Accuracy of temperature measurements and their "anomalies”

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 comprehensive and precise enough. This was already shown for some type of random and systemic errors by Frank [2] [3] hereinafter named F 10 and F 11. The same is valid for the simplified claimed SST random error cancellation by CRU[1] even for early times of 1850 when one has only few datasets available. This claim is based on the high assumption only namely that every random error of any former measurement random error have a mean of zero and are evenly distributed.

For the purpose of clearness we will 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 in literature 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. This report shows that based on the type of data available, their quantity, quality and methodology, as is the case with temperature, it is not possible to determine any global time series with the accuracy that is claimed. Thus the result is that all conclusions and correlations based on this data, which may come from causes considered to be important, lie within the range of uncertainties and thus can only be regarded as very rough estimates. Therefore every attempt to attribute any possible forcing to that variation remains scientific speculation only.

Preface

This paper is written as a report based on various papers written before by author as well as many others about the efforts done in order to measure meteorological temperature data. By this is intended to be a literature study rather than to show own research results. Nevertheless all information given in this combination and conclusions drawn are new and had not been reported by best knowledge of the author before.

Introduction

Averaging of suitable measurement data and calculation of its anomalies in order to detect otherwise hidden signals are standard tools in scientific procedures by using statistical methods. This methods are well proven as is the error propagation theory once introduced by C.F. Gauss. For global climate observations the main indicator is the local daily air temperature, measured at fixed times or by measurement of the maximum and minimum temperature at fixed locations and other parameter. The data collected by this are used as the basis to calculate a meaningful first local and

[1] Home http://www.metoffice.gov.uk/hadobs/hadsst3/uncertainty.html > HadSST3 see **Random observational errors: Cite „**Therefore, the contribution of random independent errors to the uncertainty on the global annual average SST is around 100 times smaller than the contribution of random error to the uncertainty on a single observation even in the most sparsely observed years (e.g. there were 23000 observations contributing to the average for the year 1850 and the square root of 23000 is greater than 100)“.

Draft

later global average of this temperature. One needs wants to determine

a. an absolute mean value of this data in order to allow estimations of a global mean air temperature used f.e. for estimating of energy transfer processes and
b. an anomaly of this values, i.e. the difference between a monthly or annual mean of local temperature with a mean value of same station collected over a period of 360 month or 30 years respectively. The reference period chosen is mostly set by using the WMO standard interval from 1961 to 1990.

For both calculated means the development over time is from great interest, where f.e. the x- axis show the years from 1850 to 2011 and the y axis show either the absolute mean in °C or the difference (anomaly) between actual annual or monthly mean value to a selected suitable reference value.

By using this procedures, it was shown by several climate institutes (GISS, NCDC, CRU) that the calculated mean of the entire global temperature anomaly had raised by approx. 0.7°± 0.1°C over the last century (see figure 1). Because earth surface is covered to 71 % by water, where air temperature measurements are much more difficult, climate scientist prefer to measure and use the sea surface temperature SST instead. This data vary much less over area and only little over daytime, due to the much greater heat inertia of water. Since the anomalies of SST seem to correlate quite well with LAT (land air temperature) anomalies, blending of the two seems to be easy and possible.

In order to determine smallest fluctuations of mean temperatures within few tenth of a degree C, the data itself and its statistical uncertainty treatment should allow final uncertainties with ± 0.1 °C . But the problem with which the experts had been faced with lies in the often-bad quality of often sparse historical meteorological data. Therefore, this task seems to be rather difficult if not impossible to manage. Because the general fact is, uncertainties in meteorological data do not differ in kind nor class from those in other areas of scientific work as well as in engineering and construction, therefore the normal uncertainty algorithms and uncertainty treatment mathematics have to be applied to them. As mentioned above, the papers which deal with the said problem (cited later in this paper) show that this was only partially done, and, for important data, in a rather selective manner only. In this paper an attempt is made to give an overview about the main error classes, which has to be considered and rough estimation of their potential magnitude are given. Finally a further attempt is made to quantify in brief the total uncertainty which accompanies every mean value.

Draft no 1

Problem description, examples

Local air temperatures had been observed for meteorological purpose since more than 300 years in a couple places of the world (f.e. at Berlin; Germany since 1701). Daily air temperatures on land are recorded mostly by measuring them at fixed times several times a day at and locations. In other regions only the Max and Min temperature had been determined without notation of time.

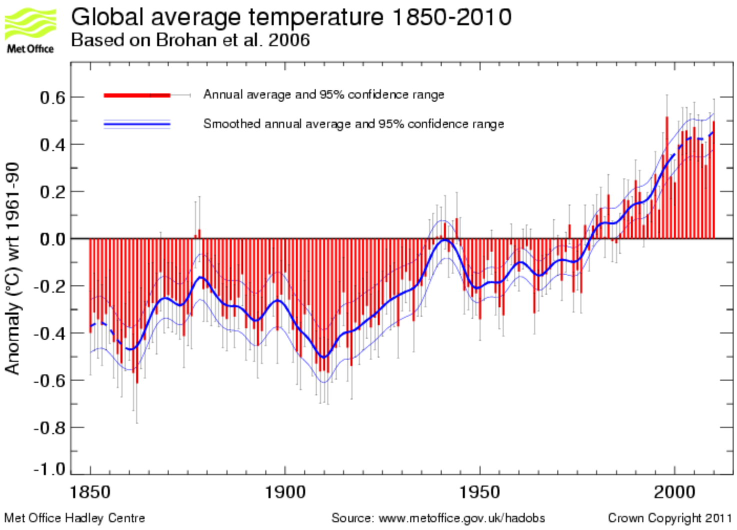


Figure 1 Global average temperature anomaly from 1850 to 2010 as published on MetOffice Webpage. The red bars show the global annual average near-surface temperature anomalies from 1850 to 2009. The uncertainty bars show the 95% uncertainty range on the annual averages. The thick blue line shows the 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

In order to prevent disturbing influences by direct solar radiation, wind flow and other perturbances the used thermometer is shielded against its environment and houses in a weather shed or screen. It is known from very beginning that this shielding has its own properties, which influence the measured temperature, thus do not allow directly measuring the true air temperature in which meteorologists are really interested. But the need for an acceptable compromise leads to these constructions matched to the area were they are sited by accepting the errors, which are created by this compromise. In addition to this known but not corrected errors introduced using 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 aim 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 than shorter data time series and as more continuously these data are measured as 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.

If such a station has been in existence for a longer period, f.e. 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 meaningful blended with all others into a global mean temperature time series as shown in fig.1. As shown there, every anomaly includes also its own uncertainty, as indicated by the gray confidence range in Figure 1[2] retrieved April 21, 2012, from http://www.cru.uea.ac.uk/cru/data/temperature/#sciref. The method used to calculate this anomalies is described by the British Met Office, and other organizations.

In the data available f.e. from the Met-Office web page (footnote 1) a number of those errors or uncertainties are defined.

Draft no 1

[2] 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,

The result of this treatment as shown in Figure 1 is 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 and precisely calibrated instruments for any local measurement. And because temperature data are collected which had been 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, this claim mean that one had 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.

Furthermore the raw data of local temperatures are subject to a variety of influences and thus entail inherent errors or deviations[3]. Only a small part of them, namely the real random errors, for which a sufficiently large data repository is available, allows to be determined and minimized or cancelled out by the application of statistical methods of the well known error propagation theory. These involve primarily and on part only reading and instrument errors. The largest portion, systematic and gross errors, must either be carefully corrected or accompanied by corresponding error ranges. The correction attempts by the scientists involved in compensating for systematic errors, for example due to the urban heat island (UHI) effect arising from altered land use, are automated and inflexible according to certain and few criteria because of the huge and often incomplete data amounts. "Inflexible" means they follow only a few and rigid requirements that do not allow real corrections because of a lack of knowledge regarding the boundary conditions. In addition the few corrective algorithms are unavoidably schematically designed. Correction often takes place at a desk, without any on-site viewing. The necessary metadata, which describe the boundary siting conditions of the raw data, are rarely available. Thus the corrections that do take place often lead to the wrong results. Other essential criteria are not recorded or are simply defined away.

As literature shows, this occurs among the differences in the various Stevenson screen designs [4]and with regards to the actual search of the real outside temperature. This problem is also well known in meteorological science.[5], [6]

Since meteorological temperature measurements were not standardized in the past and even today still aren't, different constructions [7] and methods were used globally for their determination.

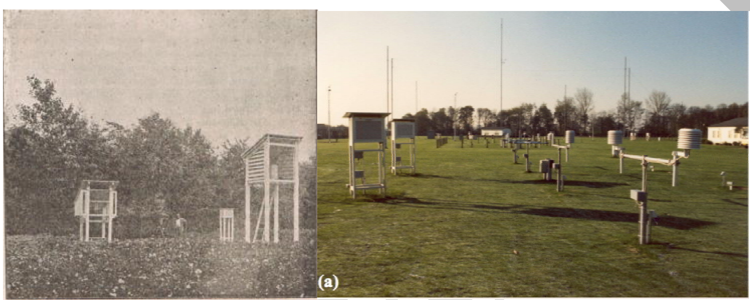


Figure 2 the left image (source [8]) show an variety of screen designs of the last century- From left to right: The French Screen, for comparison a man, than the original Stevenson Screen and the tall Wild´sche Hütte used in Russia. The right image (Source [7]) from left to right show 2 newer Stevenson Screens, somewhat bigger than the original one, and a verity of round modern housings for electronically sensors. One should realize that not only the size is rather different, but also the sensor height which varies for the Stevenson screen from roughly 1 m to 2 m. up to 3.2 m for the Russian screen.

---

3 To describe the kind of such uncertainties refer Brohan et al 2006 on the philosophical definition of that term by the former Secretary of Defense Donald Rumsfeld and makes them his own: *"A definitive assessment of uncertainties is impossible because it is always possible that some unknown error has contaminated the data, and no quantitative allowance can be made for such unknowns. There are, however, several known limitations in the data, and estimates of the likely effects of these limitations can be made (Defense secretary Rumsfeld, press conference June 6, 2002, London}* .Rumsfeld definfes this so: *"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don't know. But there are also unknown unknowns. These are things we do not know we don't know."*

Draft

This causes differing results depending on measurement conditions only but not caused by the data itself. Due to this, the historical values contain different errors originated by this. In the present study we investigate

1) whether the specified error sizes and bandwidths correspond to reality and
2) whether and how they impact the total uncertainty of global temperature anomaly on the basis of the official IPCC assessment.

By doing this we follow a bottom-up approach to address both points above. From the local measuring device to the final average. It details in short, the problems related to measuring air temperature on land LAT and over sea surface MAT NMAT and SST. It then discusses with sufficient detail, most of the possible errors which might come up during readout and further processing. It describes further the methods and algorithms for processing used by the science community in dealing with these topics, mentioning the claimed uncertainties.

Sources and types of errors

Meteorological measurements include a wide variation of potential and real errors. It is worthwhile to recall their definitions to the reader. One can group them into random errors and systematic errors or uncertainties.

Random errors

According to German VDI guidelines [4] [9] 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".* (bold added) 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 . The latter condition is hardly ever true in meteorological measurements.

Systematic errors

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*

$$x^*=\mu+er+es=\mu+er+es,b+es,u$$

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 simple that errors of data used to construct time series behave very different from single data. In order to correct them one should not only know the magnitude and direction of the single

Draft no 1

[4] see VDI 2048 p. 36

systematic error, but also the time, when it appears and how long it acts. Only than 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 that simple as with the clock example above. In addition they add 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.

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 or depending from:

1. Thermometer or sensor
    1.1. Design
    1.2. Readout
    1.3. Age
    1.4. Change of sensors. How, what?
    1.5. Shelter/housing/screen of thermometer or sensor. Subdivided into:
    1.6. The construction of the housing itself
    1.7. It's condition, painting, structure etc.
    1.8. It's measuring height above ground
    1.9. It's location in the surrounding landscape acc. f.e. to *CRN Climate reference Network* Class 1 to 5 [5]
2. The station coverage over land and sea. Subdivided into:
    2.1. coverage over land
    2.2. coverage over sea
3. Distribution of measuring stations over land and sea.
    3.1. By location
    3.2. By height above normal
4. For sea surface measures of SST own classes exists. They can be coarsely grouped into:
    4.1. Bucket take in of water (see fig. 3)
    4.2. ERI (Engine rear intake) of water
    4.3. Difference between SST and MAT (Marine Air Temperature)
    4.4. Measurement error between different temperatures and different sensors
5. The time of observation. Subdivided into:
    5.1. Time of observation due to used mean algorithm
        5.1.1. Max - Min. Method
        5.1.2. Mannheimer Method
        5.1.3. Soviet or Wild´s Method
        5.1.4. Others
6. Length of continuous observation
    6.1. With or without interruption

Draft no 1

[5] *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 c1asses of sites - ranging from most reliable to least - are defined.

figure 3 Various kinds of buckets for SST measurement s which had been used until the late 20th century to take samples of seawater from aboard a ship. A sample of the measurement sensor (LIG Thermometer; source [10]) is shown also

Identification of errors and potential treatment

The quality of individual surface stations is perhaps best surveyed in the US by the excellent and independent evaluations carried out in 2009 by Anthony Watts [11] and his group of volunteers. It is publicly available i.e.. Watts 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[6]. 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 °C. These and other[7] 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. Cite: *"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 the groups of error mentioned above has its own range of random as well as systematic errors which have to be examined very carefully. But only few of them could be considered. The papers of B 06, Karl 1994 et al [12] , Jones et al [13] ; [14] 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:
urban heat island bias, 2) changes in observing times, 3) changes in instrumentation, 4) station relocations, and 5) inadequate spatial and temporal sampling.

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,

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

[7] 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.

Draft no 1

"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 cells (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."

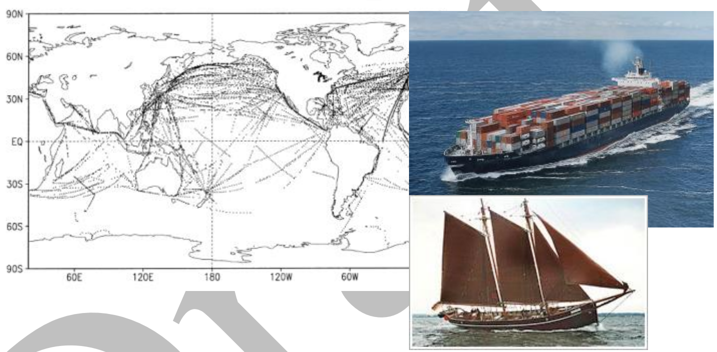
B 06 the another important study examined the following sources of uncertainties:

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 [13] [8] and Folland [15][9]. According to them, this effect is increasing gradually from 0 (1900) to 0,05 °C (1990). With this it would be much too small to have an effect. To underline this they cite Parker et. al (Parker, 2004 , Peterson, 2004).



Marine Air Temperature MAT and Sea Surface Temperature SST

Earth surface is covered to 71 % by oceans, about 80 % of the southern hemisphere, and about 60 % of the northern hemisphere. In addition the oceans with its currents and winds are the main source for our weather, short term as well as long term. As more important would be a continuous monitoring of air temperature on sea which should be geographically evenly distributed . But this is not possible to obtain. Plenty of instruments should have been installed, maintained and monitored, but doing this on ships (besides the few islands with its own microclimate) means that one have to accept uncomfortable compromises.

figure 4: the laft map with small dots show the number and distribution of SST data collected in the first week of January 2000 [16] On should take in mind, that never ever before the traffic at all the oceams had been more intense. The right upper image show container vessel with a bridge height of about 25 m and a depth of 10 to 15 m. The lower image at right show an sailing german ship of around 1900 Atalanta, with a low deck height of about 1.5 m.

[8] See there in: *Nature.* , *347* , p169 ff.

[9] 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) (Jones 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 effect is as large as ≈ 0.3◦ C /century (Kalnay & Cai, 2003, Zhou et al., 2004)"

Ships are built to travel, besides from exceptions like fire ships. So measurements steadily changes its position as well as height, when ships are travelling loaded or unloaded. In addition, the speed relative to calm air is changing and time of observation. Therefore, most climatologists prefer SST rather than MAT or NMAT although they are fully aware about the very different properties of sea water and. But the assume the anomalies of both temperatures behave rather equal looking at time series.

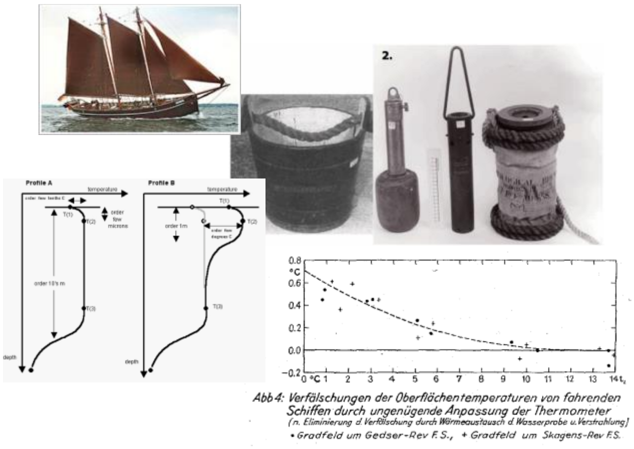


figure 5: From upper left to lower right. One of the last commercial sailing ship built around 1900 in Hamburg, Germany to drive into the northern Atlantic, with low deck height. Below: Two measurement done in calm water which show the variation of SST depending on depth. Upper right: 4 different buckets thrown out by a sailor and retreated after having touched the sea surface [10] It should collect water at around 1 m depth. The real depth remain unknown. Lower left: variation of water temperature measured in calm water depending on depth [17]. Lower right. Duration of time response of a thermometer put into the water of a bucket in the Skagerrak area in autumn [18]. After 10 minutes equilibrium has reached, this temperature should be recorded.

Jones assumes therefore that SST might behave like MAT in 10 m height. (Jones et al 1990 on p 174 and 175). But no evidence is given which supports this assumption. Instead, there are some papers showing just the opposite. F.e. Christy et al [19], showed by using data from fixed buoys in tropical areas from 1979 on, that their SST show a slight negative trend from -0.06 °C per decade, whilst simultaneously measured MAT (and NMAT) had one of +0.13 °C per decade. Also Kent et.al [20] did some research regarding this topic and reported that MAT in higher northern latitude in the north Atlantic typically has mean values 1.5 °C (p 11) above SST.

Regarding the observed or assumed uncertainties Met office scientist John Kennedy[10] wrote:
"A single SST measurement from a ship has a typical uncertainty of around 1-1.5K. (Kent and Challenor (2006), 1.2±0.4K or 1.3±0.3K; Kent et al. (1999), 1.5±0.1K; Kent and Berry (2005), 1.3±0.1K and 1.2±0.1K; Reynolds et al. (2002), 1.3K; Kennedy et al. (2011a), 1.0K; Kent and Berry (2008) 1.1K. These analyses are based on the more modern portion of the record. No studies have been done to see if there are systematic changes in the size of these errors through time. It should also be noted that not all measurements are of identical quality. Some ships and buoys take much higher quality measurements than others...and further down: ….
For ships, Kent and Berry found that the random error component was around 0.7K and the systematic observational error component was around 0.8K. Kennedy et al. (2011a) found that the random error component was around 0.74K and the systematic observational error component was around 0.71K. Adding the errors in quadrature gives a combined observational uncertainty of slightly more than 1K, consistent with earlier estimates. For drifting buoys Kennedy et al. (2011a) estimated the random error component to be around 0.26K and the systematic observational error component to be around 0.29K. The equivalent values from Kent and Berry were 0.6

[10] Source Met Office Hadley Centre observations datasets „**Uncertainty in historical SST data sets**" http://www.metoffice.gov.uk/hadobs/hadsst3/uncertainty.html

Draft no 1

and 0.3K respectively. The systematic component of the error was assumed to be different for each ship, so this model does not on its own capture the effects of ***pervasive systematic errors****.*
They assumed further that the systematic bias of more modern ERI (Engine Rear Intake, i.e. cooling water used for cooling the engines)) data against bucket data (buckets collect water manually form sailors aboard their vessels) could be -0.3 °C (referred to its monthly mean) in general. This would be the case especially than, when air temperature is remarkably cooler than water. Moreover, they noted that they prefer bucket data rather than ERI data because they seem to be more reliable. The problem with this is, that is fairly unknown what method bucket and or ERI had been used over the last 150 years. Whilst Parker, Jones and Reynolds assume much more bucket data only until 1941, and from than on bucket use ceased out pretty fast, but mostly ERI data, other observations show a solid mixture of both until the late 1980 increasing even till year 2005. But f.e. [21] Folland 2005 based his whole correction bias program on that assumption. True over time bucket method had been replaced with large variation by ERI and upcoming buoys method, but as figure 3 – especially the lower picture- show a fairly large portion of bucket data until recent years.

Results and implanted methods of correction

Averaging of absolute data, advantages and limits

Behaviour of anomalies and errors in time series and its limits.

The reason why time series of anomalies are preferred is, that in many sciences anomalies of independent quantities are calculated in order to find out common trends, correlations or in general "signals" which might be easier found and quantified than comparing absolute time series of single data. This is especially true for climatology where one deals with very different original temperature data. Due to this consequently NCDC-NOAA authors.. [22] wrote in an explanation why they prefer to use anomalies (in paragraph 3)[11] *.... reference values computed on smaller [more local] scales over the same time period establishes a baseline from which anomalies are calculated. This effectively normalises 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."* And in paragraph 4 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 in paragraph 3 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. Because both values contain random and systematic measurement errors, where the latter one varies in time, direction and magnitude. There is no reason to assume that these errors during the reference (normal) period are identical to, or even similar to the errors in any given other daily data, or its daily, monthly or annual mean. This may be the case but remains an exception. In addition, calculating the often small differences give any error a much bigger weight, than it would have comparing it with the absolute data only.
Therefore, it is worth to examine whether the calculation of differences (anomalies) between two measurement data cancel out their potential errors? Especially their systematic errors? This question is often answered with a

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

Draft no1

straightforward "yes", regardless who is asked in the scientific or engineering world. But is this answer true as many people believe? It will help if we look to the following simple formula:

Given that measurement or reference value t_2 is deducted from measurement value t_1 and t_1 as well as t_2 consists of its "real "value τ_1 or τ_2 plus an error of ε_1 or ε_2, than one can state

$$t_1 = \tau_1 \pm e_1 \text{ and } t_2 = \tau_2 \pm e_2. \quad (1)$$

The difference between the two values can be calculated as

$$\Delta t = t_1 - t_2 \text{ and therefore } \Delta t = (\tau_1 \pm e_1) - (\tau_2 \pm e_2)$$

$$\Delta t = \tau_1 - \tau_2 + (e_1 - e_2) \quad (2)$$

$$\text{with } (\pm e_1) - (\pm e_2) = \pm\sigma \quad (2a)$$

In case $e_1 = e_2$, which may be often true for measurement data of same origin and measured by same instrument $\pm\sigma = 0$ i.e. than ε_1 cancel out ε_2. This means the anomaly

$$\Delta t = \tau_1 - \tau_2. \quad (3)$$

presents the difference between the quantities without its errors

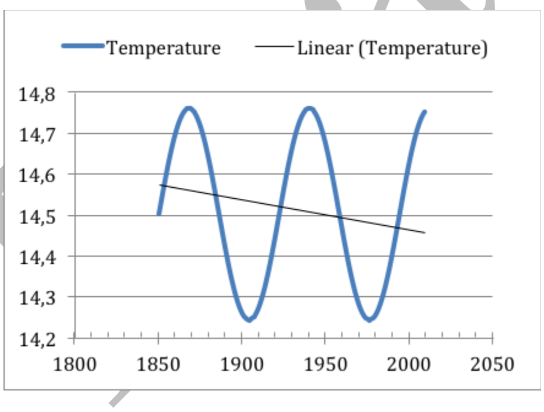
But in order to obtain this result one must be sure that the error ε_x , (precisely the systematic error ε_x) will remain constant in magnitude and direction. This is very often true in case one measure the same quantity with same instrument in close time intervals. But equation (3) shows us also the limits of this approach. What will happen if $\varepsilon_1 \neq \varepsilon_2$?

And even more interesting is, what will happen if one needs to draw time series of anomalies if $e_1 \neq e_2$?

In order to show this, we will examine the behaviour of anomalies in time series for temperature but shall having first a closer look to one of the most cited paper in recent climatology. 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 normal must also be considered."

Then in the following statement, a number of errors are defined and somewhat later for a given type of error (p.6), it is stated:



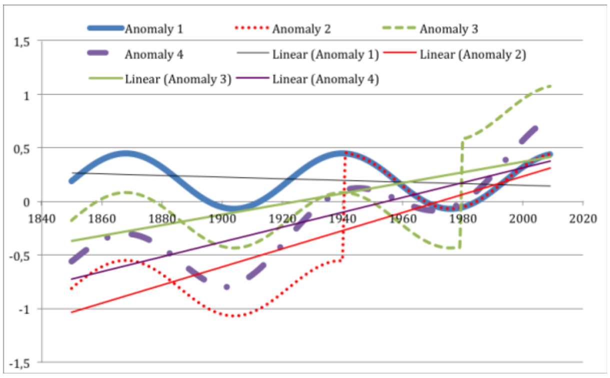
(2) "..There will be a difference between the true mean monthly temperature (i.e. from 1 minute averages) and the average calculated by each station from measurements made less often; but this difference will also be present in the station normal and will cancel in the anomaly. So this doesn't contribute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

In order to see whether this is true and what may happen and further to simplify the action of errors in anomaly time series we will use artificial sine curves instead of real temperature data and will load them by a well-defined systematic errors which appear anywhere within the span of time. Doing this the Excel formula for a sine was used with an added mean value of 14,5 °C. The sine wave has an amplitude of ± 0,25 °C with a cycle time of about 70 years. Figure 6 shows the original "temperature" time series including a trend which is slightly negative, due to selected start- end conditions- though the sine has an trend of zero.

figure 6 artificially computed time series of "global mean temperature" with ± 0,25 °C amplitude and a cycle of about 70 years. Excel calculates correctly a negative trend due to non balanced cycles although the trend for a full sine wave is zero.

On this sine wave two different systematic errors had been superimposed anywhere between 1860 and 2010 i.e. between the duration of the time of observation. The results are shown in figure 7.

The first error applied was a jump of + 1 °C and the result is shown for different times of appearance. The blue



curve shows the anomaly (1) of fig. 2, still without any error. The red curve (anomaly 2) shows the error appearing at 1940 f. e. caused by a moving of the station to a new location. That might be the situation BO 06 had in mind, when they assumed that the error would cancel out. The green curve (anomaly 2) shows the error appearing in 1980 i.e. within the reference time and f. e. caused by a different painting of the station as it has happened in the US when all paintings of existing weather screens have been exchanged from whitewash to latex. This was reported by Watts [11] in his comprehensive report about station data quality around the USA. And the violet curve (anomaly 4) shows the impact of a "creeping" error from 0 to 1 degree over 100 years, as it has happened by the impact of UHI (Urban heat island effect). Referring to the slow changes of the surrounding caused by erecting parking lots, buildings, waste processing or other human driven activities.

figure 7 : Blue (anomaly 1) shows an error free artificially computed time series anomaly of a "global mean temperature" with ± 0,25 °C amplitude and a cycle of about 70 years. The reference value was calculated by the mean of this sine within the time from 1961 to 1990. On this superimposed are separately two different errors. The red curve (anomaly 2) shows the impact of an 1 °C systematic error which appears prior to the reference time span from 1960 to 1991 and continuously acts from that time on. The green curve (anomaly 3) shows the impact of an 1 °C systematic error which appears within the reference time span from 1960 to 1991 and continuously acts from that time on- The violet curve (anomaly 4) shows the impact of a (linear) "creeping" error from 0 to 1 °C systematic error which may happen as the UHI impact which increases slowly as more and more different buildings, parking lots and other influences will impact the temperature of the local weather station. In every case we see an significant change of trend (only two shown) and form of the anomaly time series, depending of the starting point of the error and its type. It shows also, that any error is not self cancelling regardless when it appears, as long as it appears within the time span observed.

We see in all cases that the error remains fully active and impacts the time series of the anomaly value itself by 1:1. The trend change also visible in every case. There is only one exception: If the error appears prior to the begin of the time of examination, then and only then its cancels out. That is the only case where the anomalies cancel out errors within the data.

In all other cases anomalies will not cancel out the errors of its component. Calculation of anomalies requires combining the uncertainties in the temperature measured, as well as those in the calculated mean temperature, which consists of the average of all measured temperatures within the reference time. That obstacle is often not mentioned. Because the anomaly itself is calculated by subtracting from the absolute monthly or annual local average temperature a reference average temperature of the 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 usually determined by applying the formula for the error in the difference between two *i*

Draft

measurements, the root-mean-square (r.m.s.) of the errors in 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 too. But if this condition should be denied or one can´t be sure enough then the simple linear addition of said systematic errors is the right treatment for error propagation (see f.e. 1988; Miller p. 1353) [23] of such errors.

In the first case the error in Δt_i is given by $\pm\sigma_i = \pm\sqrt{e_i^2 + \varepsilon^2}$. So, consisting at least of all the errors of its component.

BO6 had mentioned such a case because in same paragraph it was written:

(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)".*

That this true for all alterations which take place during the observed period of time the authors did not mention. But it is. Furthermore no indication had been given to identify the expected "homogenization" process, how it was done and how successful is cancels the uncertainties, except the few reported above. So, it remains a hope but is not a fact. For this the uncertainties have to be defined and its magnitude estimated. This we will do next.

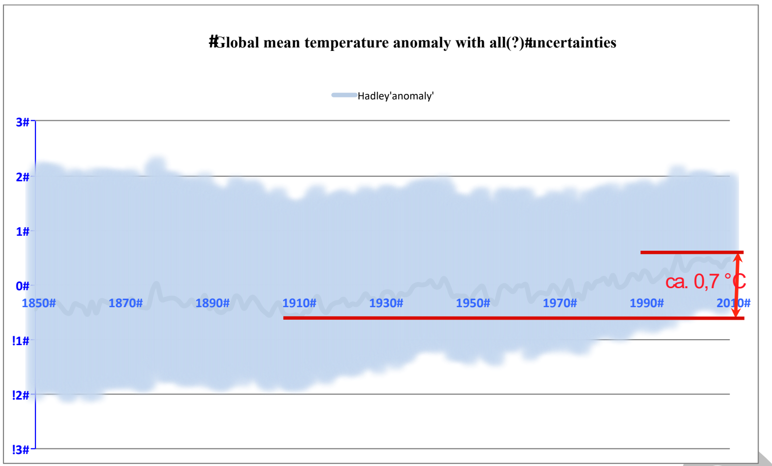
Results and discussion

From previous work of the authors a number of errors had been defined (see list at page xx) and a careful attempt has been undertaken to quantify them very cautiously. The following table show the major errors and its likely magnitude as well as direction.

name of error	**app.min. magnitude in K**	**kind of error**
Min. measurement, reading errror and normal introduced uncertainty (P. Frank 2010(2), 2011(3))	± 0,98	σ = ±0.49 C (1). The 95% confidence interval of these uncertainties alone equals ±0.98 C, discounting the entire global surface air temperature anomaly trend between 1856 and 2004. (before claimed as acc. Folland et al 2001(4) only random with ± 0,2 K); act as systematic
Painting error latex vs. whitewash	+ 0,2	Systematic; permanent; starts from ca. 1980
Instrumental change	+ 0,2	Systematic; erratic starts from ca. 1975
Height variation of Screens	± 0,2	Systematic; erratic but permanent
Sourrounding changes, f.e changes of direct ambient conditions . asphalt; parking lots etc.	+ 0,3	Systematic; creeping, starting point vague
UHI incl. changed land use	+ 0,4 bis + 0,8; 0 +0,5	Systematic; creeping, from 1900 with 0,6 K at end of 20th Ende century
Diff. Averaging algorithms	+ 0,2-0,3	Systematic; permanent (Limburg 2012
SST vs. MAT	+0,2 to 0,7	Systematic; permanent
Poor area coverage land as well as sea	+ 0,5	mean, systematic permanent, in the past much greater than today.

Draft no 1

It must be stated that most of the errors above are based on careful estimations only. But they share this fate with all the trials of error estimates and corrections within the literature as far as the author know after reading plenty of the literature. Due to this and the high magnitude of them the author believes strongly that the total uncertainties shown are rather conservative than over estimated.



Conclusion

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.

figure 8: Estimated uncertainties of major error classes of table 1 around any trend line. The basic data are the same as in fig 1. But the uncertainty is much wider than shown there. In addition, since not all systematic errors rum symmetrically over time the trend line does not follow symmetrical to the envelope boundaries. Since all errors are either stable in time or acting with slow creeping the only signal which can be extracted precisely enough are frequency signals.

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 then they will be reduced or even may cancel out. But in order to know that this may happen, one has to determine these errors fairly precisely in advance.

One may argue that the averaging process itself reduces the impact of one or single station error, regardless of whether it appears in absolute data or in its anomalies. That is true and one of the advantages of averaging. But in meteorological science there are plenty of potential errors known which came up on a great number of stations or all of them at nearly the same time. Examples can be found in the literature, here I name only o few.

1. changing screen design f.e. changing from the tall Wild type screen to the smaller English screen I Russia. Which is reported to happen around 1914.
2. Introduction of other scales from Reaumur (until late 19th century also in Germany) to Celsius.
3. Introduction of other algorithms in order to calculate the daily mean value, this happens very often in Europe as well as abroad.

Draft no 1

As reported in the US but also in other places of the world the change of painting from whitewash to latex white vanish. This list can be expanded easily.

References

[1] [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. 2006.

[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] M. Anderson, “A Field Test of Thermometer Screens,” *SHMI RMK*, vol. 62, 1991.

[5] W. Köppen, “Einheitliche Thermometeraufstellungen für meteorologische Stationen zur Bestimmung der Lufttemperatur und der Luftfeuchtigkeit,” *Meteorologische Zeitschrift*, vol. 30, 1913.

[6] R. W. SPARKS, “THE EFFECT OF THERMOMETER SCREEN DESIGN ON THE OBSERVED TEMPERATURE,” vol. 315, 1972.

[7] P. van der M. Jitze, “A Thermometer Screen Intercomparison,” *WMO Instruments and Observing Methods Reports* , 1998.

[8] A. Sprung, “Bericht über vergleichende Beobachtungen verschiedenen Thermometer - Aufstellungen zu Gr. Lichterfelde bei Berlin,” *Abhandlungen des Königlich Preussischen Meteorologischen Instituts.* , vol. Bd. I. No., 1890.

[9] VDI, “Uncertainties of measurement during acceptance tests on energy-conversion and power plants Fundamentals,” 2000.

[10] C. K. P. FOLLAND D. E., “Correction of instrumental biases in historical sea surface temperature data,” *Q. J. R. Meteorol. Soc.*, vol. 121. , pp. pp. 319–367 , 1995.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] R. W. R. REYNOLDS SMITH, NICKA.STOKES, THOMASM. WANG, DIANEC. WANQIU, “An Improved In Situ and Satellite SST Analysis for Climate,” *J O U R N A L O F C L I M A T E*, vol. 15, pp. 1609–1625, 2002.

[17] R. W. Reynolds, “Sea Surface Temperature (SST) Analyses for Climate.” National Oceanic and Atmospheric Administration National Environmental Satellite, Data & Information Service National Climatic Data Center, 2002.

Draft no 1

[18] G. Dietrich, “Über systematische Fehler in den beobachteten Wasser- und Lufttemperaturen auf dem Meere und über ihre Auswirkung auf die Bestimmung des Wärmeumsatzes zwischen Ozean und Atmosphäre,” *Deutsche Hydrographische Zeitschrift*, vol. 3 5/6, 1950.

[19] J. R. P. Christy David E. Brown,Simon Macadam, Ian Stendel,Martin and Norris, William B., “Differential Trends in Tropical Sea Surface and Atmospheric Temperatures Since 1979,” *Geophysical Research Letters,*, vol. 28, no.1 , pp. 183–186., 2001.

[20] E. C. K. A. and B. Kent David I., “Assessing Biases in Recent in situ SST andMarine Air Temperature,” 1998.

[21] C. Folland, “Assessing bias corrections in historical sea surface temperature using a climate model,” *International Journal of Climatology*, vol. 25, no. 7, pp. 895–911, 2005.

[22] 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.

[23] 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.

Draft no 1