

Research Article

Open Access

A Metrological Assessment of Precision and Accuracy of Earth's Global Mean Surface Temperature Parameter and Some Consequences

Franco Pavese

Former Research Director in Metrology (Thermodynamics) at the National Research Council of Italy

ABSTRACT

The studies on climate change are currently conducted over a specific period of time extending back to the industrial era from 1850 onwards and are intended to comprise the entire Earth's surface. In order to obtain the most recent numerical values of the parameters of interest, they are supplied by Climate Science and, in particular, by Meteorology Organizations. The paper aims to attempt to perform a *strict and full metrological analysis* concerning the specific parameter of Earth's Global Mean Surface Temperature (GMST), considered among the most important ones, as climate behavior is critically dependent on ground temperature. Information on it looks almost invariably to be provided by *non-metrological* Institutions in their published analyses on climate. In many respects, that is a reason for having failed, as the author suspects, to provide a full and comprehensive discussion on the *observation's outcoming* from the meteo Stations (see next Section), their *physical* bases being instead considered of more interest—and possibly the main competence/interest of most authors in analyzing the observed values.

*Corresponding author

Franco Pavese, Former Research Director in Metrology (Thermodynamics) at the National Research Council of Italy.

Received: August 26, 2026; **Accepted:** August 29, 2026; **Published:** September 23, 2026

Keywords: Earth's Global Surface Temperature, Metrological Measurement, GISS, Population, Temperature

Introduction

The studies on climate change are currently conducted over a specific period of time extending back to the industrial era from 1850 onwards and are intended to comprise the entire Earth's surface. In order to obtain the most recent numerical values of the parameters of interest, they are supplied by Climate Science and, in particular, by Meteorology Organizations.

A UN Committee, the IPCC, has also been specifically formed to propagate the scientific results obtained by the above Organizations, but also specifically intended for the political Agencies whose function is implementing the consequent necessary actions [1]. The language used in its reports for the communication of scientific results has been tailored to specific social and economic frameworks, rather than employing the technical terminology of science: numerical values are not provided for the analysis results [2].

Only a set of *qualitative* terms is used, forming what in metrology is called an “Ordinal Scale” (when not a “Nominal Scale”), consisting of words corresponding to ordered steps of increasing probability level of importance/urgency-of-action of the observed phenomena, here reported in Table 1.

In particular, the specific measurement science numerical terms “precision” and “accuracy” (see later) are not used, but only “uncertainty” in the generic sense of doubts about the numerical values. However, in modern civilization and science, the evaluation of acquired knowledge allowing consequent actions, scientific or economic, is always based first on *empirical/experimental* results obtained by *measurements* providing *numerical* values of the relevant physical/chemical/natural

Further, Climate Science is used in modern times to assemble new knowledge with the help of informatics in what is called Big Data. This term is intended to mean so large datasets that the overall state-of-the-art can hardly become accessible to the whole scientific Community, especially when intending to draw general analyses from these immense and heavily elaborated amounts of information—today possibly unless with the help of an indirect new tool, the AI [3,4].

The paper aims to attempt to perform a *strict and full metrological analysis* concerning the specific parameter of Earth's Global Mean Surface Temperature (GMST), considered among the most important ones, as climate behavior is critically dependent on ground temperature. Information on it looks almost invariably to be provided by *non-metrological* Institutions in their published analyses on climate. In many respects, that is a reason for having failed, as the author suspects, to provide a full and comprehensive discussion on the *observation's outcoming* from the meteo Stations

Table 1: The Likelihood Scale Introduced in the AR5 Guidance Note

Term used	Likelihood of the outcome
Virtually certain	99–100% probability
Very likely	90–100% probability
Likely	66–100% probability
About as likely as not	33 to 66% probability
Unlikely	0–33% probability
Very unlikely	0–10% probability
Exceptionally unlikely	0–1% probability

(see next Section), their *physical* bases being instead considered of more interest—and possibly the main competence/interest of most authors in analyzing the observed values.

The Proposed Analysis Method

The author, a metrologist, considered it useful to start attempting from the very beginning, before any further elaboration and contrarily to most literature, from the specific issues of the classical method of knowledge-accumulation process in science: the analysis of the *numerical values of observations quantitatively* obtained via *calibrated* measuring instruments, strictly evaluated by means of measurement-science tools and by using the metrological method and idiom—in particular, without any help from what is commonly considered the complementary/alternative way to analyse phenomena, their analysis/qualification via physical modelling.

Considering the criticality that the results may have on the subsequent political choices of the World Societies, superior confidence in data accuracy should be ensured in these Natural Science studies. The Metrology discipline and process are known to provide it, even with respect to Measurement science [5]. In fact, metrology aims to push the accuracy level of the experimental results obtained in the wider frame of Measurement Science to its best and with maximum confidence—i.e. to ensure a minimum level of *uncertainty* that *always* necessarily affects any observation/experimental results [6].

Its Major Task Is to Explore and Assess an Estimate of All Components of Uncertainty

The *reproducibility* of repeated measured values, assumed to be random, basically representing the *precision* of the results—often erroneously considered sufficient to qualify them; —The systematic *errors*, deriving from the *process* chosen for performing the measurements, whose lack of consideration would bring to non-random components of uncertainty— so often remaining implicit in the resulting data *quality*.

Consequently, the only possible way to increase *accuracy* is to perform deeper studies also on the *process* factors having assumed to influence (being the causes of) the measured value. This basically requires a deeper understanding of the phenomena being measured—also reported in [7-14]. The final result of the evaluation process consists of providing *full* information about the components of uncertainty, expressed in a *full* “uncertainty budget” (UB): a list of all the components with an estimate at 95% (or 97.5%) level of confidence §3.10 in contributing to the total uncertainty. Only a UB can provide a way to compute the accuracy [32]. The author’s position on the above concepts is exemplified in [7-16].

The Organizations Involved in the Studies and their Methodology Among the several International Organizations performing these studies, the World Meteorology Organisation (WMO) is specifically dedicated to meteorology [17]. Since its foundation by the ONU in 1946, WMO is assumed to act as the reference to all the other Organizations. The hierarchical approach becomes necessary for these in-depth studies because achieving only good data precision— not always easy—is insufficient on its own. Additionally, the highest level of accuracy currently available must be provided. In this context, the WMO is assumed to be responsible for supplying the most accurate meteorological data available for the entire Earth's surface. The National Organizations should in principle refer their studies to WMO. A network of Meteorological Stations ensures the collation of observations on the ground surface. In Fig. 1a-d the records *obtained from Stations* from various Organizations are estimated to have been a total of 81,555 in 2012, while the present number of WMO *Stations* is ≈8,000. Fig.1 shows a max of ≈5,000 of them around the ‘70s and ≈half of them as “recommended” in 2010 [18]. The number of “under development” Stations increased too almost 20,000 in the same period. The average Station density on the ground has been evaluated by the author—using his method of (red)-pixel counting (see later) in Fig.1d—to have reached in 2012 an average coverage of ≈2/3 of the ground surface, amply not uniformly distributed.

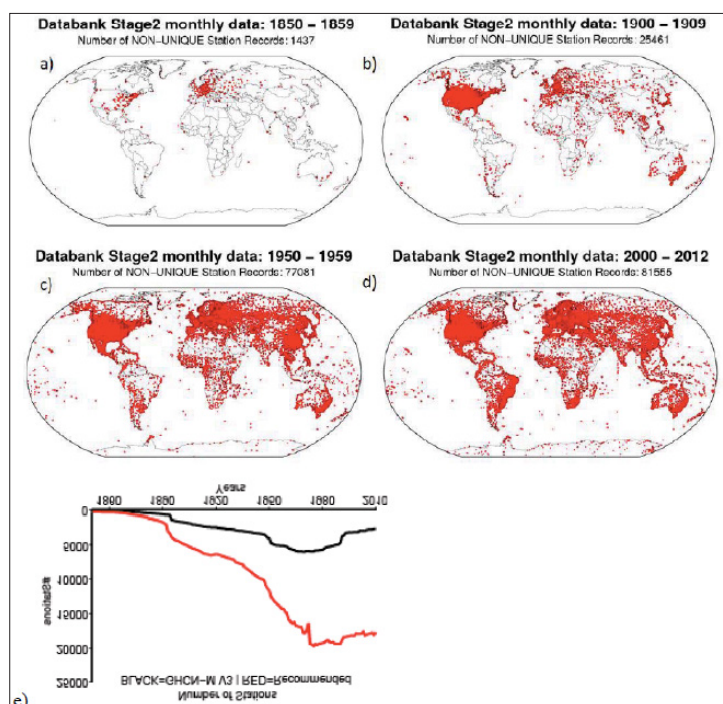


Figure 1: (a–d) Surface Coverage of the Ground Meteorological Stations, from 1850 to 2012 (Lawrimore 2013). See text (e)



Figure 2: From Fig. 1, the WMO Meteo Stations in a Portion of Europe are Shown

In Fig. 2 the distribution of the WMO ground Stations is shown, where they are quite variable in density [19,20]. It is the simplest way to access the published database (the manipulated one?). The meteo Stations comprise quite a variety of technical solutions, adopted to try to limit the effect on the accuracy of systematic errors affecting the *observational* data—also depending on their position and the age since their installation. Their density on the ground is always insufficient to the one required for specific applications (like for aviation), but, in general, their distribution is insufficient also for an accurate evaluation of the GMST when exclusively using the data they are providing. Consequently, various procedures have been adopted to make the examined network spatially denser via computational means already recalled above [21,22].

The resulting spatial database then used to compute parameters and draw overall estimates, is necessarily made of values *more uncertain* than the observational (original) ones—while certainly, the uncertainty of a mean value is significant only if computed on *repeated* data, not the case for the GMST observations. In all instances, no detailed information is usually available for comparing the uncertainty of the original observation set and the uncertainty of the manipulated/denser final set. In other words, the information to the scientific community is limited to the manipulated database, a Big Data one considered to represent the global parameter GMST. See later, e.g., Fig. 5, where the *precision of the physical models* can be deduced at least from their dispersion, while for the observational data only the annual mean values are supplied with *not* even their s.d. indicated.

The original observational datasets obtained by each Organization are normally indicated only by their name: e.g., in Fig. 3a five of these sets are reported from—in their usual form of representing the trend in time of annual (decadal) mean [23,24].

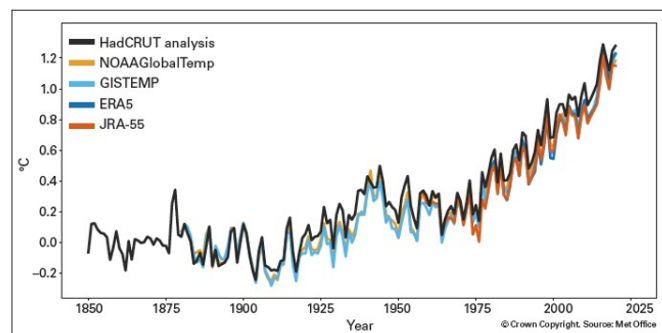


Figure 3: (a) Decadal Trend of Annual GMST as Reported by Five Databases

Many other similar collations can be found in the literature concerning other datasets, like the ones in Figs. 2—>3b—*relevant* differences among them can easily be noticed [23]. Note also the lack of indication about data/trend uncertainty, so that the *precision* of the manipulated databases (see later) can be estimated only from the variability between series—arising from a different reason: the trend in time. Similar lack can be found also in authors with different positions, like in [25].

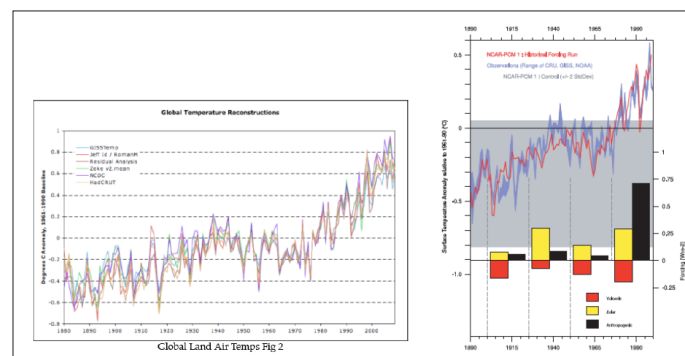


Figure 3: (b) It Reports from a Different Set of Databases: b1); b2) ISBN: 0-309-66144-7

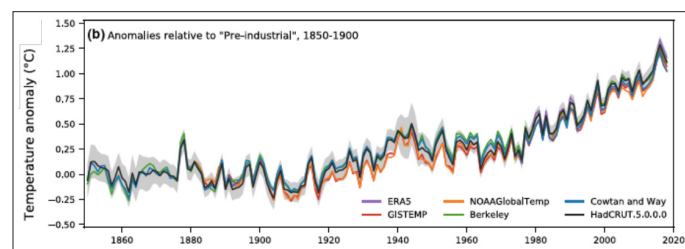


Figure 3: (c) It Reports from A Still Different Set of Databases

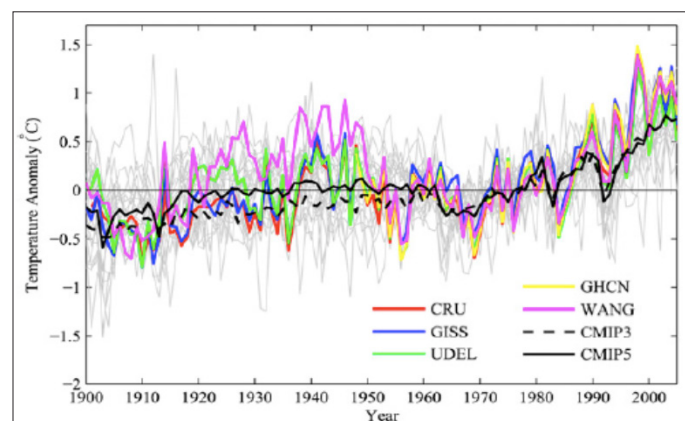


Figure 3: (d) It Reports from A Still Different Set of Databases

Note in Fig. 3d a few downward peaks ≈ -0.2 °C deep, generally due to occasional big volcanic eruptions [23,24]. Note also the less evident linear GMST increase after ≈ 1970 . The main purpose of the process evaluation is with respect to its evolution in time, i.e. concerning the possibility of predicting a reliable climate forecast. However, *only* data from WMO Stations can provide today's level of *data accuracy*. For that purpose, the most recent advancement has just been recently published as an ISO Standard (ISO19289 2014): a Project of classification of siting, the WMO Meteorological Stations, in four *certified* classes of GMST accuracy. Specifically, they correspond to four levels of accuracy for temperature: Class 1-2 ± 0.2 °C; Class 3 ± 0.6 °C; Class 4 ± 1 °C [17]. The number of *certified* Stations is presently *still quite limited* as the certification started only about 10 years ago; the previous accuracy was for all ± 1 °C—but probably was

even larger than the accuracy not so many previous decades.

Analysis by Using a Metrological Procedure

A metrological analysis of the GMST measured trends should start from the *original observational datasets* and include their uncertainty level. That resulted *impossible* because *the previous analyses do not report nor illustrate them*. Instead, it is easy to find in the literature indications that the reported results are reported by not using the original daily observations. The *measured data* have usually been amply manipulated by techniques like “data assimilation” and Foot Note 8) or data “homogenisation”, etc. and by also performing an increase of the density of the database by interpolation between Stations, extrapolation in regions where no Stations exist, etc., i.e. adding not real stations [21,22]. Each of these manipulations necessarily contributes to increasing total uncertainty, but no UB is generally generated nor made available—neither for the actual observational databases. Therefore, there are few—or no—means for a reviewer analyzing published data to check the confidence that one can assume to assign to an accuracy evaluation because never or very rarely do the papers include (full) information on both random and systematic effects.

In order to try an alternative way to keep evidence of the *original* Station datasets, the author had recently developed a method of direct analysis from Earth maps deploying graphically the distribution of the values of a specific climate parameter on them, by introducing a computer graphic method [16]. The superior total amount of information allowed by maps can be appreciated in Fig. 4, comparing two total-surface GMST distributions from the same Organization: the complex distribution and some impressing changes for the same period of time often occurring in the representations, would be difficult to be perceived from a numerical-only treatment [26].

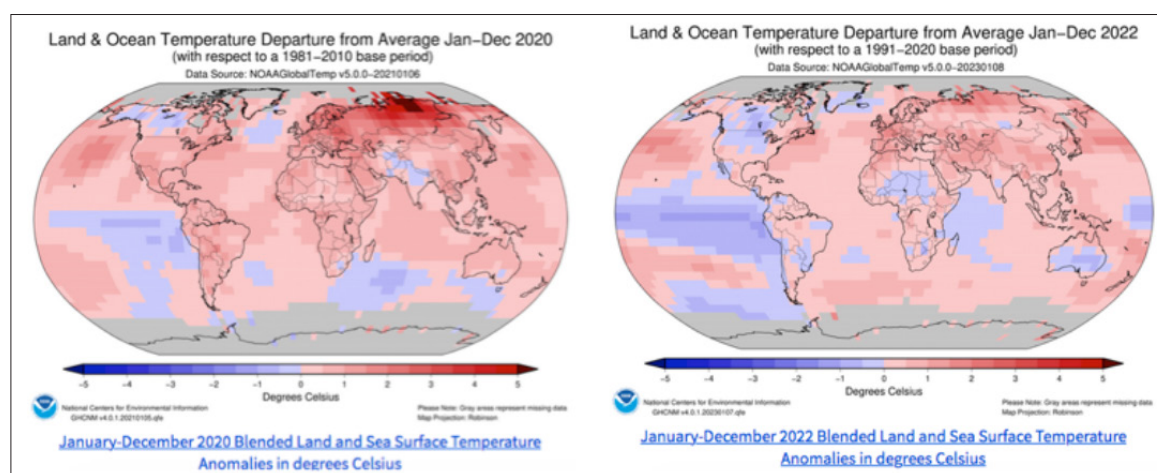


Figure 4: Maps (of the same type) from different periods: (a) 2022 (1981-2010); (b) 2022 (1991-2020). (NCEI NOAA 2020)

Author’s method basically consists of a *quantitative use of Earth maps*. The first step is the retrieval from each map of the distribution of the *numerical values* of a given parameter as qualified in narrow ranges by differently coloured *regions*—each colour representing a Δx interval of values x : e.g., for the GMST often steps $(t_2 - t_1) = 0.5^\circ\text{C}$. The pixel count within each area of a specific colour allows one to retrieve the numerical proportions between the different intervals Δt ’s regions and then to compute the GMST mean value (i.e. at 50% Δt occurrences) for that map [26].

In Fig. 4 of this paper, a similar principle and method have been used for comparing instead the GMST values obtained by several sources, starting from published figures where data trends are similar to the ones in above providing trends like in Fig. 3 [27,23,28,24]. It compares what is assumed to be the set of monthly GMST data for the same period of time, 5 as published by 5 World Organizations: HadCRUTS, MSU RSS, JMA, NOAA NCEI and GISS [29]. The full period 1850-2023 is considered and the five trend lines were extracted from the publication, as identified by their pixel colours, and compared. All indicate 10-year moving averages. The extracted trend lines were first synchronised on the same scale having 2023 for the last year and are reported in Fig. 5a. Some differences up to $\approx 0.3^\circ\text{C}$ between them can be observed. Then, in Fig. 5b the trends were *translated* with respect to each other to *minimise* their differences for the period starting from about 1980 on, where the precision of the data can be assumed to be the best, while in Fig. 5c the monthly means are added to the 10-year running mean.

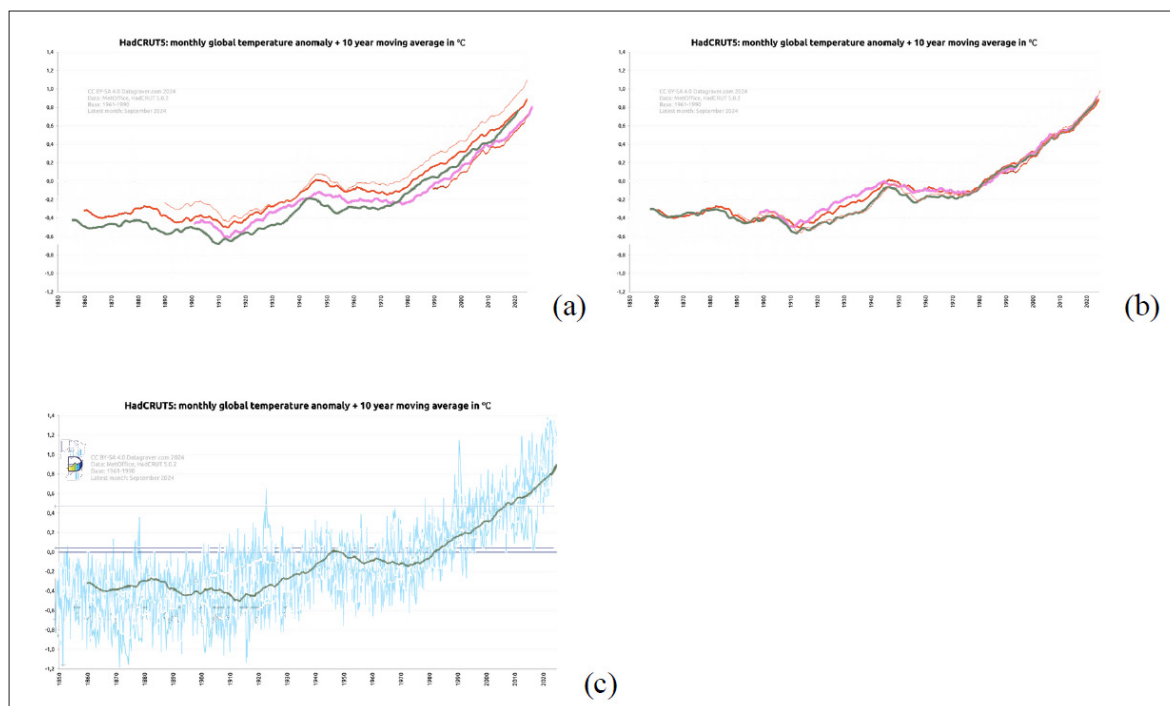


Figure 5: (a) Five trends synchronised for end at 2023: HadCRUTS, MSU RSS, JMA, NOAA NCEI and GISS (note how well the trends reproduce also the same minimal details); (b) Same trends but with differences minimised at >1980 by translation; (c) HadCRUTS trend (b) with monthly trends added.

GMST Estimated Models

The elaboration of the published GMST data is also critically confronting with *climate physical models*, not rarely even superseding the analysis of experimental data. In Fig. 6 such a kind of different approach is evident—which may facilitate metrological confusion. In such a typical case, in fact, it is quite common to find evaluations of *precision possibly only related to the models*, arising from model-set dispersion 6—see e.g. [30,31].

This fact can too easily be a reason for confounding model precision evaluation, generally provided for the forcing factors contributions, with the observational precision. The number of studies on *physical forcing* effects is by far larger than the analyses of *observational* precision (accuracy is also too often confused with precision!). In that Figure, while 44 *models* show a precision range that can even exceed a band of 1 °C, the *observational* means (visually) show a precision of a few tenths of °C, a precision inconsistent with the model mean trend—no elaboration is reported of the latter estimate and no information at all about accuracy is supplied [22].

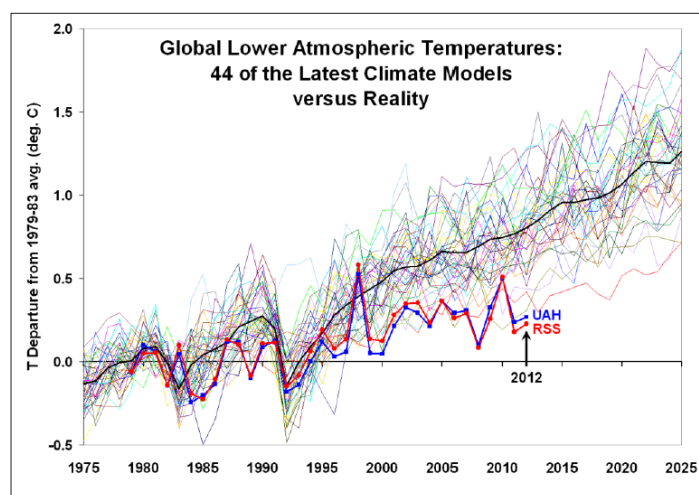


Figure 6: Comparison of 44 physical *models* for the period 1975-2025, and the annual *observational* means (the “reality”) of two different datasets for the period 1975-2012

In addition, it is common to find in the literature metrologically incorrect information. For example, in a rare study having expressed the correct position since its Abstract (“The estimated average ± 0.2 ($^{\circ}$) C station error has been incorrectly assessed as random, and the systematic error from uncontrolled variables has been invariably neglected), then *in its* Ref. 11 reported: “The random error in a single thermometer reading is about 0.2 ($^{\circ}$) C (1σ)”, just disproved [32,23]. Also reported in its Ref. 12: “The monthly average will be based on at least two readings a day throughout the month, giving 60 or more values contributing to the mean. So, the error in the monthly average will be at most $= 0.03$ ($^{\circ}$) C”: that is incorrect because these are *not* repeated measurements [33].

It is a typical example of a kind of treatment that is exactly *the reverse* of what a metrological treatment should provide/expect. Apparently, much more effort is spent in using mathematical/physical studies than they would be worth and needed for a meaningful comparison of these studies with the experimental data: the latter cannot have an accuracy lower than that of the WMO-stations period prior to 2015, i.e. ± 1 $^{\circ}$ C, amply consistent with model sets ... but, on the other hand, making an accurate forecast impossible.

Finally, the author found in the literature explicitly reported the GMST numerical uncertainty estimate only in (IPCC 2022) as follows:

IPCC AR6 Technical Summary

For the decade 2011–2020, the increase in global surface temperature since 1850–1900 is assessed to be 1.09 [0.95 to 1.20] $^{\circ}$ C. Throughout the WG1 report and unless stated otherwise, uncertainty is quantified using 90% uncertainty intervals. The 90% uncertainty interval, reported in square brackets [x to y], is estimated to have a 90% likelihood of covering the value that is being estimated. The range encompasses the median value and there is an estimated 10% combined likelihood of the value being below the lower end of the range (x) and above its upper end (y). Often the distribution will be considered symmetric about the corresponding best estimate, but this is not always the case. In this Report, an assessed 90% uncertainty interval is referred to as a ‘very likely range’. Similarly, an assessed 66% uncertainty interval is referred to as a ‘likely range’. (The term “uncertainty” is ambiguous, as it does not clarify whether it refers to *precision* or *accuracy*, however, given its subsequent labelling as “very likely”, it can be inferred that it is an assessment of accuracy).

Incidentally, the author performed a fit on a HadCRUTS-published numerical database for the same time interval and found its standard deviation (s.d.) 0.2-0.3 K, similar to the one reported as “uncertainty” in most publications. A fit however only allows evaluating the dispersion of the dataset, i.e. the precision of the fit, *not the accuracy*.

GMST Precision and Accuracy Evaluation of Observations According to Metrological Analysis

In the following, the case of the GMST parameter will now be considered based exclusively on meteo Station *observations*, in order to exemplify the situation for that global parameter—but the resulting considerations are possibly valid also for other parameters. Instead, the precision/accuracy of climate *mathematical models* will never be considered nor discussed anymore.

Accuracy assessment is the subject matter of the present analysis. In line with this aim, the comments provided here are directed

solely at estimating *accuracies*, without addressing aspects related to (stated) *precision*, except when strictly related to accuracy. The following features can thus be reported (in the absence of information about the identity of the used databases), when considering the above Fig. 5:

- While an *exact* synchronisation i.e. the consistency of the 5 sets is difficult, the trends look quite similar even in small details, though often showing in some parts a difference with each other higher than the reported uncertainty;
- The minimisation of the trend differences since about 1980 up and below 1910 required shifts up to ≈ 0.3 $^{\circ}$ C to make all trends consistent to within 0.1 $^{\circ}$ C, while a divergence remains higher in the intermediate period;
- The bump between ≈ 1940 –1955 of about $+0.2$ $^{\circ}$ C is due to the effect of II World War, especially evident in Fig. 4b of Section 3: if one takes that anomaly due to II World War away, the increase between 1980 and 2000 looks basically linear, with fluctuations certainly within the true accuracy [23]. The period 1945-1975 only shows the recovery from the anomaly, making dubious the normally considered size of the anthropogenic effect. The bump is less evident only in some cases, as HadCRUTS (Fig. 5a), where a mere maximum at 1947 is shown, followed by a subsequent decrease until 1970. However, notice from the monthly data in Fig. 4c that the *real* peak was during the war 1940-45, followed by a stepping lowering down to the same pre-war value. That is most clear in NOAA NCEI (Fig. 5b), according to the monthly data. The following period up to 1970 shows the same trend as the period since 1910, the minimum for the full period being at -0.6 $^{\circ}$ C, for a total increase of $+0.4$ $^{\circ}$ C
- The manipulation of the original databases and the strong averaging of short period masks divergences amply within the accuracy of the data before 1950 that could have occurred related to shorter-term fluctuations and the actual accuracy of the data. Account being taken of the latter respect, the full period can be subdivided into only *three periods*: 1855-1910 decrease -0.3 $^{\circ}$ C (0.055 $^{\circ}$ C/decade, oscillations well within accuracy); 1910-1975 increase $+0.4$ $^{\circ}$ C with only the II World war *anomaly* (0.062 $^{\circ}$ C/decade, still well within accuracy); 1975-2023 further $+0.7$ $^{\circ}$ C (total $+1.1$ $^{\circ}$ C, 0.146 $^{\circ}$ C/decade).
- Considering issue 4), there is *no real increase* of the GMST until about 1975, only fluctuations, excluding the real and justified II World War anomaly.
- Contrasting with the previous time periods, the almost linear increase of the GMST after ≈ 1975 is basically identical in all treatments (but see also Fig.3c), looks as the only clear indication of a GMST almost linear real increase since ≈ 1970 and up to 2020+ (if real see Footnote 10), i.e. during ≈ 50 years, by about 0.8 $^{\circ}$ C.

There is *no divergence*, e.g. in Fig. 3d, between the trends of the five organizations from 1970 up. This is a strong indication that the manipulations (e.g. see Foot Notes 10-11) performed on the original database were determinant in *cancelling* all information about the actual *accuracy*, which has certainly lowered by a factor of at least $\times 10$ in 160 years: that consequence looks incorrect from a scientific point of view. In fact, the accuracy of most of the measurements is necessarily still closer to the original WMO ± 1 $^{\circ}$ C accuracy, for the full period until after the ISO/WMO decision of ≈ 10 years ago—where the present dispersion of almost half the one measured before ≈ 1920 [34]. A similar *larger* data dispersion can also be observed from the simple fit of the annual GMST observations, as done by the author on a couple of public databases.

Thermometry Accuracy as a Basic Factor

The Meteo Stations are normally not equipped for only measuring temperature. However, *from the vast experience of the author on this type of metrological measurement*, this quantity is probably the most *difficult to measure accurately*. The reason is that in temperature measurements systematic errors are probably prevailing on any other component of uncertainty. For modern thermometers, obtaining a calibration precision better than 0.1 °C is no longer a problem, and also the stability in the time of the calibration in principle is not difficult to ensure over reasonably long periods. However, the required measurements basically concern assigning a temperature value to a “portion of air” *typically* that of a small fraction near the ground of the chosen location where the Station is

Air Problems

- The same air composition “external” and “internal” to the calorimeter is not strictly required.
- The sample of (calorimeter) internal air, let's say ≈ 1 L, must have the *same* temperature of the external atmosphere of the station location within, say, ≈ 0.1 °C, uniformly.
- The internal air sample must track, also in time, the composition/physical properties of the external air.
- Internal air convection should *not* occur according to correct “isothermal” conditions of the calorimeter, i.e. no *external* energy sources should reach the internal volume (thermometer dissipation when active can be reduced to low enough value).
- Condition also involves calorimeter/thermometer shielding from external radiative energy, differently from the external air always involving surroundings radiation, one of the major causes of temperature systematic errors.
- Easier is the exclusion of direct effects of rain/snow/wind on the internal air.
- Several models of Meteo Stations have been devised in order to get much better approximations to the above conditions, which are so difficult to obtain, namely b), that the very physical- and geo- location of a meteo Station is very critical in many respects—one important reason for making necessary considering the three Classes of accuracy and to certify it.

Thermometry Problems

- Calibration of the thermometer and its stability for a reasonably long time are necessary conditions: no problems today for up to better than 0.05 °C.
- If frequent re-calibration cannot be ensured, stability at the time of the calibration cannot be proved. It may be less costly to use *two* thermometers (min), of different models (so possibly behaving differently, allowing a cross-check). This solution is often used in the field of air pollution for the relevant apparatuses, doubled just for that reason.
- Thermometers observed to drift in time should be excluded temporarily and then replaced.
- By using two thermometers one increases confidence in the results.
- There are *pros and cons* in placing the two thermometers close to each other, as implicit from the above considerations.

However, spatially placing them variously in a different way in the calorimeter can increase confidence about a limitation of possible systematic errors due to residual systematic temperature errors in the Station.

m) 3 thermometers would allow both above checks.

Use of the Temperature Observations

- A single measurement made daily is the minimum, if its daily time can be considered sufficient to take into account the meaning attributed to it, typically the *day-mean value*; increasing accuracy to 0.2 °C makes this choice questionable concerning the significance of the daily value.
- Repeating the measurements daily does not necessarily solve all the problems. In fact, multiple measurements made during a single day *cannot* be considered “replicated measurements”, because, differently from the oceans, the air temperature on the ground normally changes continuously and naturally. Thus, also the mean of these values cannot be treated as “replicated measurements” on different days.
- Repeated measurements could be obtained *only* in rapid time sequences: actually, their results only check the *repeatability* of the thermometer, basically useful only for detecting anomalies due to any deterioration in the apparatus (the author would not include natural causes, apart ... birds/humans interfering with the station, lightning, a wind gust, measuring equipment, ...).
- The number of anomalies may increase in stations also depending on their location, e.g., not inhabited, close to water/sea, strong-wind location, and in general arising from the turbulence level of the external air and its temperature gradients—which should be taken into account in estimating the actual accuracy.

Comparison of the Analyses & Conclusions

The main problem with the GMST accuracy, arising from the above literature analysis, is the fact that there is not a reported univocal or clear origin of the precision/accuracy estimate of the SAMT value and uncertainty.

In principle, after the '50s, the WMO should have been considered not only paralleling the local/national organizations but also contributing to that estimate with its own Meteo Stations. The other organizations should be considered only “participants” with their own Stations to the worldwide reference data. The author was unable to find a clear and univocal identification of the distribution on the surface of different sets of Stations corresponding to the reported different sets of data.

The above problem is vital for two reasons: since its foundation in 1946 only the WMO is supposed to be universally recognized as *an* International Organization in that respect, so the WMO only should be considered to provide an international reference. Only the WMO has been responsible since about a decade of the Project for assessing in 4 Classes the *accuracy* of *its* Stations, where their *accuracy* may result to be substantially *better than precision*—thanks to only the *two* classes deserving the ± 0.2 °C accuracy, probably being at present still a minority—see Section 3.

In addition, it comes quite evident from the literature that the original numerical values provided by each Station, forming the *real* observational set, are “*as lost*”, since they are subjected to extremely heavy manipulations, without the details being available for independent evaluations. In principle, the reasons for that are sound and scientific, as intended to increase the reliability of the whole dataset by subtraction of doubtful data and addition of new data to obtain a reasonably uniformly dense dataset. However, from a metrological viewpoint, any kinds of data manipulation concerning single stations/datasets will alter/increase its UB. As a *minimum* requirement, both datasets should be reported: the *original* (“*raw*”) for the originating set of Stations/organizations, and the *final* after manipulation, intended to be the basis of the

published analyses. Otherwise, the accuracy of the information, if any, of the original Station will go lost and back-untraceable.

In addition, the geographical distribution of the Stations contributing to each GMST estimate must be reported on a map, if an estimate of human forcing on the climate is a goal of the analysis and would be taken into account.

In fact, one needs to check the numerical consistency of human population and its distribution on ground, e.g., as reported in Fig. 7 below, where it can be visually compared with Fig. 1 and the GMST in Fig. 8—of the same Mercator/Robinson map type: see why the same type and the needed best choice of a map type [29,19,24,27].

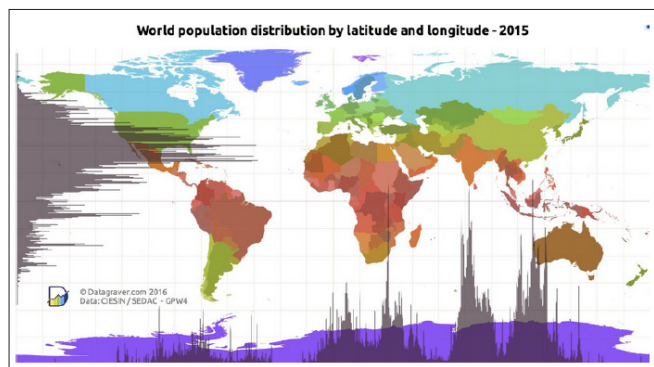


Figure 7: Human Population Distribution on Land

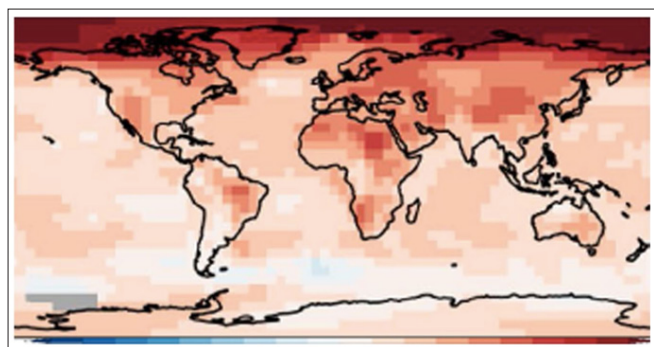


Figure 8: The low resolution is original of the data as reported in the map, in small surface regions with the same temperature value, according to the colour scale reported at the bottom (t step width $0.5\text{ }^{\circ}\text{C}$)

The main suggestion that arises from the above analysis is that the uncertainty attributed (e.g. by IPCC) to the global parameter GMST is *not* scientifically based if intended to report the *accuracy* of the available data, resulting today still being closer to $\pm 1\text{ }^{\circ}\text{C}$ than to $\pm 0.2\text{ }^{\circ}\text{C}$ (and even to $\pm 0.64\text{ }^{\circ}\text{C}$) [35].

This fact makes the reliability of the present forecasts questionable: scientists should become more *cautious in the prediction* of future trends, also considering that the constraints to future trends imposed by the knowledge of history extend back by max only ≈ 50 years, not by 100-160 years: in fact, the only sufficiently long period of time showing a consistent trend, not justified by a mere increase of precision of the measurements, possibly started only from ≈ 1880 . 480.

In conclusion, from the previous metrological analysis the author is still considering, at least for the next couple of decades, the accuracy of the present GMST increase more reliably if limited to $[+1.1 \pm (0.5 \text{ to } \sim 1)]\text{ }^{\circ}\text{C}$ —also based on WMO indications [20,36,37].

The main consequence is that the reliability of the forecasts is strongly limited to a band of uncertainty where most of the present models are already comprised [38-40]. In all instances, reaching the goal of a GMST accuracy of $\pm 0.2\text{ }^{\circ}\text{C}$ with sufficient confidence looks like “a vast programme” as also model precision should become higher [41-45]. On the other hand, also in the metrological respect, the possible trend of GMST linear increase since 1970-80 cannot be so far found justified in the literature even by models: the author was able to find only a single Reference about that [46-50]. The author suggests that climate scientists should supplement their physical/chemical analyses with shared reasons also for such a fact [51].

APPENDIX A (Global Climate Highlights 2023)

The Paris Agreement and uncertainty in the warming from 1850-1900 to recent decades “It is clear from the report of the dialogue and parts of the Working Group I (WGI) and II (WGII) contributions to AR5 that it was understood that an average temperature for 1850-1900 (or a nominal temperature for 1880) had to serve as the pre-industrial reference due to the absence of adequate instrumental data back to the start of industrialization in the 18th century; the temperature thresholds were based on model projections of future change relative to a 1986-2005 baseline; there was only one dataset available that related the 1986-2005 average global temperature to the 1850-1900 temperature; it gave a difference between the two periods of $0.61\text{ }^{\circ}\text{C}$, with an uncertainty range (90% confidence interval) from $0.55\text{ }^{\circ}\text{C}$ to $0.67\text{ }^{\circ}\text{C}$ [see this paper text].

The Paris Agreement could thus equivalently have been made with the aim of “holding the increase in the global average temperature to well below $1.39\text{ }^{\circ}\text{C}$ above its 1986-2005 level and pursuing efforts to limit the temperature increase to $0.89\text{ }^{\circ}\text{C}$ above the 1986-2005 level”. This formulation would have rendered the target levels explicitly independent of changes in estimates of the (then and now) rather uncertain 1850-1900 “pre-industrial” level.

Since 2015, these targets of the Paris Agreement have generally been interpreted literally. In particular, the IPCC’s Sixth Assessment Report (AR6) estimated a temperature rise of $0.69\text{ }^{\circ}\text{C}$ from 1850-1900 to 1986-2005, an increase of $0.08\text{ }^{\circ}\text{C}$ over AR5. The uncertainty range was larger than in AR5: $0.52\text{ }^{\circ}\text{C}$ to $0.82\text{ }^{\circ}\text{C}$ for surface air temperature [see this paper text]. WGI of the AR6 chose to use the $0.69\text{ }^{\circ}\text{C}$ value to estimate projected times of reaching the $1.5\text{ }^{\circ}\text{C}$ threshold, although WGII recognized on page 142 of its contribution to AR6 that the impacts of climate change associated with a given threshold level of warming would change due to the changed estimate of temperature rise from 1850-1900 to the recent past.

C3S has followed the consensus literal interpretation of the targets of the Paris Agreement. In particular, it adds $0.88\text{ }^{\circ}\text{C}$ to annual-mean temperatures relative to the 1991-2020 reference period to convert them to temperatures relative to 1850-1900. This estimate derives from the approach used in the WMO statements on “The state of the global climate” from the Preliminary Statement for 2021 onwards. The value $0.88\text{ }^{\circ}\text{C}$ is based on adjusting the AR6 figure of $0.69\text{ }^{\circ}\text{C}$ to account for the difference between 1991-2020 and 1986-2005 temperatures.

It appears from the above that the intended targets of the Paris Agreement have been shifted by $0.08\text{ }^{\circ}\text{C}$, making them more stringent, by interpreting the wording of the Agreement literally rather than by taking account of its scientific foundation. A further important result from AR6 needs to be considered, however. The

Summary for Policymakers of the WG II contribution states in finding B3.3: “Levels of risk for all Reasons for Concern (RFC) are assessed to become high to very high at lower global warming levels than in AR5 (high confidence)”. The comparison between the AR5 and AR6 assessments presented explicitly in Figure 3.3 of the AR6 Synthesis Report shows that the two assessments differ by several tenths of a °C in their estimates of the temperature rise since 1850-1900 that corresponds to a given level of risk, with the precise difference dependent on the RFC. The increased stringency of the targets of the Paris Agreement that results from higher estimates of historical warming thus only partially accounts for the need for increased stringency due to the perceived risks of global warming being larger now than at the time the Agreement was made.

Uncertainty in estimates of recent global-average temperatures evidently has to be considered in assessing the approach of the atmosphere to the target warming levels of the Paris Agreement. However, the considerations presented above, most especially those outlined in the WGII contribution to AR6, and an assessment of the legal character of the Paris Agreement's 1.5°C temperature goal, indicate that the quite large uncertainty in the warming from 1850-1900 to the recent past need not be taken into consideration, as any future revision of this warming should be cancelled by a corresponding change in the target warming levels, assuming the assessment of risks does not change.”

References

1. IPCC- Intergovernmental Panel on Climate Change (2021) Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Masson-Delmotte, V. et al. (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA 1-2391.
2. Mastrandrea MD (2011) Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). IPCC guidance notes, Jasper Ridge, CA, USA 7 <https://wmo.int/about-wmo/overview>.
3. Hilbert M, López P (2011) The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 332: 60-65.
4. Wikipedia (2025) Artificial intelligence. Consulted February https://en.wikipedia.org/wiki/2025_in_artificial_intelligence.
5. Peirce CS (2011) Philosophical Writings of Peirce. Buchler, J., Ed.; Courier Dover Publications: New York, NY, USA <https://ia801507.us.archive.org/24/items/in.ernet.dli.2015.191017/2015.191017.Philosophical-Writings-Of-Peirce.pdf>.
6. Peterson TC, Vose R, Schmoyer R and Razuvaev V (1998) Global historical climatology network (GHCN) quality control of monthly temperature data. *Int. J. Climatol* 18: 1169-1179.
7. Pavese F (2007a) Accounting for systematic effects in metrology and testing, namely in comparisons' Proceedings of ENBIS-DEINDE General Conference, Dortmund.
8. Pavese F (2010a) Comparing statistical methods for the correction of the systematic effects and for the related uncertainty assessment. 13th IMEKO TC1-TC7 Joint Symposium, Journal of Physics: Conference Series 238: 012041.
9. Pavese F, Baer M, Filtz JR, Forbes AB, Pendrill (2012a) On the difference of meanings of 'zero correction: zero value vs. no correction, and the associated uncertainty. AMCTM IX Eds. L. and Shirono K., Series on Advances in Mathematics for Applied Sciences World Scientific, Singapore 84: 297-309.
10. Pavese F (2012b) Corrections and input quantities in measurement models, *IJMQE* 1: 155-159.
11. Pavese F (2016) The probability in throwing dices and in measurement. *Acta IMEKO* 5: 1-6.
12. Forbes AB, Zhang NF, Chunovkina AG, Eichstädt S, Pavese F (2018) On the classification in random and systematic effects, AMCTM XI, 2018, in Eds. 'Advanced Mathematical and Computational Tools in Metrology and Testing XI', Series on Advances in Mathematics for Applied Sciences World Scientific, Singapore 89: 58-69.
13. Pavese F (2020a) About a new concept diagram for the measurement process and related terms. *Ukrainian Metrological Journal* file:///C:/Users/admin/Downloads/217689-%D0%A2%D0%B5%D0%BA%D1%81%D1%82%D1%81%D1%82%D0%B0%D1%82%D1%82%D1%96-492652-1-10-20201202.pdf.
14. WMO (2021) Report No. 1264. State of the Global Climate 2020 World Meteorological Organization, 2021. Copernicus Climate Change Service, European Centre for Medium-Range Weather Forecasts (ECMWF) <https://public.wmo.int/en/ourmandate/climate/wmo-statement-state-of-global-climate>.
15. Pavese F (2009) About the treatment of systematic effects in metrology. *Measurement* 42: 1459-1462.
16. Pavese F (2020b) Graphic method for retrieval of quantitative data from computer-mapped qualitative information, with a NASA video as an example. *ESIN* 13: 655-662.
17. WMO (2011) The Secretariat: Overview <https://wmo.int/about-wmo/overview>.
18. Lawrimore JH, Rennie J, Gambi de Almeida W, Christy J, Flannery M, et al. (2013) The International Surface Temperature Initiative's Global Land Surface Databank. *AIP Conf. Proc* 1552: 1036-1104.
19. Databasin (2024) <https://databasin.org/maps/new/#datasets=de8f7f71e3334eba863ff6003484364f>.
20. WMO (2018) Guide to Instruments and Methods of Observation. Volume I – Measurement of Meteorological Variables. WMO <https://community.wmo.int/site/knowledge-hub/programmes-and-initiatives/instruments-and-methods-of-observation-programme-imop/guide-instruments-and-methods-of-observation-wmo-no-8>.
21. Farguella A, Moré J, Cortés A, Miró JR, Margalef T, et al. (2016) Reducing Data Uncertainty in Surface Meteorology using Data Assimilation: A Comparison Study. *Procedia Computer Science* 80: 1846-1855.
22. Gillespie I (2021) Investigating the potential use of sparse-input reanalyses to homogenise long-term land surface air temperature records. PhD Thesis submitted to Irish Climate Analysis and Research Units (ICARUS), Department of Geography, Maynooth University, National University of Ireland 1-243.
23. Jeff Id (2011) Global Temperatures and Incomplete Rationale of My Own Skepticism: the Air Vent <https://noconsensus.wordpress.com/2011/04/05/234-5/>.
24. Morice CP, Kennedy JJ, Rayner NA, Winn JP, Hogan E, et al. (2019) An updated assessment of near surface temperature change from 1850: the HadCRUT5 dataset. *J. Geophys. Res. Atmos* 126: 1-28.
25. Lindzen RS and Christy JR (2024) Reassessing the climate change narrative. *Asia-Pacific Journal of Atmospheric Sciences* 60: 319-327.
26. NCEI.NOAA Report (2020) Assessing the Global Climate in 2020. Available at: <https://ncei.noaa.gov/news/global-climate-202012>.

27. Pavese F, Forbes AB, Zhang NF and Chunovkina AG (2021) Series on Advances in Mathematics for Applied Sciences, World Scientific, Singapore 12: 347364.
28. Brohan P, Kennedy JJ, Harris I, Tett SFB and Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. J Geophys. Res 111: 1-21.
29. Datagraver (2024) Digging data. Consulted February <https://datagraver.com/>.
30. Hansen J, Fung I, Lacis A, Rind D, Lebedeff S, et al. (1988) Global climate changes as forecast by Goddard Institute for Space Studies. Three-dimensional model. Journal of Geophysical Research-Atmospheres 93: 9341-9364.
31. Farmer GT and Cook J (2023) Climate Change Science: A Modern Synthesis. Book Volume 1- The Physical Climate: Impossible Expectations 481.
32. Folland CK, Rayner NA, Brown SJ, Smith TM, Shen SSP, et al. (2001) Global Temperature Change and its Uncertainties Since 1861. Geophys. Res. Lett 28: 2621-2624.
33. Hubbard KG, Lin X (2002) Realtime data filtering models for air temperature measurements Geophys. Res. Lett 29: 1-4.
34. ISO (2014) Air quality - Meteorology - Siting classifications for surface observing stations on land 1-3.
35. Frank P (2010) Uncertainty in the global average surface air temperature index: a representative lower limit. Energy & Environment 21: 969-989.
36. NASA GISTEMP (2015) v4 Figures. www.data.giss.nasa.gov. Available at: <https://data.giss.nasa.gov/gistemp/>.
37. Vinòs J (2023) Solving the Climatic Puzzle. Book, Critical Science Press 1-390.
38. Chen L and Frauenfeld OW (2013) Surface Air Temperature Changes over the Twentieth and Twenty-First. J of Climate 27: 3920-3937.
39. Global Climate Highlights (2023) Copernicus, <https://climate.copernicus.eu/GCH2023-Paris-Agreement>.
40. Haustein K, Allen MR, Forster PM, Otto FEL, Mitchell DM, et al. (2017) A real-time Global Warming Index 7: 1-6.
41. IPCC (2022) Sixth Assessment Report of the Intergovernmental Panel on Climate Change. AR6 Climate Change 2022, Mitigation of Climate Change, Technical Summary, Working Group III 102.
42. JCGM-WG2 (2023) VIM-2023-011. International Vocabulary of Metrology, Fourth edition – Second Committee Draft' 31 July 2023. BIPM, Paris 1-77.
43. Mari L, Maul A, Wilson M (2022) Measurement Across the Sciences, Springer Series in Measurement Science and Technology, Chapter 7 Modeling Measurement and Its Quality.
44. Mari L, Maul A, Wilson M (2019) 'Can there be one meaning of "measurement" across the sciences?' Journal of Physics: Conference Series 1379: 213-264.
45. NASA Science Editorial Team (2021) Can scientists use global temperature data as is? <https://science.nasa.gov/climate-change/faq/can-scientists-use-global-temperature-data-as-is/>.
46. NASA Maps (2022) Global Temperature Anomalies from 1880 to 2020 <https://svs.gsfc.nasa.gov/4964>.
47. Pavese F and Ichim D (2007) Design for Measurement Quality and Results Evaluation in the Calibration of Large Batches of Similar Artefacts by Computer-assisted Data Acquisition and Processing (CERN) Proceedings of ENBIS-DEINDE Conference 193-201.
48. Pavese F (2021) Measurement in science: between Evaluation and Prediction. AMCTM XII Eds.
49. Pavese F (2023) 'The choice of cartographic system on the calculation of Earth's surface parameters from maps, namely of the GMST' Earth Science Informatics, METHODOLOGY, 4/2023 ESIN 14, <https://doi.org/10.1007/s12145-023-01061-0>.
50. Peirce CS (1908) Selected writings (Values in a universe of chance). New York: Dover Publications 358-379.
51. WMO Siting Classification (2012) <https://community.wmo.int/en/activity-areas/imop/siting-classification>.