26 June 2023

The Verdict of Instrumental Methods

Pat Frank

*LiG Metrology, Correlated Error, and the Integrity of the Global Surface Air Temperature Record* has passed peer-review and [is now published](https://www.mdpi.com/1424-8220/23/13/5976) in the *MDPI* journal, *Sensors* ([pdf](https://www.mdpi.com/1424-8220/23/13/5976/pdf?version=1687874552)).

The paper complements Anthony's revolutionary [Surface Stations project](https://doi.org/10.1029/2010JD015146), in that the forensic analysis focuses on ideally located and maintained meteorological sensors.

The experience at *Sensors* was wonderfully normal. Submission was matter-of-fact. The manuscript editor did not flee the submission. The reviewers offered constructive criticisms. There was no defense of a favored narrative. There was no dismissive language.

MDPI also has an [admirable approach to controversy](https://www.mdpi.com/about/controversial-articles). The editors, "*ignore the blogosphere*." The contest of ideas occurs in the journal, in full public view, and critical comment must pass peer-review. Three Huzzahs for MDPI.

*LiG Metrology...* (hereinafter *LiG Met.*) returns instrumental methods to the global air temperature record. A start-at-rock-bottom 40 years overdue forensic examination of the liquid-in-glass (LiG) thermometer.

The essay is a bit long and involved. But the take-home message is simple:

1. The people compiling the global air temperature record do not understand thermometers.
2. The rate or magnitude of climate warming since 1900 is unknowable.

Global-scale surface air temperature came into focus with the [1973 Science paper](https://doi.org/10.1038/242310a0) of Starr and Oort, *Five-Year Climatic Trend for the Northern Hemisphere*. By 1983, the [Charney Report](http://www.nap.edu/catalog.php?record_id=12181), *Carbon Dioxide and Climate* was 4 years past, Stephen Schneider [had already weighed in](http://www.jstor.org/stable/3822965) on CO2 and climate danger, Jim Hanson was [publishing on his climate models](https://doi.org/10.1175/1520-0493(1983)111%3c0609:ETDGMF%3e2.0.CO;2), ice core [CO2 was being assessed](https://doi.org/10.1021/j100244a025), and the [trend in surface air temperature](https://doi.org/10.1175/1520-0493(1982)110%3c0059:Visatp%3e2.0.Co;2) came to focused attention.

Air temperature had become central. What was its message?

To find out, the reliability of the surface air temperature record should have been brought to the forefront. But it wasn't. Air temperature measurements were accepted at face value.

Errors and uncertainty were viewed as external to the instrument; a view that persists today.

*LiG Met.* makes up the shortfall, 40 years late, starting with the detection limits of meteorological LiG thermometers.

The paper is long and covers much ground. This will short summary starts with an absolutely critical concept in measurement science and engineering, namely:

*I. Instrumental detection limits*: The detection limit registers the magnitude of physical change (*e.g.,* a change in temperature, ΔT) to which a given instrument (*e.g.,* a thermometer) is able to reliably respond.

Any read-out below the detection limit has no evident physical meaning because the instrument is not reliably sensitive to that scale of perturbation. (The subject is complicated; see [here](https://doi.org/10.1021/ac60259a007) and [here](https://doi.org/10.1016/j.sab.2007.11.015).)

The following Table provides the lower limit of resolution -- the detection limits -- of mercury LiG 1C/division thermometers, as determined at the National Institute of Standards and Technology (NIST).

NIST 1C/division Mercury LiG Thermometer Calibration Resolution Limits (2σ, ±C)

|  |  |  |
| --- | --- | --- |
|  | Eye alone | Magnifying Lens |
| accuracy limit (resolution) | 0.30 | 0.114 |
| visual repeatability | 0.144 | 0.144 |
| Per-measurement uncertaintya | 0.33 | 0.18 |
| Anomaly uncertaintyb | 0.46 | 0.26 |

a. root-sum-square of resolution and visual repeatability. b. Uncertainty in an anomaly is the root-sum-square of the uncertainties in the differenced magnitudes.

These are the laboratory ideal lower limits of uncertainty one should expect in measurements taken by a careful researcher using a good-quality LiG 1C/division thermometer. Measurement uncertainty cannot be less than the lower limit of instrumental response.

The [NASA/GISS air temperature anomaly record](https://doi.org/10.1029/2010RG000345) begins at 1880. However, the largest uncertainties in the modern global air temperature anomaly record are found in the decades 1850-1879 published by [HadCRU/UKMet](https://doi.org/10.1029/2011JD017139) and [Berkeley BEST](https://essd.copernicus.org/articles/12/3469/2020/). The 2σ root-mean-square (RMS) uncertainty of their global anomalies over 1850-1880 is: HadCRU/UKMet = ±0.16 C and Berkeley BEST = ±0.13 C. Graphically:

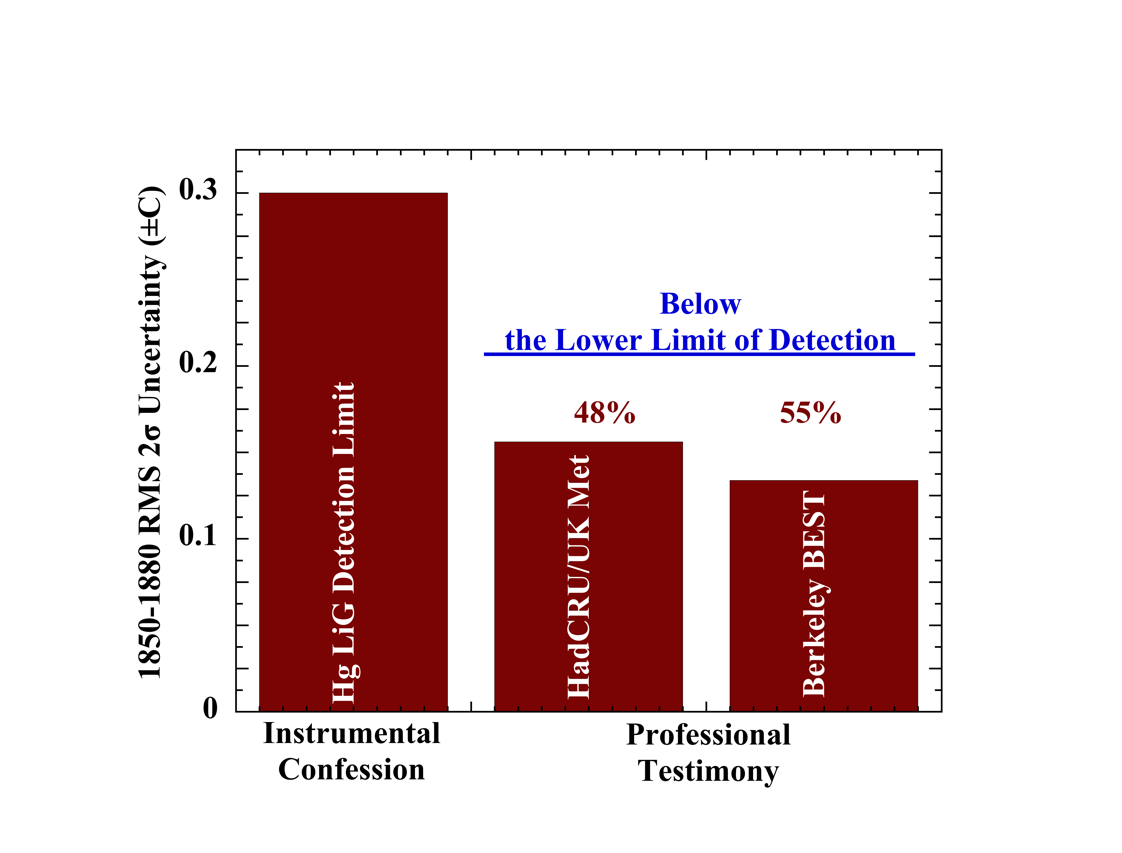


Figure 1: The LiG detection limit and the mean of the uncertainty in the 1850-1880 global air temperature anomalies published by the Hadley Climate Research Unit of the University of East Anglia in collaboration with the UK Meteorological Office (HadCRU/UKMet) and by the Berkeley Earth Surface Temperature project (Berkeley BEST).

That is, the published uncertainties are about half the instrumental lower limit of detection -- a physical impossibility.

The impossibility only increases with the decrease of later uncertainties (Figure 6, below). This strangeness shows the problem that ramifies through the entire field: neglect of basics.

Summarizing (full details and graphical demonstrations in *LiG Met.*):

*Non-linearity:* Both mercury and especially ethanol (spirit) expand non-linearly with temperature. The resulting error is small for mercury LiG thermometers, but significant for the alcohol variety. In the standard surface station prior to 1980, [an alcohol thermometer provided *Tmin*](https://library.wmo.int/index.php?id=12407&lvl=notice_display), which puts 2σ = ±0.37 C of uncertainty into every daily land-surface *Tmean*. Temperature error due to non-linearity of response is uncorrected in the historical record.

*Joule-drift*: Significant bulb contraction occurs with aging of thermometers manufactured before 1885, and is most rapid in those made with lead-glass. Joule-drift puts a spurious 0.3-0.7 C/century warming trend into a temperature record. Figure 4 in *LiG Met.* presents the Pb X-ray fluorescence spectrum of a 1900-vintage spirit meteorological thermometer purchased by the US Weather Bureau. Impossible-to-correct error from Joule drift makes the entire air temperature record prior to 1900 unreliable.

*The Resolution message:* All of these sources of error and uncertainty -- detection limits, non-linearity, and Joule drift -- are inherent to the LiG thermometer and should have been evaluated right at the start. Well before making any serious attempt to construct a record of historical global surface air temperature. However, they were not. They were roundly neglected. Perhaps most shocking is professional neglect of instrumental detection limit.

Figure 2 shows the impact of the detection limit alone on the 1900-1980 global air temperature anomaly record.

Land surface temperature means include the uncorrected error from non-linearity of spirit thermometers. Sea surface temperatures (SSTs) were measured with mercury LiG thermometers only (no spirit LiG error). The resolution uncertainty for the global air temperature record prior to 1981 was calculated as,

*2σT = 1.96 × sqrt[0.7 × (SST resolution)2 + 0.3 × (LST resolution)2]*

*= 1.96 × sqrt[0.7 × (0.136)2 + 0.3 × (0.195)2] = ±0.306 C*, where *LS* is Land-Surface.

But global air temperature change is reported as an anomaly series relative to a 30-year normal. Differencing two values requires adding their uncertainties in quadrature. The resolution of a LiG-based 30-year global temperature normal is also 2σ = ±0.306 C. The resolution uncertainty in a LiG-based global temperature anomaly series is then

*2σ = 1.96 × sqrt[(0.156)2 + (0.156)2] = ±0.432 C*

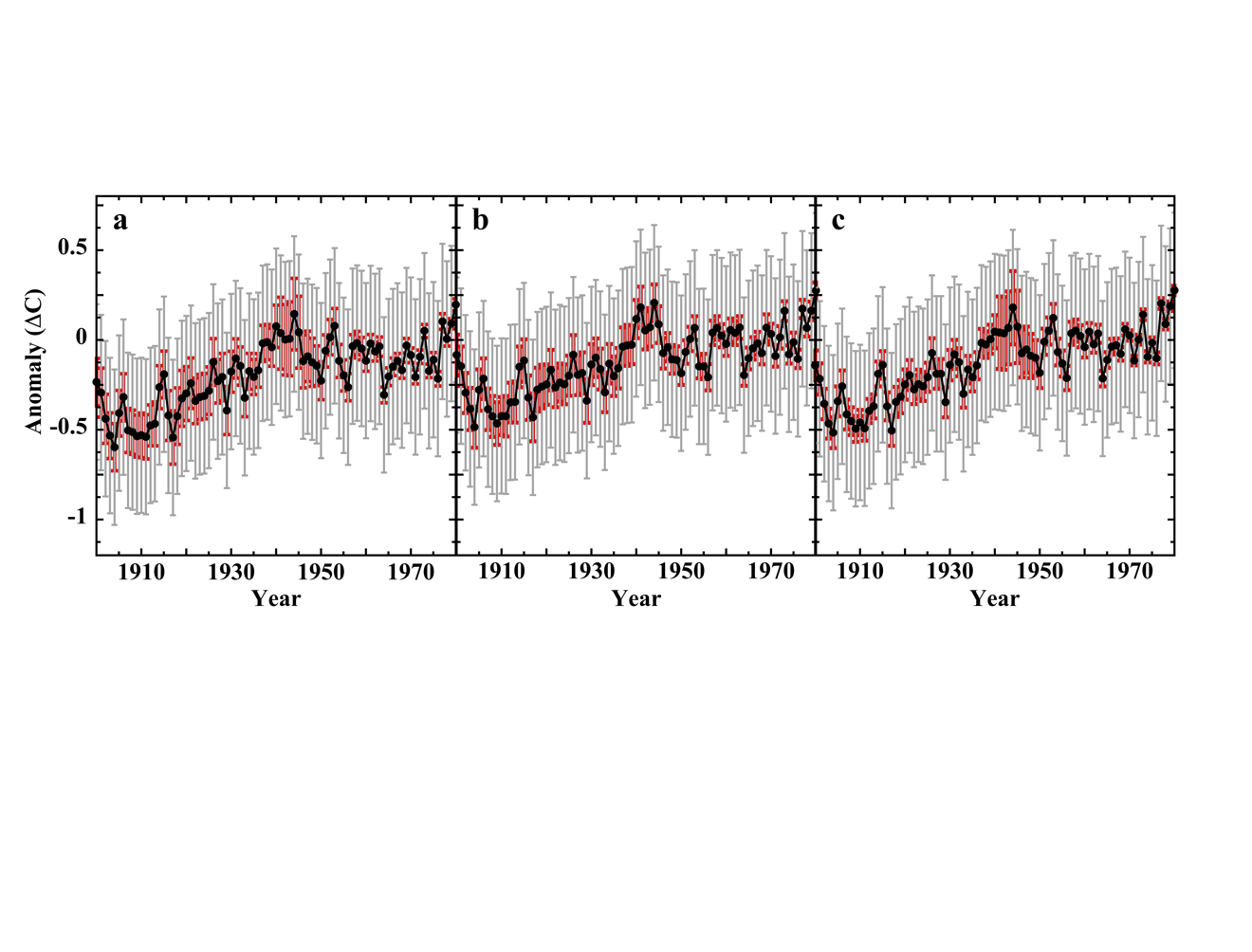


Figure 2: (Points), 1900 - 1980 global air temperature anomalies for: panel a, HadCRUT5.0.1.0 (published through 2022); Panel b, GISSTEMP v4 (published through 2018), and; Panel c, Berkeley Earth (published through 2022). Red whiskers: the published 2σ uncertainties. Grey whiskers: the uniform 2 σ = ±0.432 C uncertainty representing the laboratory lower limit of instrumental resolution for a global average annual anomaly series prior to 1981.

In Figure 2, the mean of the published anomaly uncertainties ranges from 3.9× smaller than the LiG resolution limit at 1900, to 5× smaller at 1950, and nearly 12× smaller at 1980.

*II. Systematic error enters into global uncertainty. Is temperature measurement error random?*

Much of the paper tests the assumption of random measurement error; an assumption absolutely universal in global warming studies.

*LiG Met. Section 3.4.3.2* shows that differencing two normally distributed data sets produces another normal distribution. This is an important realization. If measurement error is random, then differing two sets of simultaneous measurements should produce a normally distributed error difference set.

*II.1 Land surface systematic air temperature measurement error is correlated*: Systematic temperature sensor measurement calibration error of proximately located sensors turns out to be pair-wise correlated.

Matthias Mauder and his colleagues [published a study](https://doi.org/10.1175/2008JTECHA1046.1) of the errors produced within 25 naturally ventilated HOBO sensors (gill-type shield, thermistor sensor), relative to an aspirated Met-One precision thermistor standard. Figure 3 shows one pair-wise correlation of the 25 in that experimental set, with correlation r = 0.98.

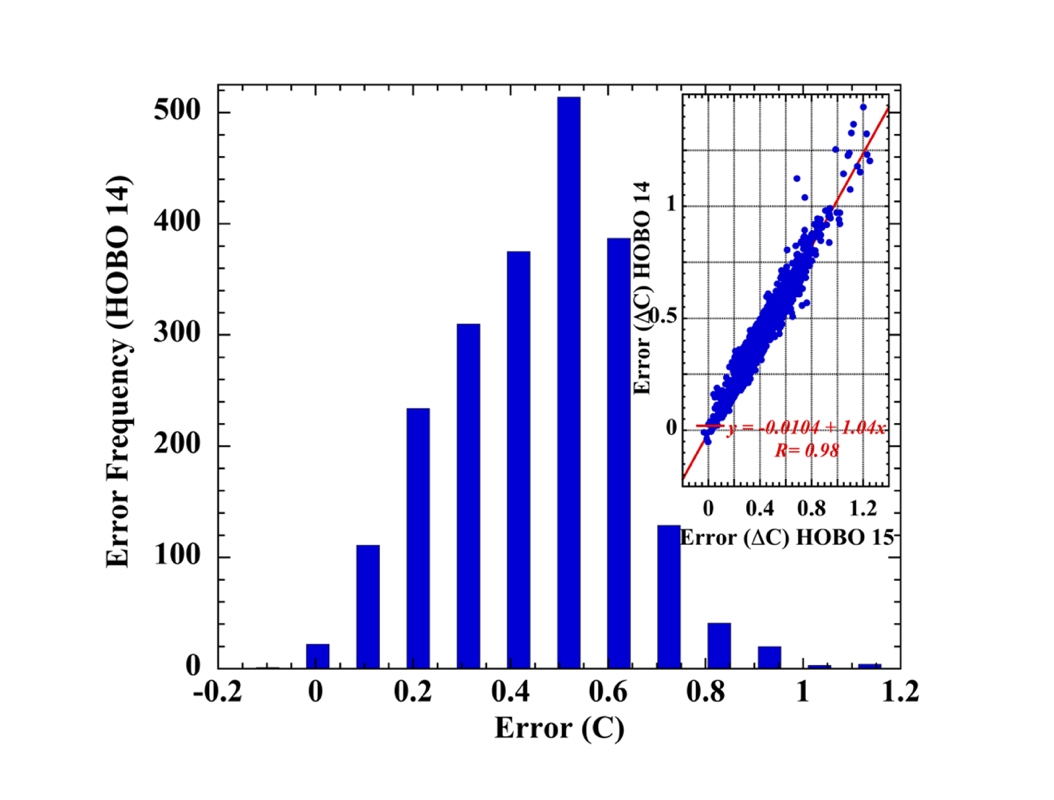


Figure 3: Histogram of error in HOBO number 14 (of 25). The StatsKingdom online [Shapiro-Wilk normality test](https://www.statskingdom.com/shapiro-wilk-test-calculator.html) (2160 error data points) yielded: 0.979, p < 0.001, statistically non-normal. Inset: correlation plot of measurement error -- HOBO #14 versus HOBO #15; correlation r = 0.98.

High pair-wise correlations were found between all 25 HOBO sensor measurement error data sets. The Shapiro-Wilk test, [has the greatest statistical power](https://doi.org/10.1080/00949655.2010.520163) to indicate or reject the normality of a data distribution and showed that every single measurement error set was non-normal.

*LiG Met.* and the Supporting Information provide multiple examples of independent field calibration experiments, that produced pair-wise correlated systematic sensor measurement errors. Shapiro-Wilk statistical tests of calibration error data sets invariably indicated non-normality.

Inter-sensor correlation in land-surface systematic measurement field calibration error, along with non-normal distributions of difference error data sets, together falsify the general assumption of strictly random error. No basis in evidence remains to permit diminishing uncertainty as 1/√N.

*II.2.1 Sea-Surface Temperature measurement error is not random*: Differencing simultaneous bucket-bucket and bucket-engine-intake measurements again yields the measurement error difference, *Δε2,1*. If measurement error is random, a large SST difference data set, *Δε2,1,* should have a normal distribution*.*

Figure 4 shows the result of a World Meteorological Organization project, [published in 1972](https://library.wmo.int/index.php?lvl=notice_display&id=5811), which reported differences of 13511 simultaneously acquired bucket and engine-intake SSTs from all manner of ships, at low and high N,S latitudes and under a wide range of wind and weather conditions. The required normal distribution is nowhere in evidence.

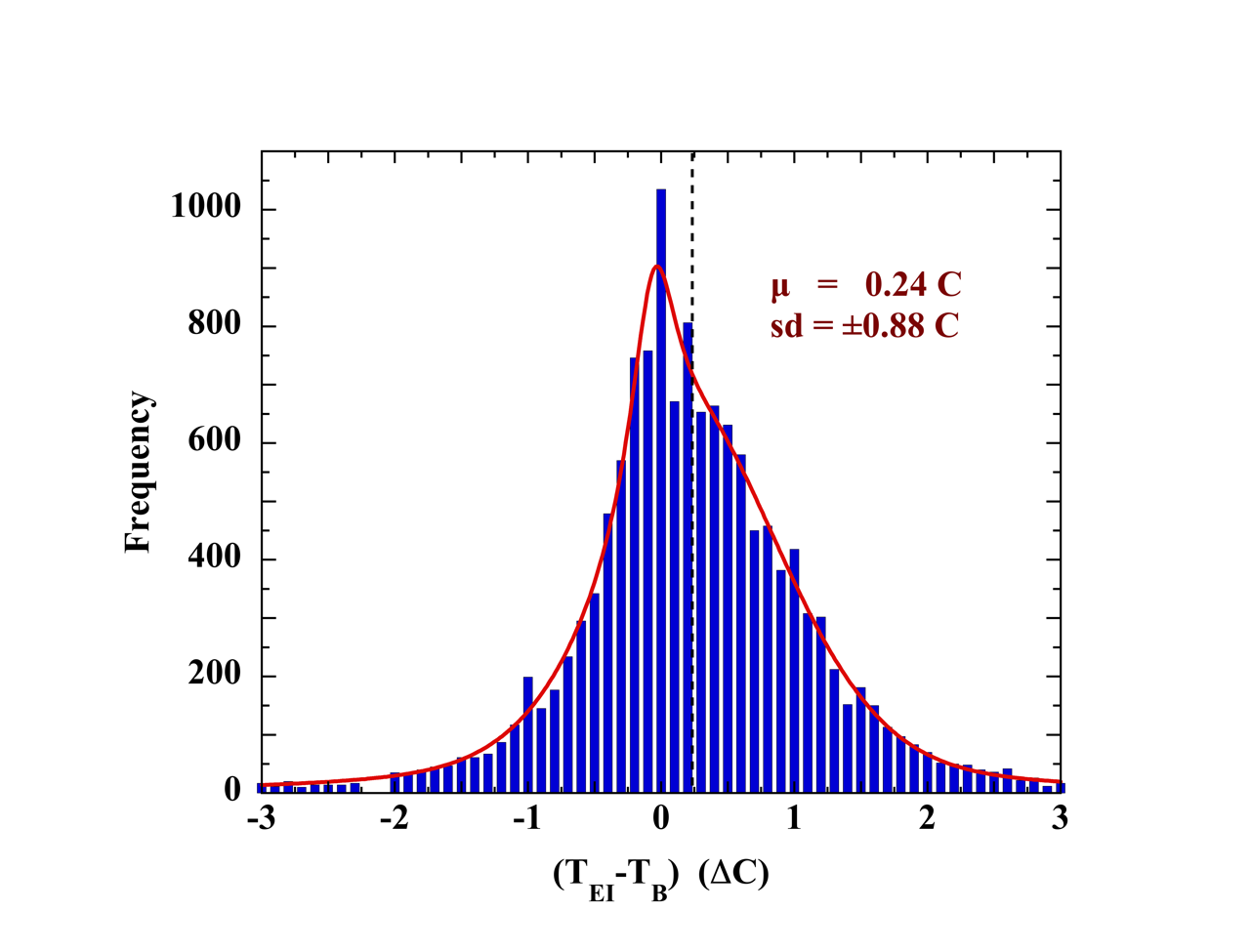


Figure 4: Histogram of differences of 13511 simultaneous engine-intake and bucket SST measurements during a large-scale experiment carried out under the auspices of the World Meteorological Organization. The red line is a fit using two Lorentzians and a Gaussian. The dashed line marks the measurement mean.

*LiG Met.* presents multiple independent large-scale bucket/engine-intake difference data sets of simultaneously measured SSTs. The distributions were invariably non-normal, demonstrating that SST measurement errors are not random.

*II.2.2 The SST measurement error mean is unknown*: The semivariogram method, taken from [Geostatistics](https://doi.org/10.1016/B0-12-227410-5/00869-3), has been used to [derive the shipboard SST error mean](https://journals.ametsoc.org/view/journals/atot/16/7/1520-0426_1999_016_0905_asdotr_2_0_co_2.xml), *±εmean*. The assumption again is that SST measurement error is strictly random, but with a mean offset.

Subtract *εmean*, and get a normal distribution of with a mean of zero, and an uncertainty diminishing as 1/√N.

However, *LiG Met. Section 3.4.2* shows that the semivariogram analysis doesn't produce *±εmean*, but instead *±0.5Δεmean*, half themean of the error *difference*. Subtraction does not leave a mean of zero.

Conclusion about SST: *II.2.1* shows measurement error is not strictly random. *II.2.2* shows ignorance of the error mean. No grounds remain to diminish SST uncertainty as 1/√N.

*II.2.3 The SST is unknown*: In 1964 (*LiG Met. Section 3.4.4*) [Robert Stevenson](https://scripps.ucsd.edu/news/obituary-notice-father-space-oceanography-robert-e-stevenson) carried out an [extended SST calibration experiment](https://doi.org/10.1175/1520-0450(1964)003%3c0115:TIOASO%3e2.0.CO;2) aboard the VELERO IV oceanographic research vessel. Simultaneous high-accuracy SST measurements were taken from the VELERO IV and from a small launch put out from the ship.

Stevenson found that the ship so disturbed the surrounding waters that the SSTs measured from the ship were not representative of the physically true water temperature (or air temperature). No matter how accurate, the bucket, engine-intake, or hull-mounted probe temperature measurement did not reveal the true SST.

The only exception was an SST obtained using a prow-mounted probe, but *iff* the measurement was made when the ship was heading into the wind "*or cruising downwind at a speed greater than the wind velocity*."

Stevenson concluded, “*One may then question the value of temperatures taken aboard a ship, or from any large structure at sea. Because the measurements vary with the wind velocity and the orientation of the ship with respect to the wind direction no factor can be applied to correct the data. It is likely that the temperatures are, therefore, useless for any but gross analyses of climatic factors, excepting, perhaps, those taken with a carefully-oriented probe.*”

Stevenson's experiment may be the most important investigation ever carried out of the veracity of ship-derived SSTs. However, the experiment generated scant notice. It was never repeated or extended, and the reliability question of SSTs the VELERO IV experiment revealed has generally been by-passed. The journal shows only 5 citations since 1964.

Nevertheless, ship SSTs [have been used to calibrate satellite SSTs](https://doi.org/10.1175/1520-0477(2001)082%3c2773:ESSTFI%3e2.3.CO;2) probably [through 2006](https://doi.org/10.1175/JTECH1828.1). Which means that earlier satellite SSTS are not independent of the large uncertainty in ship SSTs.

*III. Uncertainty in the global air temperature anomaly trend*: We now know that the assumption of strictly random measurement error in LSTs or SSTs is unjustified. Uncertainty cannot be presumed to diminish as 1/√N.

*III.1 For land-surface temperature,* uncertainty was calculated from:

* LiG resolution (detection limits, visual repeatability, and non-linearity).
* systematic error from unventilated CRS screens (pre-1981).
* interpolation from CRS to MMTS (1981-1989).
* unventilated Min-Max Temperature System (MMTS) sensors (1990-2004).
* Climate Research Network (CRN) sensors self-heating error (2005-2010).

Over 1900-1980, resolution uncertainty was combined in quadrature with the uncertainty from systematic field measurement error, yielding a total RMS uncertainty 2σ = ±0.57 C in LST.

*III.2 For sea-surface temperature* uncertainty was calculated from Hg LiG resolution combined with the systematic uncertainty means of bucket, engine intake and bathythermograph measurements scaled by [their annual fractional contribution](https://doi.org/10.1002/wcc.55) since 1900.

SST uncertainty varied due to the annual change in fractions of bucket, engine-intake and bathythermograph measurements. Engine intake errors dominated.

Over 1900-2010 uncertainty in SST was RMS 2σ = ±1.38 C.

*III.3 Global:* Annual uncertainties in land surface and sea surface again were combined as:

*2σT = 1.96 × sqrt[0.7 × (SST uncertainty)2 + 0.3 × (LST uncertainty)2]*

Over 1900-2010 the RMS uncertainty in global air temperature was found to be, 2σ = ±1.22 C.

The uncertainty in an anomaly series is the uncertainty in the air temperature annual (or monthly) mean combined in quadrature with the uncertainty in the selected 30-year normal period.

The RMS 2σ uncertainty in the NASA/GISS 1951-1980 normal is ±1.48 C, and is ±1.49 C in the HadCRU/UEA and Berkeley BEST 1961-1990 normal.

The 1900-2010 mean global air temperature anomaly is 0.94 C. Using the NASA/GISS normal, the overall uncertainty in the 1900-2010 anomaly is,

*2σ = 1.96 × sqrt[(0.755)2 + (0.622)2] = ±1.92 C*

The complete change in air temperature between 1900-2010 is then 0.94±1.92 C.

Figure 5 shows the result applied to the annual anomaly series. The red whiskers are the 2σ quadratic annual combined RMS of the three major published uncertainties (HadCRU/UEA, NASA/GISS and Berkeley Earth). The grey whiskers include the combined LST and SST systematic measurement uncertainties. LiG resolution is included only through 1980.

The lengthened growing season, the revegetation of the far North, and the poleward migration of the northern tree line provide evidence of a warming climate. However, the rate or magnitude of warming since 1850 is not knowable.

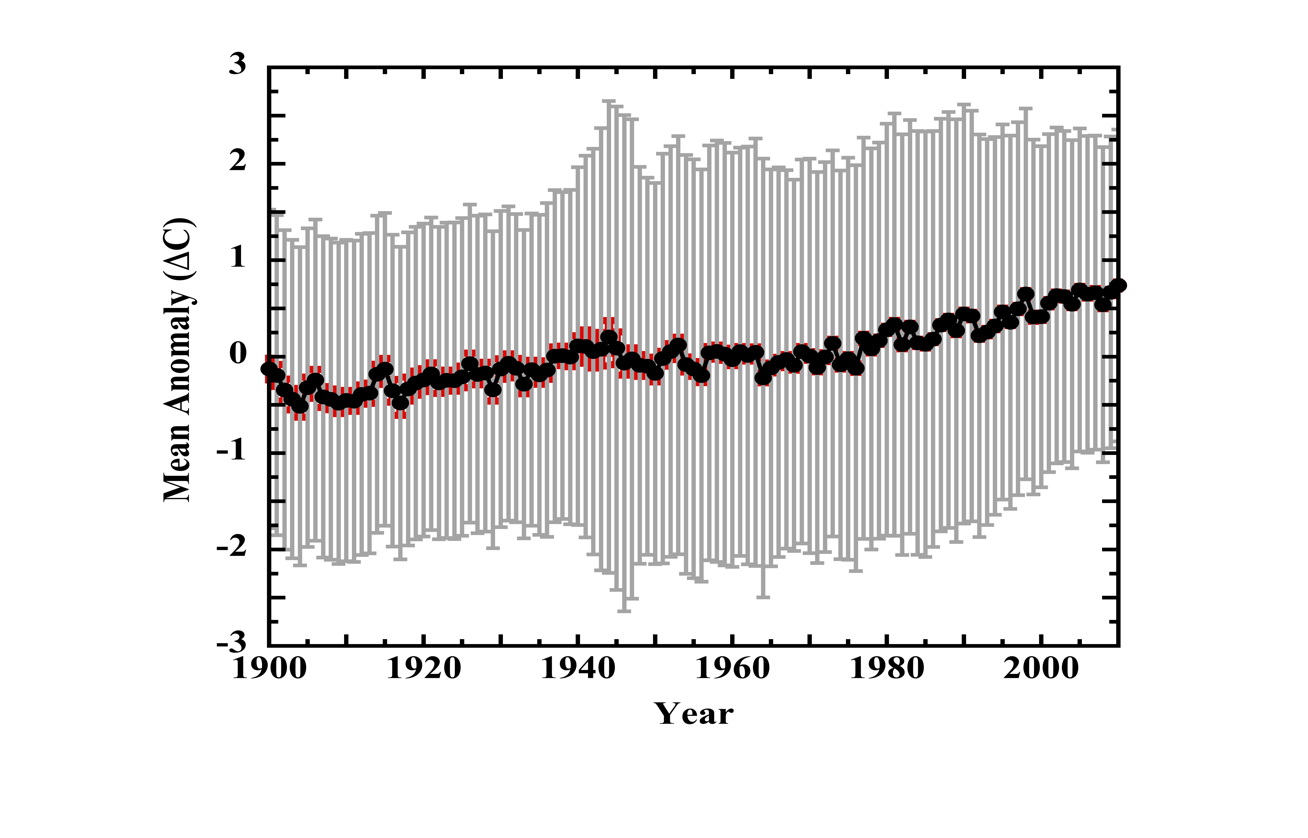


Figure 5: (Points), mean of the three sets of air temperature anomalies published by the UK Met Office Hadley Centre/Climatic Research Unit, the Goddard Institute for Space Studies, or Berkeley Earth. Each anomaly series was adjusted to a uniform 1951-1980 normal prior to averaging. (Red whiskers), the 2σ RMS of the published uncertainties of the three anomaly records. (Grey whiskers), the 2σ uncertainty calculated as the lower limit of LiG resolution (through 1980) and the mean systematic error, combined in quadrature. In the anomaly series, the annual uncertainty in air temperature was combined in quadrature with the uncertainty in the 1951-1980 normal. The increased uncertainty after 1945 marks the wholesale incorporation of ship engine-intake thermometer SST measurements (2σ = ±2 C). The air temperature anomaly series is completely obscured by the uncovered uncertainty bounds.

*IV. The 60-fold Delusion*: Figure 6 displays the ratio of uncovered and published uncertainties, illustrating the extreme of false precision in the official global air temperature anomaly record.

Panel a is (LiG ideal laboratory resolution) ÷ Published. Panel b is total (resolution plus systematic) ÷ Published.

Panel a covers 1850-1980, when the record is dominated by LiG thermometers alone. The LiG lower limit of detection is a hard physical bound.

Nevertheless, the published uncertainty is immediately (1850) about half the lower limit of detection. As the published uncertainties get ever smaller traversing the 20th century, they get ever more unphysical; ending in 1980 at nearly 12× smaller than the LiG physical lower limit of detection.

Panel b covers the 1900-2010 modern period. Joule-drift is mostly absent, and the record transitions into MMTS thermistors (1981) and CRN aspirated PRTs (post-2004). The comparison for this period includes contributions from both instrumental resolution and systematic error.

The uncertainty ratio now maxes out in 1990, with the published version about 60× smaller than the combined instrumental resolution plus field measurement error. By 2010, the ratio declines to about 40× because ship engine-intake measurements make an increasingly small contribution after 1990 ([Kent, et al., (2010)](https://doi.org/10.1002/wcc.55)).

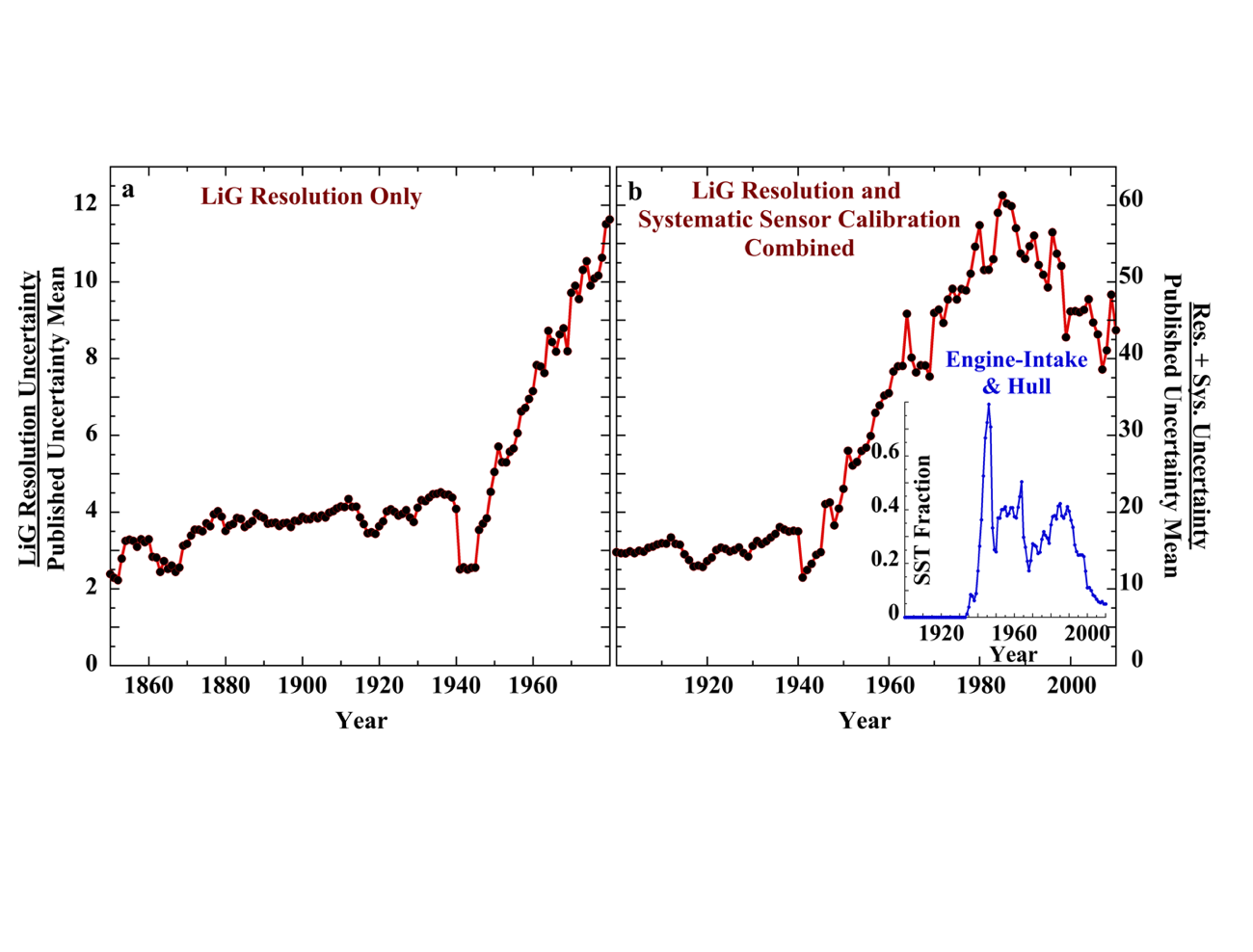


Figure 6: Panel a. (points), the ratio of the annual LiG resolution uncertainties divided by the RMS mean of the published uncertainties (2σ, 1850-1980). Panel b. (points), the ratio of the annual total measurement uncertainties divided by the RMS mean of the published uncertainties (2σ, 1900-2010). Inset: the fraction of SSTs obtained from engine-intake thermometers and hull-mounted probes (a minority). The drop-off of E-I temperatures in the historical record after 1990 accounts for the declining uncertainty ratio.

*V. The verdict of instrumental methods*:

Inventory of error and uncertainty in the published air temperature record:

[NASA/GISS](https://doi.org/10.1029/2010RG000345): incomplete spatial coverage, urban heat islands, station moves.

[Hadley Centre/UEA Climate Research Unit](https://doi.org/10.1029/2005JD006548): random measurement error, instrumental or station moves, changes in instrument type or time-of-reading, sparse station data, urban heat island, bias due to changes in sensor exposure (screen type), bias due to changes in methodology of SST measurements.

[Berkeley Earth](https://doi.org/10.4172/2327-4581.1000103): non-climate related noise, incomplete spatial coverage, and limited efficacy of their statistical model.

No mention by anyone of anything concerning instrumental methods of analysis, in a field completely dominated by instruments and measurement.

Instead, one encounters an analysis conveying no attention to instrumental limits of accuracy or to the consequences attending their technical development, or of their operational behavior. This, in a field where knowledge of such things is a pre-requisite.

Those composing the air temperature record display no knowledge of thermometers. Perhaps the ultimate irony.

No appraisals of LiG thermometers as instruments of measurement despite their preponderance in the historical temperature record. Nothing of the very relevant history of their technical evolution, of their reliability or their resolution or detection limits.

Nothing of the known systematic field measurement errors that affect both LiG thermometers and their successor temperature sensors.

One might expect those lacunae from mathematically adept science dilettantes, who cruise shallow numerical surface waters while blithely unaware of the instrumental depths below; never coming to grips with the fundamentals of study. But not from professionals.

We already knew that climate models [cannot support any notion of a torrid future](https://www.frontiersin.org/articles/10.3389/feart.2019.00223/full). Also, [here](https://doi.org/10.1080/02626667.2010.513518) and [here](https://doi.org/10.1623/hysj.53.4.671). We also know that [climate modelers do not understand physical error analysis](https://doi.org/10.13140/RG.2.2.34218.70083). Predictive reliability: a mere bagatelle of modern modeling?

Now we know that the air temperature record cannot support any message of unprecedented warming. Indeed, almost no message of warming at all.

And we also now know that compilers of the air temperature record evidence no understanding of thermometers, incredible as that may seem. Instrumental methods: a mere bagatelle of modern temperature measurement?

The climate effects of our CO2 emissions, if any, are invisible. The rate or magnitude of the 20th century change in air temperature is unknowable.

With this study, nothing remains of the IPCC paradigm. Nothing. It is empty of content. It always was so, but this truth was hidden under the collaborative efforts of administrative embrace, partisan shouters, character assassins, media propagandists, and professional abeyance.

All those psychologists and sociologists who [published their profoundly learned insights](https://www.psychologicalscience.org/observer/the-subterranean-war-on-science?utm_content=buffer52b77&utm_source=buffer&utm_medium=twitter&utm_campaign=Buffer&utm_content=buffer52b77&utm_source=buffer&utm_medium=twitter&utm_campaign=Buffer) into the [delusional minds](https://doi.org/10.1192/bjb.2021.7), [psychological barriers](https://doi.org/10.1037/a0023566), and [inadequate personalities](https://www.sciencedirect.com/science/article/abs/pii/S0191886915003827) plaguing [their notion of climate/science deniers](https://journals.sagepub.com/doi/abs/10.1177/0956797612457686) are left with egg on their faces or in their academic beards. In their professional acuity, they inverted both the order and the perceivers of delusion and reality.

We're faced once again with the enormity of contemplating a science that has collapsed into a partisan narrative; partisans hostile to ethical practice.

And the professional societies charged with embodying physical science, with upholding ethics and method -- [the National Academies](https://www.nationalacademies.org/topics/climate), the [American Physical Society](https://www.aps.org/policy/reports/popa-reports/energy/climate.cfm), the [American Institute of Physics](https://history.aip.org/climate/index.htm), the [American Chemical Society](https://www.acs.org/policy/publicpolicies/sustainability/globalclimatechange.html) -- collude in [the offense](https://encyclopedia.pub/entry/32380). Their negligence is beyond shame.