

Climate Signal Detection Times and Constraints on Climate Benchmark Accuracy Requirements

STEPHEN S. LEROY AND JAMES G. ANDERSON

Harvard School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts

GEORGE OHRING

NOAA/National Environmental Satellite, Data, and Information Service, Camp Springs, Maryland

(Manuscript received 12 March 2007, in final form 3 July 2007)

ABSTRACT

Long-term trends in the climate system are always partly obscured by naturally occurring interannual variability. All else being equal, the larger the natural variability, the less precisely one can estimate a trend in a time series of data. Measurement uncertainty, though, also obscures long-term trends. The way in which measurement uncertainty and natural interannual variability interact in inhibiting the detection of climate trends using simple linear regression is derived and the manner in which the interaction between the two can be used to formulate accuracy requirements for satellite *climate benchmark* missions is shown. It is found that measurement uncertainty increases detection times, but only when considered in direct proportion to natural variability. It is also found that detection times depend critically on the correlation time of natural variability and satellite lifetime. As a consequence, requirements on satellite climate benchmark accuracy and mission lifetime must be directly related to the natural variability of the climate system and its associated correlation times.

1. Introduction

The U.S. National Research Council, in its decadal survey of the National Oceanic and Atmospheric Administration (NOAA) and the National Aeronautics and Space Administration (NASA) (National Research Council, Committee on Earth Science and Applications from Space 2007), has called for the following new philosophy in monitoring climate change from space:

Design of climate observing and monitoring systems from space must ensure the establishment of global, long-term climate records, which are of *high accuracy, tested for systematic errors on-orbit, and tied to irrefutable standards* such as those maintained in the U.S. by the National Institute of Standards and Technology. For societal objectives that require long-term cli-

mate records, the accuracy of core benchmark observations must be verified against absolute standards *on-orbit* by fundamentally independent methods, such that the accuracy of the record archived today can be verified by future generations. Societal objectives also require a long-term record not susceptible to compromise by interruptions in that data record.

In this note we call observations that satisfy these demands *climate benchmarks*.

Climate benchmarks mark a departure from the current paradigm of climate monitoring in which space instruments are assumed to be “stable.” In the stability paradigm, the accuracy of an instrument is assumed to be unknown yet unchanging. A climate data record is formed, then, by adding offsets to each in a series of satellite instruments so that there is no difference between satellites’ measurements during periods of overlap (National Research Council, Committee on Climate Data Records from NOAA Operational Satellites 2004). The result is a time series of measurements with no obvious discontinuities. Prime examples of this para-

Corresponding author address: Stephen Leroy, Anderson Group, 12 Oxford St., Link Building, Cambridge, MA 02138.
E-mail: leroy@huarp.harvard.edu

digm are the records of upper-air temperature established using the Microwave Soundings Units (MSUs) aboard the NOAA series satellites (Spencer and Christy 1990). Some evidence is given to support the stability of the MSU instruments (Spencer and Christy 1993), but complications in bias adjustment during periods of instrument temporal overlap—a procedure that must be undertaken because the instrument accuracy is unknown—provide the ultimate uncertainty in the climate record established by MSU (Climate Change Science Program 2006).

The paradigm of climate monitoring using instruments traceable to international measurement standards (National Research Council, Committee on Earth Science and Applications from Space 2007; Ohring 2007) calls for a different analysis technique. After calibration, every measurement obtained is known to be accurate to within an uncertainty determined by the multiple calibration pathways to the international system of units on board the spacecraft. No statement can or need be made regarding whether the uncertainty of the measurements is changing or unchanging with time. Conservative error analysis requires that the worst-case scenario for error propagation be considered the relevant one: if changing uncertainty leads to greater error in the result, then changing uncertainty must be assumed; if unchanging uncertainty leads to greater error in the result, then unchanging uncertainty must be assumed. Ordinary linear regression is the obvious analysis method for climate benchmark data. With a series of climate benchmarks, one obtains a time series of data, with or without gaps in time, wherein each data point is accurate to within an uncertainty established by its traceable pathways to international standards. In the problem of detecting slowly evolving trends in the climate system, one must also consider the shorter time-scale natural fluctuations of the climate as an additional source of uncertainty. Thus, the uncertainty associated with each point in a time series of data is estimated as the standard deviation of the data points from the best-fit line, and both measurement uncertainty and natural variability contribute to that standard deviation. In considering their contributions through a proper error analysis, one arrives at minimum signal detection times and requirements for instrument accuracy that are dictated by nature. In this paper, we present such an error analysis.

2. Derivation by linear regression

To arrive at an equation that can be used for calculating a requirement for climate benchmark accuracy,

we start with standard linear regression (Williams 1959; von Storch and Zwiers 1999). Ultimately, a climate benchmark time series will be used to determine whether there are trends in the climate system. We give the solution for the trend m in an N -element time series of data d_i at times t_i :

$$m = \left[\sum_{j=1}^N (t_j - \bar{t})^2 \right]^{-1} \sum_{i=1}^N d_i (t_i - \bar{t}), \quad (1)$$

where \bar{t} is the mean of the times t_i . A determination of the slope is inevitably corrupted by natural variability of the climate system, which adds scatter to the data away from any fitted line. A determination of the slope is also corrupted by measurement uncertainty, which adds scatter to the data, too. In data analysis one computes the uncertainty in the estimate of the trend using the data's residuals. In deriving accuracy requirements for a climate benchmark instrument, we estimate the uncertainty of a yet-to-be-measured trend m through conventional error propagation techniques. A deviation δm to the slope estimate is caused by uncertainty δd_i in the data:

$$\delta m = \left[\sum_{j=1}^N (t_j - \bar{t})^2 \right]^{-1} \sum_{i=1}^N \delta d_i (t_i - \bar{t}). \quad (2)$$

The mean-square estimate of the uncertainty in the slope $\langle (\delta m)^2 \rangle$ is

$$\langle (\delta m)^2 \rangle = \left[\sum_{k=1}^N (t_k - \bar{t})^2 \right]^{-2} \sum_{i=1}^N \sum_{j=1}^N (t_i - \bar{t})(t_j - \bar{t}) \langle \delta d_i \delta d_j \rangle. \quad (3)$$

Typically in time series analysis, the uncertainty in the data is understood to be completely uncorrelated, namely, that $\langle \delta d_i \delta d_j \rangle = \sigma^2 \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta function. In climate signal detection this does not hold because there is serial correlation in the time series (von Storch and Zwiers 1999).

The fluctuations of the climate system are not white noise: they have associated length and time scales. A temporal anomaly of the climate system is bound to last a finite amount of time, and if that amount of time is comparable to or greater than the interval between data points $dt = t_{i+1} - t_i$, then the expectation value $\langle \delta d_i \delta d_j \rangle$ takes the following form:

$$\langle \delta d_i \delta d_j \rangle = \sigma_{\text{var}}^2 \text{Corr}_{i-j}(\text{var}), \quad (4)$$

where $\text{Corr}_n(\text{var})$ is the correlation function of natural variability at lag $n dt$ in time and σ_{var}^2 is the zero-lag

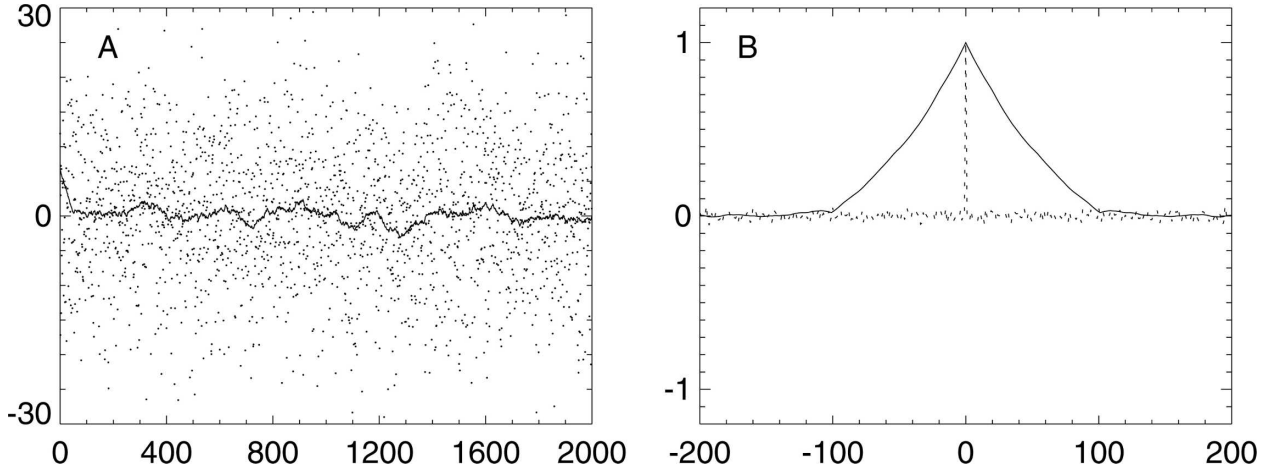


FIG. 1. The time constant τ_{var} of a smoothed random process. (a) A time series of serially uncorrelated, normally distributed random data with std dev of 10 (points) and the smoothed data after applying a boxcar smoothing with a width of 100 (solid curve). (b) The autocorrelations as a function of lag for the data points (dashed line) and for the smoothed data (solid line). The area under the solid line is ≈ 100 . By definition, the area under the solid curve in (b) is τ_{var} . Theoretically, the constant τ_{var} of Eq. (7) is estimated to be the width of the boxcar filter.

variance associated with natural variability. (By natural variability, we mean all the naturally occurring variations internal to the climate thought to not be associated with the response to a prescribed forcing.) A similar argument holds for measurement uncertainty. It is very likely that an uncertainty in measurements of a climate benchmark instrument lasts a finite amount of time as well, and therefore errors at near-adjacent instances in time may be correlated. This appears as an extra term in the expectation value $\langle \delta d_i \delta d_j \rangle$:

$$\langle \delta d_i \delta d_j \rangle = \sigma_{\text{var}}^2 \text{Corr}_{i-j}(\text{var}) + \sigma_{\text{meas}}^2 \text{Corr}_{i-j}(\text{meas}), \quad (5)$$

where σ_{meas} is the measurement uncertainty and $\text{Corr}_\mu(\text{meas})$ is its time-lagged correlation function.

We simplify the summations in Eq. (3) after inserting Eq. (5), setting $j = i + \mu$, and summing over μ . Except in the immediate vicinity of the first and last elements of the time series, the uncertainty in the estimate of the trend becomes

$$\langle (\delta m)^2 \rangle = \left[\sum_{k=1}^N (t_k - \bar{t})^2 \right]^{-1} \left[\sigma_{\text{var}}^2 \sum_{\mu=-\infty}^{\infty} \text{Corr}_\mu(\text{var}) + \sigma_{\text{meas}}^2 \sum_{\mu=-\infty}^{\infty} \text{Corr}_\mu(\text{meas}) \right]. \quad (6)$$

The summations are exactly the normalized Fourier transforms of the natural variability and of the measurement uncertainty time-lagged covariance functions at zero frequency, and so they can be rewritten in terms of correlation times for the natural variability τ_{var} and for the measurement uncertainty τ_{meas} :

$$\tau_{\text{var}} \equiv dt \sum_{\mu=-\infty}^{\infty} \text{Corr}_\mu(\text{var}), \quad (7)$$

$$\tau_{\text{meas}} \equiv dt \sum_{\mu=-\infty}^{\infty} \text{Corr}_\mu(\text{meas}). \quad (8)$$

While infinite lags in these correlation functions are not practically realizable, we anticipate that the only significant lags will be no greater than a few years for decadal-scale trend detection. To illustrate why Eqs. (7) and (8) are meaningful time constants, consider the case of smoothing of a serially uncorrelated random process. A serially uncorrelated random process has an autocorrelation of 1 at zero lag and an autocorrelation of 0 otherwise. If it is smoothed with a boxcar filter with width T , the autocorrelation function becomes triangular. It is 1 at zero lag and linearly drops to 0 at $+T$ and $-T$ lag. The summation of that correlation function given by Eq. (7), or the integral of the autocorrelation function in lag time, is exactly $\tau_{\text{var}} = T$. See Fig. 1 for an illustration.

We note for an unbroken but discretized time series t_i that

$$\sum_{i=1}^N (t_i - \bar{t})^2 = (dt)^2 (N^3 - N)/12. \quad (9)$$

If we assume a long time series ($N \gg 1$), then the uncertainty in the determination of the trend in the data reduces to

$$\langle (\delta m)^2 \rangle \approx 12(\Delta t)^{-3} (\sigma_{\text{var}}^2 \tau_{\text{var}} + \sigma_{\text{meas}}^2 \tau_{\text{meas}}), \quad (10)$$

where $\Delta t = N dt$ is the length of the time series. Equation (10) is useful both for deriving science requirements for climate benchmark missions and for estimating the signal-to-noise ratio in detecting climate signals.

Equation (10) bears a simple interpretation. It is obvious that the longer the time series, the easier it should be to distinguish a trend from natural variability (and measurement uncertainty). We call this the baseline effect. If we have just two data separated by Δt in time, the uncertainty in the trend determination is just the uncertainty in the measurements divided by the baseline Δt . The mathematical expression for this is $|\delta m| \approx \sigma_{\text{var}}/\Delta t$, wherein the error in the determination of the slope is inversely proportional to the baseline of the time series. Advantage is gained, though, from the fact that a continuous time series offers the possibility of averaging out some of the natural variability. We call this the averaging effect. The advantage is proportional to the inverse square root of the number of *independent* measurements in the time series. Because fluctuations associated with natural variability in the climate system last about a correlation time τ_{var} , the number of independent measurements in the time series is $\Delta t/\tau_{\text{var}}$. Putting the baseline and averaging effects together, the error in the slope becomes $|\delta m| \approx \sigma_{\text{var}}/\Delta t \times \sqrt{\tau_{\text{var}}/\Delta t}$, in agreement with Eq. (10).

3. Accuracy requirements

A series of climate benchmark missions will be used to detect trends in the climate system of any of a variety of quantities. Before a climate benchmark is designed it is customary to estimate the size of the trend m_{est} . We define the factor s to be the signal-to-noise ratio of detection $s \equiv m_{\text{est}}/|\delta m|$, and the greater the s the greater the confidence with which one can declare a measured trend statistically different from zero. The signal-to-noise ratio s is also the inverse of the fractional precision of the estimate of a trend in a time series of data. We estimate the time it takes for a signal to emerge above natural variability and measurement uncertainty with a signal-to-noise ratio s ($s = m_{\text{est}}/|\delta m|$),

$$\Delta t = \left[\frac{12s^2}{m_{\text{est}}^2} \sigma_{\text{var}}^2 \tau_{\text{var}} \right]^{1/3} (1 + f^2)^{1/3}, \quad (11)$$

where the measurement uncertainty factor f is

$$f^2 = (\sigma_{\text{meas}}^2 \tau_{\text{meas}}) / (\sigma_{\text{var}}^2 \tau_{\text{var}}). \quad (12)$$

From Eq. (11) it is clear that there is a lower bound on the time it takes to detect a signal with a prescribed level of confidence that is dictated by nature alone, and measurement uncertainty amplifies that detection time

only in proportion to natural variability. The cube-root quantity in square brackets in Eq. (11) gives the minimum time to detect a climate signal, which can be easily seen by setting the measurement uncertainty, and hence f , to zero. One consequence of a full consideration of the correlation time constant of natural variability is that detection time remains unchanged after smoothing a time series of data. By smoothing data, one can reduce the departures (σ_{var}) from a fitted line by the inverse square root of the smoothing interval, but the effective increase in the time constant of the departures [τ_{var} as defined by Eq. (7)] from the fitted line cancels the reduction in departures when considering the two in combination through the product $\sigma_{\text{var}}^2 \tau_{\text{var}}$.

Second, from Eq. (11) it is also clear that measurement uncertainty amplifies the time to detection only when considered in direct relation to natural variability (through the factor f). If one requires that $f < 0.5$, then the time to detection is only increased by $< 8\%$ over what nature allows. If one assumes that the correlation time constant of the measurement uncertainty is the lifetime of the climate benchmark instrument in question, then the accuracy requirement for the mission must be inversely proportional to the square root of the instrument lifetime. The constant of proportionality $\sigma_{\text{var}} \sqrt{\tau_{\text{var}}}$ is dictated by nature in the form of the natural variability of the climate system.

Third, from Eq. (11) it is also clear that detection time is decreased when the time constant associated with measurement uncertainty is decreased. The time constant associated with measurement uncertainty τ_{meas} is difficult to quantify, so we use the “worst-case scenario” concept described in the introduction to argue for its magnitude. In short, a dataset can be no more credible (or accurate) than one can experimentally demonstrate. If an instrument is designed to be traceable to international standards (Pollock et al. 2003) with uncertainty σ_{meas} , it is possible that the residual difference between truth and measurement can wander within the limits of σ_{meas} during the lifetime of the mission, and this would certainly aid the cause of detection by effectively decreasing the product $\sigma_{\text{meas}}^2 \tau_{\text{meas}}$ [see Eq. (8)]. It is impossible, though, to experimentally demonstrate that the residual difference between truth and measurement wanders on time scales shorter than the mission lifetime, because no measurement can be made with an uncertainty smaller than σ_{meas} . On the other hand, if the residual difference between truth and calibrated measurement persists for the duration of a satellite mission, then the uncertainty in the determination of the estimated trend is greatly enhanced through the product $\sigma_{\text{meas}}^2 \tau_{\text{meas}}$. It is impossible to demonstrate experimentally that the residual

TABLE 1. Dependence of signal detection time on measurement uncertainty (σ_{meas}) and instrument lifetime (τ_{meas}) is shown. An interannual variability of 500-hPa temperature of $\sigma_{\text{var}} = 0.18$ K with a correlation time constant of $\tau_{\text{var}} = 1.54$ yr and a trend of $m = 0.2$ K decade $^{-1}$ is assumed.

Measurement uncertainty (σ_{meas}) (K)	Detection time ($\tau_{\text{meas}} = 2$ yr) (yr)	Detection time ($\tau_{\text{meas}} = 6$ yr) (yr)
0.00	33.4	33.4
0.02	33.6	34.0
0.05	34.5	36.5
0.10	37.4	43.5
0.20	46.0	60.1
0.50	74.4	105.1

difference between the truth and measurement remains constant over an instrument's lifetime, but because the latter case is the worst-case scenario, which one cannot disprove experimentally; we argue that τ_{meas} must be the mission lifetime.

4. Example: Upper-air temperature

As an example, we assume that a satellite-borne climate benchmark instrument obtains global average temperature measurements centered at 500 hPa. We wish to assess the dependence of a global warming detection time given two possible instrument lifetimes ($\tau_{\text{meas}} = 2, 6$ yr) and an array of possible accuracy requirements for instrument uncertainty (σ_{meas}). We wish to test climate models by constraining the sensitivity of the climate to an uncertainty of 20%, so we set $s = 5$.

In Table 1 we give $s = 5$ detection times for a global warming signal of $m = 0.2$ K decade $^{-1}$. We assume that natural variability has $\sigma_{\text{var}} = 0.18$ K with time constant $\tau_{\text{var}} = 1.54$ yr for the global average temperature of the 500-hPa surface, consistent with a realistic preindustrial control run of the Met Office's Third Hadley Centre Coupled Ocean-Atmosphere General Circulation Model (HadCM3), taken from the Third Coupled Model Intercomparison Project (CMIP3) archive of climate model runs hosted by the Program for Climate Model Diagnosis and Intercomparison (PCMDI). The minimum detection time is found to be 33.4 yr. A series of climate benchmark missions, each with a 2-yr lifetime, clearly has shorter detection times than the series of missions with 6-yr lifetimes. If one chooses to follow a $f = 0.5$ rule for a science requirement, the detection time becomes 36.0 yr, just 2.6 yr over the minimum; the series of 2-yr missions must have an accuracy of 0.079 K, and the series of 6-yr missions must have an accuracy of 0.046 K. [If natural variability is uncorre-

lated from year to year ($\tau_{\text{var}} = 1$ yr), then the minimum time to detection would be 29.0 yr.]

5. Conclusions

Using standard linear regression and error propagation techniques, we have derived a formula that can be used to compute accuracy requirements for a climate benchmark mission. The requirements relate measurement uncertainty and instrument lifetime directly to the natural variability of the climate and the correlation time scale of that variability. Explicitly, those requirements are governed by Eq. (12) with $f < 0.5$.

We have found that more precise trend estimates are associated with shorter instrument lifetimes given the same measurement uncertainty. This is best understood by a simple example. If one desires a precise trend estimate over a 20-yr time series of data, flying 10 satellites with a 2-yr lifetime is preferred to flying 4 satellites with a 5-yr lifetime. The reason is that measurement uncertainties associated with the satellites' instruments can be assumed to be uncorrelated, and thus average out with more satellites. In the absence of natural variability, the uncertainty in trend determination from the 10-satellite time series will be a factor of $\approx \sqrt{10/4}$ less than the trend determination from the 4-satellite time series.

Other factors enter when establishing accuracy requirements for a mission. One must decide acceptable signal detection times given the expense involved in deploying climate benchmark satellite instruments. The minimum detection time dictated by nature must be considered. While shorter lifetimes (and hence more frequent deployment) for satellite instruments decreases detection times, the increase in the number of satellite instruments will dramatically inflate the expense of a signal detection program. On the other hand, the shorter lifetime missions generally require less accurate instruments, which are expected to be lower in cost per instrument than more accurate (and longer lifetime) instruments.

Acknowledgments. We thank Dr. Jerry Sullivan and Dr. John Dykema for their helpful suggestions. We acknowledge the modeling groups for making their simulations available for analysis, the PCMDI for collecting and archiving the CMIP3 model output, and the WCRP's Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The WCRP CMIP3 multimodel dataset is supported by the Office of Science, U.S. Department of Energy. This work was supported by Grant ATM-0450288 of the National Science Foundation.

REFERENCES

- Climate Change Science Program, 2006: Temperature trends in the lower atmosphere: Steps for understanding and reconciling differences. U.S. Climate Change Science Program, Synthesis and Assessment Product 1.1 Final Report, 164 pp.
- National Research Council, Committee on Climate Data Records from NOAA Operational Satellites, 2004: *Climate Data Records from Environmental Satellites*. National Academies Press, 102 pp.
- National Research Council, Committee on Earth Science and Applications from Space, 2007: *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. National Academies Press, 418 pp.
- Ohring, G., Ed., 2007: *Achieving Satellite Instrument Calibration for Climate Change*. National Oceanographic and Atmospheric Administration, 144 pp.
- Pollock, D., T. Murdock, R. Datla, and A. Thompson, 2003: Data uncertainty traced to SI units. Results reported in the International System of Units. *Int. J. Remote Sens.*, **24**, 225–235.
- Spencer, R., and J. Christy, 1990: Precise monitoring of global temperature trends from satellites. *Science*, **247**, 1558–1562.
- , and —, 1993: Precision lower stratospheric temperature monitoring with the MSU: Technique, validation, and results 1979–1991. *J. Climate*, **6**, 1194–1204.
- von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Williams, E., 1959: *Regression Analysis*. John Wiley and Sons, 214 pp.