking data as well as screen designs had been used.

Since one aims to extract climate relevant signals presumed inherent in locally measured data the requirement for a temperature station is that it is better to use longer rather than shorter measurement time series and as more continuously it measures the better it is. For better recognition and comparison of this, signals anomalies are calculated. i.e. that from the observed data a suitable reference value will be subtracted. In this paper special attention is given to the behavior of those anomalies.

If a station has been in existence for a longer period of time, for example for the last 50 better 120 to 150 years of continuous measurement, then a time series and anomaly calculation is useful for every single month or year. The result may be meaningfully blended with others into a global mean time series as shown in Fig.1. As shown, every anomaly includes also its own uncertainty, as indicated by the gray confidence range in Figure 1[1] retrieved April 21, 2012, from http://www.cru.uea.ac.uk/cru/data/temperature/#sciref. The method used to calculate anomalies is described by the British Met Office, and other organizations.

In the data available for example from the Met-Office web page (footnote 1) a number of those errors or uncertainties are defined. Such uncertainties are inherent in every measurement process and do not differ in kind or class from those in other areas

[1] Source http://www.cru.uea.ac.uk/cru/data/temperature/#sciref 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. Geophysical Research* **111**, D12106, — Jones, P.D., New, M., Parker, D.E., Martin, S. and Rigor, I.G., 1999: Surface air temperature and its variations over the last 150 years. *Reviews of Geophysics* **37**, 173-199. Rayner, N.A., P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T. Ansell and S.F.B. Tett, 2006: Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *J. Climate*, **19**, 446-469. Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V, Rowell, D.P., Kent, E.C. and Kaplan, A., 2003: Globally complete analyses of sea surface temperature, sea ice and night marine air temperature, 1871-2000. *J. Geophysical Research* **108**, 4407,

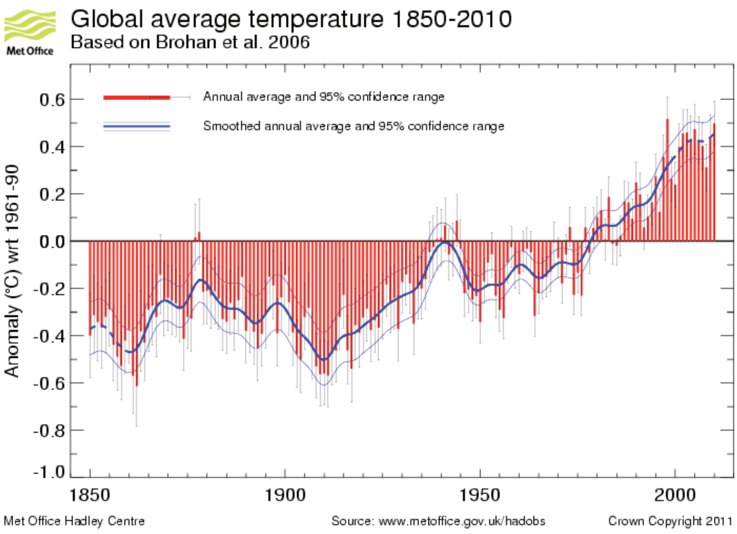

Figure 1 Global average temperature anomaly from 1850 to 2010 as published on MetOffice Webpage. The red bars show the global annual average near-surface temperature anomalies from 1850 to 2009. The uncertainty bars show the 95% uncertainty range on the annual averages. The thick blue line shows the annual values after smoothing with a 21 point binomial filter. The dashed portion of the smoothed line indicates where it is influenced by the treatment of the end points. The thin blue lines show the 95% uncertainty on the smoothed curve.
Source: http://www.metoffice.gov.uk/hadobs/hadcrut3/diagnostics/comparison.html

of scientific work, as well as in engineering and construction. As a result, the normal uncertainty algorithms and uncertainty treatment mathematics can be applied to them. As mentioned above, the papers explaining methods and results (cited later in this paper) which deal with the methods and results show that this is partially done only, and for important data in a rather selective manner.

Figure 1 shows that the claimed overall uncertainty in the global temperature anomaly is about ±0.15 °C to ±0.1 °C over the whole time span. This combined uncertainty represents a claim of accuracy to the limit of single instrumental precision (±0.1 C) for land air temperature (LAT), as well as for sea surface (water) temperature (SST). But it is equal or exceeds those, which may be obtained by trained personnel only who use well-maintained instruments for any local measurement. And because temperature data are collected measured up to more than 100 years ago, and are obtained in various climate zones with annual mean values which vary from about + 30 °C to about -35 °C, thus covering a range of 65 °C. So, this claim means that one has been able to obtain an overall accuracy of ±0.23 % to ±0.15 %, respectively, in determination of global mean temperature anomaly. However, normal temperature measurements are commonly much less precise, or, using the term “uncertainty”, their uncertainties are much higher.

The definition of global mean temperature anomaly

In general, climate studies recommend determination of a global average temperature, calculated by determining its arithmetic mean of all useful local data. In order to do this the recorded single temperature data t_{ix} at time $_x$ from day $_i$ had been used to calculate a daily mean $\bar{t}_i$. For meteorological purposes the single temperature had been measured at (mostly) different times of the day, hence adding them, to obtain some arithmetic mean, using very different algorithms, selected by practical and economical based reasons. This basic arithmetic mean $\bar{t}_i$ was calculated historically by use of more than 100 different algorithms as Griffith [4] reports in 1997. The daily mean than is added up over a month and divided by the number of days in order to calculate an arithmetic monthly mean. All available monthly mean values are than used to construct a time series of local air temperature of this location.

It is well known that these raw data are rather difficult to compile over space and time. This is due to often no continuous data over time being available, as interruptions took place due to known or unknown reasons. Also site movements may have taken place, sometimes reported, sometimes not. And the location of the screens at different heights must be considered as well. To compile them over the local area often suffers additionally from sparse distribution of measurement stations, like for example most of Africa (especially the Sahara desert and also south of equator, but also in vast areas in South America and Asia, as well as in Australia) their different construction and often unknown maintenance status, but also overall quality. This gives an impression that it is very problematic to compare absolute time series of measurement stations with each other to find the potential departure from the mean, precise enough to allow any determination of long-term climate trends for a wider area.

So in order to find a common signal in this absolute temperature data which differ widely from each other due to influences of location, seasons, elevation and time, a monthly or annual anomaly is calculated from the single monthly or annual mean by deducting from the absolute a reference value. The reference value is the mean of all monthly (sometimes also annual) means within a fixed 30 years time span of same station. It is often named the **"station normal"** $\bar{T}$. This practice is chosen because it is well known that anomalies do not deviate as much from a common base as the absolute data do. The reference value or "**station normal"** named $\bar{T}$ is recommended by the WMO World Meteorological Organization as the temperature average for the years from 1961 to 1990. As WMO Arguez et al [5] defined: $\bar{T}$ is defined as $\bar{T} = \frac{1}{30}\sum_{i=1961}^{1990} \bar{t}_i$; where $\bar{T}$ is here the thirty year average annual mean values; $\bar{t}i$ is the yearly average station temperature.

Knowing this consequently NCDC-NOAA authors....[6]. wrote in paragraph 3[2] about the preference for anomalies*"(*using)*.... reference values computed on smaller* [more local] *scales over the same time period establishes a baseline from which anomalies are calculated. This effectively normalizes the data so they can be compared and combined to more accurately represent temperature patterns with respect to what is normal for different places within a region."* .

[2] Source: http://www.ncdc.noaa.gov/cmb-faq/anomalies.php paragraph 3

But calculations of anomalies create in general the same problems with error handling which arise by calculation of absolute monthly or annual average values of t_{ix}. But in this case twofold, because both datasets contain errors. The local temperature t_{ix} is never free from measurement errors, neither is the reference value $\overline{T}$. Therefore the reasoning of NCDC-NOAA authors further down in..[6] in paragraph 4 needs a critical examination. There one reads: *"Anomalies **more accurately** describe climate variability over larger areas than absolute temperatures do, and they give a frame of reference that allows more meaningful comparisons between locations **and more accurate calculations of temperature trends.**" (bold added)*

The first part of the sentence explains precisely the reason why anomalies are preferred, but the bold part of paragraph 4 is incorrect in general. Anomalies will be in most cases less accurate than the absolute temperatures. This is because both values contain random and systematic measurement errors, where the latter one varies in time, direction, magnitude. There is no reason to assume that these errors during the normal period are identical to, or even similar to the errors in any given month. This may be the case, but remains an exception. In addition, calculating the – often small – differences gives any error a much bigger weight, than it would have comparing it with the absolute data only.

Calculation of anomalies requires combining the uncertainties in the temperature measured, as well those in the calculated mean temperature, which consist of the average of all measured temperatures within the reference time. That obstacle is often not mentioned. This is because the anomalies themselves are calculated by subtracting from the absolute monthly or annual local average temperature t_i a reference average temperature of same station. That is: every local average temperature is $\overline{t_i} \pm e_i$, where "e_i" is the total error in the *ith* averaged temperature "$\overline{t_i}$". And the temperature normal is $\overline{T} \pm \varepsilon$, where $\varepsilon = \pm\sqrt{\frac{\sum_{n=1}^{n=360}\varepsilon_i^2}{360}}$ is the average error from the thirty years of measured monthly temperatures that are averaged into $\overline{T}$. Every anomaly temperature is $\Delta t_i = \overline{t_i} - \overline{T}$, and the error in the anomaly Δt_i can be determined by applying usually the formula for the error in the difference between two *i* measurements, the root-mean-square (r.m.s.) of the errors in $\overline{t_i}$ and $\overline{T}$ but only as long as the error could be considered as fairly normally distributed. That is the case not only for random errors, but often for systematic errors also. But if this condition should be denied or one can not be sure enough than the simple linear addition of said systematic errors is the right treatment for error propagation (see for example 1988; Miller p. 1353)[7] of such errors.

In the first case the error in Δt_i is given by $\pm\sigma_i = \pm\sqrt{e_i^2 + \varepsilon^2}$. By using this, every anomaly is properly reported as $\Delta t_i \pm \sigma_i$.

So, the central reason for using an anomaly is that the excursions away from the mean temperature emerge in a coherent way. Thus, temperature departures from very many local mean temperatures can be combined in order to highlight the trends in climate change over a wide area. Anomalies are calculated so as to compare or compile the net temperature trends of the huge variety of local climates. However, the price one pays for this convenience is that the anomalies will carry a larger uncertainty than the original temperature measurements, as shown above. The only exception is, due to the subtraction of data, if their errors have the same direction and magnitude than they will be reduced or even may cancel out. But in order to know that this may happen, one have to determine these errors fairly precise in advance.

In the following sections the general remarks are applied to real observations and further a method is described which had been used to define the magnitude of one of the numerous potential errors, here named as "algorithm error".

The scope of this study

This study aims to show that in addition to the underestimated random uncertainty from B 06, which have been shown by F 11, additional systematic uncertainties are acquainted with every measurement data that are named. One of it is defined in detail introduced by the various algorithms used to calculate a daily mean on land and monthly mean on sea.

The uncertainty definitions, groups and method used

Meteorological measurements include a wide variation of potential and real errors. One can group them into random errors and systematic errors or uncertainties.

1. Random uncertainty

According to German VDI guidelines [3] [8] about *"Uncertainties of measurement"* …. *"Random measurement error are thus typically defined as fluctuations in magnitude, which have to have a Gaussian distribution and a mean of zero"*. In other words random uncertainty is defined by the population characteristics of the uncertainty itself, and not by the effect it has on measurements. Errors that behave differently are defined as not random. And all this definitions are valid only when making several to many repeated measures of **same value** - around the "true" value μ, which in our case is named t_i .

In climate research the single max. t_{imax} or min. temperature t_{imin} or any other daily regime for determining the daily mean temperature measurement is always treated as a measure which may contain random uncertainties according to the definition above. for example in B06, one of the most cited papers, describing methods and results for global mean temperature calculations, the only measurement error mentioned is a "read error." This is the readout-error made by an observer when recording the level in a liquid-column thermometer. This error is claimed to be random.

However, no evidence is presented to support this claim. This claim may be true, or it may not be true. Since there is no evidence either way, no one knows whether read-error is random, or not. Therefore, the claim of random read-error is an assumption. But this assumption has a powerful impact on the magnitude of the final error. The assumption is used to justify the use of statistics that reduces the final value of uncertainty. But this assumption is entirely unjustified, and therefore cannot be used for an estimation of error. Error estimates must always be conservative, so as to produce the maximum likely uncertainty. One cannot make unjustifiable assumptions that permit one to reduce the final uncertainty.

---

[3] see VDI 2048 p. 36

In reality the measured temperature t_i is always different from measurement to measurement and from day to day. If we name this value τ_i then $\tau_i \pm \tau_j$. Where $\tau_i = t_i + e_i$ is the real measured value at day i and $\tau_j = t_j + e_j$ is the real measured value at day j, and t and e are the components of that value. F 11 examined in detail the properties of these additional uncertainties. He states*: The mean temperature,* $\bar{T}$*, will have an additional uncertainty,* ±*s, reflecting the fact that the* τ_i *magnitudes are inherently different*". Whilst Folland et al [9] as reported by B 06 treated this uncertainty as fully random, following the formula below, with a remaining random uncertainty of only ± 0.03 °C; they overlooked that it also contains a *"magnitude uncertainty* ± *s"* that comes from the spread of sub-states in a heterogeneous system component. F 11 states (p 408) "*B06 (Brohan et al 2006) assumed that temperature measurement error is random and declines as* $1/\sqrt{N}$ *and disregarded magnitude uncertainty,* ±*s, which should have followed the description of station normal error,* ε_N*. Magnitude uncertainty also made no appearance in any monthly or annual mean surface air temperature anomaly* [1]". [1] refers to the cited paper of B 06.

Applying the results found by F 11, the minimum added magnitude uncertainty due to different magnitudes measured day by day over the entire global time-series is s = ±0.28 °C (p. 416). For data observed over sea – here the sea surface temperature SST is used as a proxy with own properties different from air, instead of the required air temperature MAT (Marine Air temperature) – the final uncertainty must be greater, due to much to less data per day being available, so that only monthly mean temperatures can be determined often only by few or no daily measurements. The averaging is done not considering the position (although the position data exists) of the measurement device, except it must belong to the grid within it was measured. But this particular problem may be covered in some length in another paper.

2. Systematic uncertainties

Systematic uncertainties are different from random uncertainties but appear also with almost every measurement. According to the before cited VDI Guidelines ((VDI 2048 p. 36) "*….a systematic measurement uncertainty* e_s *(see DIN 1319-1*[1]*) is the term given to that portion of the uncertainty of a measured value x* from the true value* μ *which arises as a result of the imperfection of the sensors and the measuring method due to constant influences and causes. The systematic measurement uncertainty* e_s *would assume the same value. The systematic measurement uncertainty* e_s *is composed of the known and measurable systematic measurement uncertainty* $e_{s,b}$ *and the unknown and unmeasurable measurement uncertainty* $e_{s,u}$*. The measured value is described as follows*

$$x^* = \mu + er + es = \mu + e_r + e_{s,b} + e_{s,u}$$

In other words they are biases in measurement which lead to the situation where the mean of many separate measurements differs significantly from the actual value of the measured attribute. If they are known or can be precise enough and clearly estimated than they can be corrected, by applying the necessary correction to the measured value. A very simple example for a systematic error is a clock, which is always 5 minutes late, whereby one can easily add 5 minutes to the readout and get the precise time. If this is not possible they have to be properly described, estimated in direction and quantity and added as a bar of uncertainty to the measured value. But this clock example shows also quite simplythat errors of data used to construct time series behave very differently from single data. In order to correct them one should not only know the magnitude and direction of the single systematic error, but also the time, when it appears and how long I acts. Only then it can be canceled out by anomaly computation, because otherwise it distorts the time series and its trend 1:1.

Unfortunately the kind of systematic uncertainties in climatology are not as simple as the clock example above. In addition they appear in various multiple forms to the variable and can hardly be distinguished from the true value. Looking especially at time-series, as one does in climatology, they can also develop as a function of time, as is typically indicated by the UHI. (Urban heat island effect) for example.

This paper deals in detail only with one of them, but it will aim to name most of them. The systematic uncertainties in meteorological measurements can be grouped into various classes from which often all or some will appear when measuring local temperatures. They are related to:

1. Thermometer or sensor
   - 1.1. Design
   - 1.2. Readout
   - 1.3. Age
   - 1.4. Change of sensors. How, what?
2. Shelter/housing/screen of thermometer or sensor. Subdivided into:
   - 2.1. The construction of the housing itself
   - 2.2. Its condition, painting, structure etc.
   - 2.3. Its measuring height above ground
   - 2.4. Its location in the surrounding landscape acc. for example to *CRN Climate reference Network* Class 1 to 5 [4]
3. The station coverage over land and sea. Subdivided into:
   - 3.1. coverage over land
   - 3.2. coverage over sea
4. Distribution of measuring stations over land and sea.
   - 4.1. By location
   - 4.2. By height above normal
5. For sea surface measures of SST own error classes exist. They can be coarsely grouped into:
   - 5.1. Bucket take in of water
   - 5.2. ERI (Engine rear intake) of water
   - 5.3. Difference between SST and MAT (Marine Air Temperature)
   - 5.4. Measurement error between different temperatures and different sensors
6. The time of observation. Subdivided into:
   - 6.1. Time of observation due to used mean algorithm
     - 6.1.1. Max – Min. Method
     - 6.1.2. Mannheimer Method

6.1.3. Soviet or Wild's Method

6.2. Others

7. Length of continuous observation

7.1. With or without interruption

The quality of individual surface stations is perhaps best surveyed in the US by the excellent and independent evaluations carried out by Anthony Watts and his group of volunteers. It is publicly available i.e.. Watts [10] and covers in extent the entire USHCN surface station network. Due to this comprehensive study about 69% of the USHCN stations were reported to enjoy a site rating of poor, and a further 20% only fair[5] . Poor means that they may have a deviation of the real temperature according to the US Climate Reference Network Rating Guide (CRN) classification of >2 to > 5 °C. Fair means acc. same classification of > 1 and < 2 °C. These and other[6] more limited published surveys of station deficits have indicated far from ideal conditions governing surface station measurements in the US. But as F10 reports this is true in Europe also. *"A recent wide-area analysis of station series quality under the European Climate Assessment* [JCOMM 2006]*, did not cite any survey of individual sensor variance stationarity, and observed that, "it cannot yet be guaranteed that every temperature and precipitation series in the December 2001 version will be sufficiently homogeneous in terms of daily mean and variance for every application."*

Each of these groups mentioned above has its own range of systematic errors which have to be examined very carefully. The papers of B 06, Karl 1994 et al [11] , Jones et al [12] ; [13] name only the most cited ones and examined only a few of them.

According to Karl et al the uncertainties which need to be corrected can be summed up as follows:

1) urban heat island bias, 2) changes in observing times, 3) changes in instrumentation, 4) station relocations, and 5) inadequate spatial and temporal sampling.

---

[4] *US NOAA für CRN Climate reference Network* NOAAs National Climatic Data Center:Climate Reference Network (CRN). Section 2.2. of the Climate Reference Network CRN. Site Information Handbook, "thc most desirable local surrounding landscape is a relatively large and flat open area with low local vegetation in order that the sky view is unobstructed in all directions except at the lower angles of altitude above the horizon." Five classes of sites - ranging from most reliable to least - are defined.

[5] look for details here Watts, A., Is the U.S. Surface Temperature Record Reliable?, The Heartland Institute, Chicago, IL 2009

[6] All studies derived from Frank where they are listed: 10: Pielke Sr., R., Nielsen-Gammon, J., Davey, C., Angel, J., Bliss, O., Doesken, N., Cai, M., Fall, S., Niyogi, D., Gallo, K., Hale, R., Hubbard, K.G., Lin, X., Li, H. and Raman, S., Documentation of Uncertainties and Biases Associated with Surface Temperature Measurement Sites for Climate Change Assessment, *Bull. Amer. Met. Soc.*, 2007, 913- 928; doi: 10.1175/BAMS-88-6-913.; Davey, C.A. and Pielke Sr., R.A., Microclimate Exposures of Surface-Based Weather Stations, *Bull. Amer. Met. Soc.*, 2005, 86(4), 497-504; doi: 10.1175/BAMS-86-4-497. Runnalls, K.E. and Oke, T.R., A Technique to Detect Microclimatic Inhomogeneities in Historical Records of Screen-Level Air Temperature, *J. Climate*, 2006, 19(6), 959-978. Pielke Sr., R.A., Davey, C.A., Niyogi, D., Fall, S., Steinweg-Woods, J., Hubbard, K., Lin, X., Cai, M., Lim, Y.-K., Li, H., Nielsen-Gammon, J., Gallo, K., Hale, R., Mahmood, R., Foster, S., McNider, R.T. and Blanken, P., Unresolved issues with the assessment of multidecadal global land surface temperature trends, *J. Geophys. Res.*, 2007, 112 D24S08 1-26; doi: 10.1029/2006JD008229.

But as shown above, with this list only few of the potential uncertainties have been identified. In addition these authors reduce their correction activities to the last uncertainty: *inadequate spatial and temporal sampling*. They explain,

"Our analysis focuses on this last item because until the last few decades most of the globe was not sampled.

We will consider two types of errors: the errors that arise owing to an absence of any observations (type I errors of incomplete geographic coverage) and the errors that arise owing to imperfect sampling within grid ceIls (or averaging areas) with observations (type 2 within-grid cell errors).

....Clearly, they are only a portion of all the errors and biases affecting the calculation of hemispheric and global temperature trends that must be considered in any comprehensive error analysis."

That statement is very true, because on page 1162 it is stated:

"Unfortunately, it will never be possible to be certain about the magnitude of the errors that may have been introduced into the historical record owing to incomplete spatial sampling because we will never know the true evolution of the spatial patterns of temperature change."

B 06 the most recent study examined the following sources of uncertainties:

Station Error: *the uncertainty of individual station anomalies.*

Sampling Error: *the uncertainty in a grid-box mean caused by estimating the mean from a small number of point values.*

Bias Error: *the uncertainty in large-scale temperatures caused by systematic changes in measurement methods.*

But regardless of this statement, the only systematic uncertainties the authors look for are the Urban Heat Island Effect (UHI) and a potential change of sensors only. In respect to UHI they follow Jones [12][7] and Folland [9][8]. According to them, this effect is increasing gradually from 0 (1900) to 0,05 °C (1990). With this it is much too small to have an effect. To underline this they cite Parker et. al (Parker, 2004 , Peterson, 2004).

The uncertainty caused by the time of observation and the algorithms used.

From the very beginning meteorologists have wanted to measure true air temperature data, showing how weather has developed during the day, and how this will develop over months and years. Therefore the daily temperature as a major component of weather has to be measured. As a significant indicator for daily temperature the arithmetical daily mean seems to be a good index to show its variation over time. Undoubtedly the best mean for easy comparison would be a mean of diurnal time-series where

[7] See *Nature*. *347* , p169 ff.

[8] Cited from "Bias correction uncertainties are estimated following (Folland et al., 2001) who considered two biases in the land data: urbanisation effects (Jones et al., 1990) and thermometer exposure changes (Parker, 1994). Urbanisation effects The previous analysis of urbanisation effects in the HadCRUT dataset (Folland et al., 2001) recommended a 1σ uncertainty which increased from 0 in 1900 to 0.05° C in 1990 (linearly extrapolated after 1990) (Jone et al., 1990). Since then, research has been published suggesting both that the urbanisation effect is too small to detect (Parker, 2004, Peterson, 2004), and that the e?ect is as large as ≈ 0.3° C /century (Kalnay & Cai, 2003, Zhou et al., 2004)"

temperature data are measured in very short intervals. Today, enabled by remote electronic sensor techniques, one might use 5 minutes intervals; even 15 minutes intervals may be fine and sufficient. The measured values are used to calculate a daily mean. This is intended as an approximation to the temperature integral over the hours of the day. Historically, worldwide more than 100 algorithms for calculation of daily mean temperature have been used as Griffith [4] reported in 1997. Summing up the measured values and divide them by their number would give the best available mean, or "true" mean[9]. It can be calculated in general by the formula

$$\bar{t}_i = \sum_1^i \frac{t_{in}}{n} \tag{1}$$

where $\bar{t}_i$ is the daily mean temperature of measured variable t_{in}, which is the local temperature (note it is not the true temperature in free air, but the temperature within the shelter) at day *i* and time *n* measured at every time interval as short as is feasible. Since this demand would have been too time consuming in pre-electronic times, one had to reduce the number of measurements in order to make it practically viable. Therefore one decided to measure either only twice a day with maximum and minimum temperatures, or at fixed hours 3 or 4 times a day, or some other frequency.

The most used algorithm in continental Europe and some of its former colonies is named "Mannheimer Stunden" because this algorithm was introduced by one of the first meteorologic societies, "Societas Meteorologica Palatina", also named "Mannheimer Meteorologische Gesellschaft". It was founded in 1780 from Elector Count Karl Theodor from Mannheim, Germany. Its daily mean value is calculated according to formula (2).

$$\bar{t}_i = \frac{t_{i7} + t_{i14} + 2t_{i21}}{4} \tag{2}$$

where t_{in} is the local temperature measured at day at n = 7, 14 and 21.00 hrs.

And, mostly in Anglo-Saxon countries and their former colonies in Africa and Asia the Max & Min algorithm was, and is in use.

The daily mean is calculated for the daily mean by formula (3):

$$\bar{t}_i = \frac{t_{i\min} + t_{i\max}}{2} \tag{3}$$

Also widely used in Russia and latterly the Soviet Union as well as its former satellites, but not considered here, is a method which was developed by the meteorologist Wild in St. Peterborough in the 19^{th} century at four fixed hours. Four measurements are taken and the mean is calculated by formula (4):

---

[9] The "true" mean should not be mixed with "true" temperature. The latter is the temperature outside the screen, in which meteorologists are really interested in.

$$\bar{t}_i = \frac{t_{i0} + t_{i6} + t_{i12} + t_{i18}}{4} \quad (4)$$

where t_{in} is the local temperature measured at day i with n = 0, 6, 12 and 18.00 hrs.

It is known, that all these algorithms give a different daily mean if compared to the "true" mean, as pointed out by B 06; but also to each other, as can be calculated easily.

But which one is right, which one is wrong? The answer is, there is no possible correct answer. Since all of them do not have any physical meaning, this question cannot be answered by using first physical principles. Therefore it is a free choice following practical considerations. And as long as meteorologists concentrated their efforts only on daily weather or defining local climate zones, always using the same algorithm for daily calculations, it did not really matter either.

But as soon as the data needs to be combined in order to calculate a global mean, the magnitude of uncertainty introduced by different algorithms become important. As B 06 stated (p. 2 and 4):

(1) *"The station normal (monthly averages over the normal period 1961–90) are generated from station data for this period where possible....The values being gridded are anomalies, calculated by subtracting the station normal from the observed temperature, so errors in the station normals must also be considered."*

Then in the following statement, a number of errors are defined, but the potential algorithm errors are not. Instead, somewhat later (p.6), it is stated why:

(2) *"..There will be a difference between the true mean monthly temperature (i.e. from 1 minute averages) and the average calculated by each station from measurements made less often; but this difference will also be present in the station normal and will cancel in the anomaly. So this doesn't contribute to the measurement error."*

And later in same paragraph it was stated :

(3) *"If a station* ***changes the way mean monthly temperature is calculated*** *it will produce an inhomogeneity in the station temperature series, and uncertainties due to such changes will form part of the homogenization adjustment error. (bold marked by the author)"*

From the statements above No. 2 is true for one single case only, namely when the systematic error introduced for example by the algorithm itself remains not only constant but also is present over the whole time-series. And this is hardly ever true. Not only because changes of the algorithms used have taken place fairly often over time, if one considers only the breakdown of Soviet Union and the following changes in applied standards. But also if one compares any anomaly derived from one algorithm to any other by blending them with each other with the real wanted "true" mean. Therefore some efforts have to be undertaken to isolate and hence correct the error introduced by changing the algorithm for mean calculation. But this hasn't been done in the procedures described for correction of (see statement 3) *homogenization adjustment error.*

So the error models used by B 06 does not take into account this uncertainty but believes in most cases instead it will cancel out, or had been corrected by later homogenization. But this has not happened due to the fact that only monthly mean values have been used.

These now follows an attempt to correct this.

RESULTS

In order to do this one has first to try to quantify the uncertainty caused by the algorithm used. To get an idea about this one should look at real data. Regarding the impact and magnitude of algorithm errors there is unfortunately a sparse literature available, which may allow some more precise mean value to quantify. In Allisow et al [14]-textbook about climate science (German edition "Lehrbuch der Klimatologie") – they name three places namely, Leningrad (now St. Peterborough, Europe Baltic Sea) 59° 56`N, 30°16E, Barnaul (West Siberia) 53° 20´N, 83° 46´E and Tbilisi (Georgia) 41° 34´N, 44° 48´E, to explain the impact of this different mean calculation. In some detail they showed the difference between the ideal mean value calculated by hourly observed records with 3 (applied since 1871) and 4 time observations (since 1936) in the whole Soviet Union. They reported about mean differences obtained just by used algorithm against the ideal mean in the warm period from 0.4 °C° to 1.0 °C (3 daily measurements) and from 0.4 °C to 0.5 °C (4 daily measurements). In the cold period from 0.1°C to 0.3°C (3 as well as 4 daily measurements). But – as they write – as an exception in some months in spring and fall the 3x and 4 x algorithm may also deliver negative differences to the ideal mean.

And the authors had been fully aware about the impact of this change and described difficult measures to correct those errors. But as they explained that this is hard to correct, because just very few stations allow a 24 hour recording of temperature, and for some others one has to rely for correction on a more generalized formula developed by Wild. So in that case homogenization took place for few stations only. And this shows that one could not expect many of them.

Furthermore Germany for example used the Mannheim Method (t7 + t 14 + 2 x t 21)/4 until 1945 to 1950 in its whole area. Until this time this contains also part of todays Poland. When divided in 1945 East Germany (as well as Poland and all other east block countries) took over the soviet method (a little different from the one mentioned above, (t0 + t6 + t 12 + t 18)/4; but West Germany continued with the Mannheim Method. In 1990 this was unified to West Germanys algorithm only. In 2001 Germany as a whole switched to 5 minute intervals and since that time it becomes – in comparison – 0.1 °C warmer. The latter change of algorithm is under observation. Austria used for some time (t7 + t14 + t 21)/3 as well as for some time (t7+t19+tmy+tmin)/4. , also Asia (Russia, China etc). South America probably too, due to their strong cultural and economic bindings to Spanish Europe.

Though there are few observations are available one can use the data published by Aguilar et. al [15]. This paper is a very careful homogenization guidance named "Guidance on Metadata and Homogenization". It defines how homogenization of local weather data should be done.

Within this study the influence of algorithms on calculated means from data collected at Station Puchberg, Austria (p. 19) is shown. Puchberg is located at 47° 47′ N, 15° 54′ E, Height: 585 m above sea-level. It locates a very well maintained weather station which was therefore probably selected for this reason. Puchberg station is based in the middle northern latitude of Central Europe and observations used for this purpose are full nine years long and for each day 24 hourly observation. One can argue that in higher and lower latitudes the observable differences may be smaller or greater, but as reported by Allisow et al with bigger deviations observed by them the Puchberg data appear as a very cautious common mean.

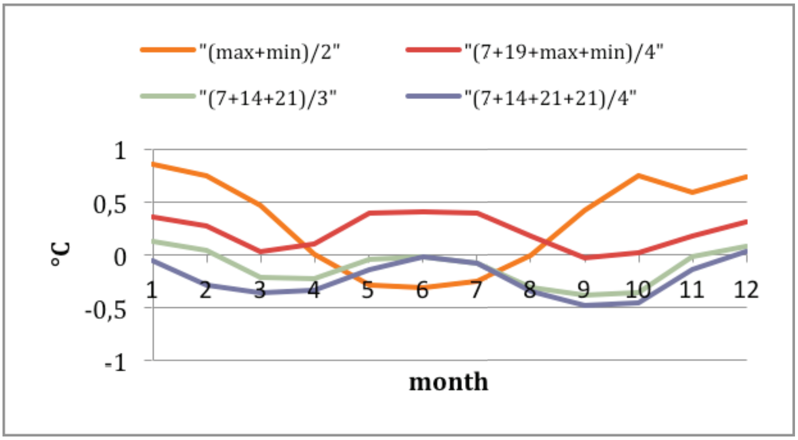

Figure 2 Evolution through the year of the difference between different ways of calculating daily mean temperature and 24-hourly observations average for the inner-alpine station Puchberg in Austria, 1987-1996. Data source: Central Institute for Meteorology and Geodynamics, Vienna, Austria.

As figure 2 shows, the monthly mean of same screen temperature is different for every month for each algorithms looked in. For simplicities sake, we will further compare only two of them, namely the "Max-Min" and "Mannheimer Method" with the "true" mean (the base line at zero degrees).

As figure 3 show if the used averaged algorithm results are compared with this "true" mean, the annual mean using the max-min method differs by + 0.3°C, the Mannheimer Method by -0.23 °C. Against each other the difference between Max-Min method and Mannheimer Method is 0.53 °C.

Due to its location, good condition and precise measurements the data shown appear to be usable for defining a lower range of algorithm uncertainties. They will act as a systematic bias, which has to be taken into account when estimating systematic errors in annual anomaly time-series.

As was previously shown in B 06, the station anomaly is calculated by subtracting the station normal from the mean monthly value of that station. For our purpose of estimating the algorithm error, we will yet not take into account the systematic errors of the station normal nor the error of the monthly value as shown by Frank 2010, 2011.

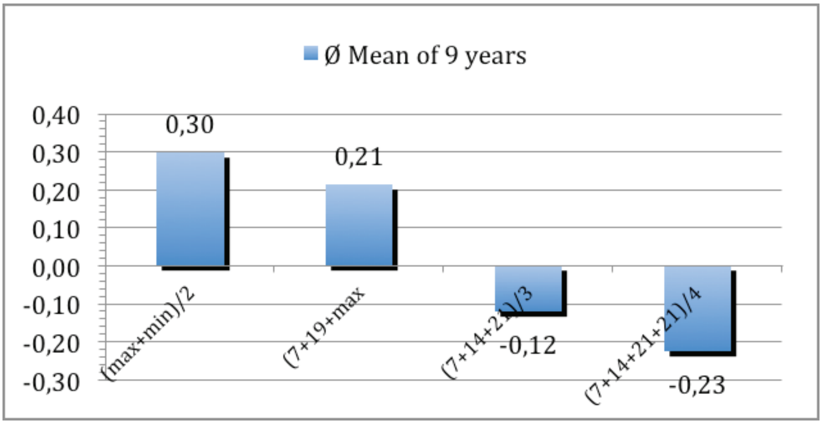

Figure 3 From Figure 2 Same data as above but averaged over 9 years for different ways of calculating daily mean temperature and 24-hourly observations average for the inner- alpine station Puchberg in Austria, 1987-1996. Calcuation done by the author. Data source: Central Institute for Meteorology and Geodynamics, Vienna, Austria.

Though the resulting anomaly "ceteris paribus (c.p.)" based on Brohans calculation may have cancelled out the intrinsic algorithm error (but only as long as this error is constant over time, for which no evidence is given nor is it even likely), looking at fig. 2 and 3 again, it can be seen that this is not very likely, since every monthly mean value is different from the next within the same algorithm. So this assumption can be regarded as to be wrong in all cases, where changes took place. That is true in for all places of the world. It also applies to the US and other English speaking countries, which may have maintained the unchanged max/min algorithm over time. But due to the fact that hardly any station in the world remained unchanged the assumption that the "station normal" may contain the same error as the running monthly mean is rather unrealistic. In addition, if the station normal and the monthly actual mean is referred to the "true" mean, what implicitly is done by blending the data calculated by various algorithms, all anomalies carry just by this a systematic error of + 0.3 °C or -0.23 °C respectively, depending on how they are calculated, f.e. by the max/min method or Mannheimer method respectively . So we have to add at least this error as an + 0.3 °C and -0.23 °C uncertainty to the global temperature anomaly.

Although the described algorithms may cover about 70 % of world land based temperature data (Europe started in the 17th Century with repeated meteorological measurements, the United States in the 19th Century) we are not certain about the real distribution applied to every measurement worldwide. Therefore the choice must be the smallest error of ± 0.23 °C which in all cases considered limits the lower value, regardless which algorithm had been used.

Now we may add this value to the previous findings of Frank to the total minimum error. Systematic errors arising from independent sources can be combined within some limits as the root-mean-square-errors (rmse) combined in quadrature. As this is explained in some detail in the textbook "An Introduction to Error Analysis: The Study of Uncertainties if Physical Measurements" by John Robert Taylor"[16]. Frank showed in F 10 and in F 11 that the minimum instrumental uncertainty cannot be smaller under best conditions than ± 0,46 °C. In F 11 he added another ±0,17 °C error from the annual anomaly trend. Now we have to ad further + 0,3 °C and – 0,23°C due to algorithm errors which certainly show up. This than is calculated to:

$$+1\hat{\sigma} = \sqrt{0.283^2 + 0.359^2 + 0.17^2 + 0.3^2} = 0.573 \tag{5}$$

and

$$-1\hat{\sigma} = \sqrt{0.283^2 + 0.359^2 + 0.17^2 + 0.23^2} = -0.542 \tag{6}$$

But as was stated before, in order to simplify the equation we take only into account the value of equation (6) and this will now show

$$\pm 1\hat{\sigma} = \sqrt{0.283^2 + 0.359^2 + 0.17^2 + 0.23^2} = \pm 0.542 \tag{7}$$

A 95 % confidence interval of these four defined errors alone equals roughly 2 x ± 0.542 = ± 1,084 °C.

Because with these errors the Gaussian distribution cannot not be guarantied one has to determine additionally the upper boundary of the total minimum error by computing the linear addition of said errors which can not be exceeded (see Taylor, John Error Analysis, p. 60)

$$\pm\sigma = 0.283 + 0.359 + 0.17 + 0.23 = \pm 1.042 \tag{8}$$

or ± 2.084 °C respectively

Using this limits the combined error may be expressed to lie within the limits for 1σ as

$$\pm 0.542\ (7) < \pm\sigma < \pm 1.042\ (8)\ °C$$

or with 95 % confidence range of 2 σ

$$\pm 1.084\ (7) < \pm 2\sigma < \pm 2.084\ (8)\ °C$$

But when we consider the multiple error potential from the beginning (p 111), this is by far not all. In addition it may also be noted that the common error taken all algorithms into account is not equally distributed around the mean value of time-series, because it has different magnitudes in plus or minus directions.

CONCLUSION

It was shown that the limited uncertainty assessment of B 06 and his predecessors did not describe the real situation in respect to uncertainties of historical temperature measurements. In order to demonstrate this a single but wide spread error, named "algorithm error" was used. It adds at least another ± 0.23 °C (1 σ !) systematic error to the error developed by F 11 but it was also shown that this error is very likely much greater in reality.

Additionally it was shown that this is only one of a variety of other systematic errors which have been named in the previous pages, but could never been examined carefully, and thus have never been corrected. Due to this, any attempt to develop a time-series of a global temperature anomaly must fail, if not accompanied by a much wider band of uncertainty. Furthermore the anomaly itself is hidden within this broader band of uncertainty, meaning that also the resulting envelop curve is not symmetrical to the hidden anomaly time series. Thus makes it impossible to attribute any potential cause for the variation. Any attempt to attribute a probable cause to this global temperature time series therefore remains pure speculation. The only but very important exception may be the influence of a strong driving force which oscillates around a given mean. Its oscillating signal can be much more easily discriminated from the uncertainty band described before, due to its repetitive nature. This is because all uncertainty sources examined above appear either as jumps up to a new stable error value over time, (for example change of algorithm used) or may induce a creepingly in- or decreasing uncertainty, (like alteration of properties over time for example by aging of painting) but do not show oscillating properties. For purpose of climate research by looking at the results above one can ask which scientific value an anomaly mean calculation for the whole globe as done for example in B 06 may have once one consider the vast variations of earth temperatures depending from its location (see p.2; from +30 ° to -35 °C) and from season to season. That does not mean that no warming could be observed generally in some periods of the last century, because other proxies world wide indicate that there was a moderate warming, but it means that this time series construction does not allow to deliver the data for which it is intended for.

ACKNOWLEDGEMENT

This paper was initiated and written in full by the author alone. Regarding the proper treatment of systematic uncertainties and its often unknown distribution as well as magnitude and direction an intensive and very helpful discussion took place with Pat Frank PhD. He helped a lot to understand how systematic and random errors have to be estimated and calculated within the rules of error propagation. For this and his patience I want to express my thanks. Equal thanks I owe to Dr. Sonja Böhmer-Christiansen helping in preparing the manuscript. Furthermore the treatment and behavior of anomalies was examined and discussed with two gentlemen who acted as reviewers, but wanted to stay anonymous. That has to be respected. Their questions and hints guided to a more detailed description how anomalies behave if they are used for time series computation, for this I want to express my thanks also.

REFERENCES

1. P. Brohan, J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. . Jones "Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850," *Journal of Geophysical Research*, vol. 111, pp. D12106+},

2. P. Frank, "Uncertainty in the Global Average Surface Air Temperature Index: A Representative Lower Limit," *Energy & Environment*, vol. 21 , no. 8, pp. 969-989, 2010.
3. P. Frank, "Imposed and Neglected Uncertainty in the Global Average Surface Air Temperature Index," *Energy & Environment* · , vol. Vol. 22, N, pp. 407-424, 2011.
4. J. F. Griffiths, "Some problems of regionality in application of climate change. ," *J. Int. Soc. Biomet.*, 1997.
5. A. Arguez and R. S. Vose, "The Definition of the Standard WMO Climate Normal: The Key to Deriving Alternative Climate Normals," *American Meteorological Society*, pp. 699-704, Jul. 2011.
6. E. R. Long, "Contiguous U. S. Temperature Trends Using NCDC RAW and Adjusted Data for One-Per-State Rural and Urban Station Sets," *Sppi Original Paper*, 2010.
7. J. C. M. and J. H. Miller, "Basic Statistical Methods for Analytical Chemistry Part I; Statistics of Repeated Measurements," *Analyst*, vol. 113, pp. 1351-1356, 1988.
8. VDI, "Uncertainties of measurement during acceptance tests on energy-conversion and power plants Fundamentals," 2000.
9. C. K. R. Folland N. A. Brown,S. J. Smith3, T. M. P. Shen, S. S. Parker1, D. E. and I. J. Macadam P. D. Jones, R. N. Nicholls, N. and Sexton, D. M. H., "Global temperature change and its uncertainties since 1861 ," *Hadcrut 2, HadCRUT2 error analysis* , 2001.
10. A. Watts, *Is the U.S. Surface Temperature Record Reliable?*, vol. 1. The Heartland Institute 19 South LaSalle Street #903 Chicago, Illinois 60603 www.heartland.org, 2009.
11. J. R. Karl, T.R., Knight, R.W. & Christy, "Global and Hemispheric Temperature Trends: Uncertainties Related to Inadequate Spatial Sampling," *Journal of Climate*, vol. 7, p. 1144 bis 1163, 1994.
12. P. D. G. Jones P. Ya Coughlan, M; , Plummer, M. C; Wang W-C, Karl, T.R, "Assessment of urbanization effects in time series of surface air temperature over land," *Nature.*, vol. 347, p. 169 ff, 1990.
13. P. D. N. Jones M; Parker, D. E. S.Martin, Rigor, I. G., "Surface Air Temperature and its Changes Over the Past 150 Years," *Reviews of Geophysics*, vol. 37, pp. 173-199, 1999.
14. E. S. Allisow, B.P.; Drosdow, O.A; Rubinstein, *Lehrbuch der Klimatologie*, Geman Edit. Berlin: Wissenschaften, VEB Deutscher Verlag der, 1956, pp. 403-410.
15. E. A. Aguilar I. Brunet, M. Peterson, Thomas C.Wieringa,V., "Guidance on Metadata and Homogenization ," 2003.
16. John Robert Taylor, *An Introduction to Error Analysis: The Study of Uncertainties if Physical Measurements.*, Second edi. Mill Valley, California: University Science Books, 1997, p. 327.