Accurate Spectrally Resolved Infrared Radiance Observation from Space: Implications for the Detection of Decade-to-Century–Scale Climatic Change

David W. Keith

Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania

James G. Anderson

Department of Earth and Planetary Science, Harvard University, Cambridge, Massachusetts

(Manuscript received 10 March 1999, in final form 3 November 1999)

ABSTRACT

The character of data required to measure decade-to-century–scale climatic change is distinctly different from that required for weather prediction or for studies of meteorological processes. The data ought to possess the accuracy to detect the small secular climate changes of interest. To be useful to future investigators, the data must include convincing proof that a given level of accuracy was in fact attained.

Spectrally resolved infrared radiance is one of the most important quantities to measure accurately from space—it contains much of the fingerprint of climate response and of the forcing that causes it. The authors describe the physics of infrared radiance measurements, and demonstrate that trade-offs exist between instrument accuracy (required for climate data) and sensitivity (required for weather prediction). No such simple trade-off exists between spectral resolution and accuracy; in fact, spectral resolution can improve accuracy. The authors analyze the implications of these trade-offs for the design of climate-observing systems based on observation of infrared radiance. It is argued that convincing demonstrations of sensor accuracy requires a measurement approach founded on the overdetermination of instrument calibration, an approach that aims to reveal rather than conceal instrumental error. It is argued that the required accuracy can by achieved in simple instruments that provide spectral resolution if high sensitivity is not simultaneously demanded. Laboratory data are presented to illustrate the means by which radiometric calibration with the accuracy required for climate observation—about 0.1 K in the midinfrared—might be achieved in a practical instrument.

1. Introduction

Spectrally resolved infrared radiance is one of the most important quantities to measure accurately from space—it contains the fingerprint of both climate response and the forcing that causes it. Kiehl (1983) and Goody et al. (1996) showed that IR radiance observations could be used directly to detect the climatic response to greenhouse gas forcing. More recently, Iacono and Clough (1996) used high-resolution radiance data from Infrared Interfrometer Spectometer (Hanel et al. 1971) to investigate the seasonal cycles of tropical SSTs and water vapor. Using the same dataset, Haskins et al. (1999, 1997) have demonstrated that IR radiances can be used directly to construct powerful tests of the fidelity of climate models.

Our focus is on the measurement of climate variability on decade-to-century timescales. This focus places a premium on accuracy for two reasons. First, because the magnitude of climate response on long timescales (such as the response to solar-cycle forcing) is small compared to the variability on shorter timescales (such as the tropical response to ENSO). And second, because long-duration observation requires that data from independent instruments—operated at different times by independent investigators—be assembled into a homogeneous dataset.

If we are serious about collecting data for measurement of long timescale climatic variation, then we must pose the question: What character of data will be required by future generations? This framing implies that a data record must not only be sufficiently accurate, it must also demonstrate to future investigators that a given level of accuracy was in fact achieved. We must convince future investigators that our estimate of systematic error is robust. Meeting this challenge demands a new approach to the calibration of space-based sensors, one which is rooted in the tradition of accurate metrology in the physical sciences, which relies on the

Corresponding author address: Dr. David W. Keith, Department of Engineering and Public Policy, Carnegie Mellon University, 129 Baker Hall, Pittsburgh, PA 15213.
E-mail: keith@cmu.edu

estimation of error by means of measurements that are overdetermined or redundant.

Karl et al. (1995) contains an excellent summary of deficiencies in existing space-based meteorological datasets with respect to the accuracy and homogeneity required for climate observation. They note that, in contrast to meteorological observations for forecasting, ''long-term climate monitoring, capable of resolving decade-to-century–scale changes in climate, requires different strategies of operation.'' Our focus is on the physics that limits instrument accuracy, particularly for infrared radiometers, and on the trade-offs that emerge between the sensitivity required for operational meteorology and the accuracy required for climate observation.

The remainder of this paper is organized as follows. Section 2 discusses the desirable attributes of climate data, and examines the trade-offs between them. In section 3 we describe the advantages of direct analysis of radiance for climate studies. The instrumental physics that govern the accuracy of infrared radiance measurements are analyzed in section 4. We focus on the trade-offs between accuracy and sensitivity, and describe a calibration methodology that is appropriate for climate-observing sensors. In section 5 we describe calibration experiments performed on an aircraft spectrometer as an illustration of the general arguments made in section 4. Finally, section 6 examines the implications of the design trade-off that emerges from the instrument physics for the design of space-based observing systems.

2. Requirements for climate observation

Key characteristics of data intended for the detection of decadal-to-century timescale climatic change are as follows.

- *Accuracy:* The measurement should be accurate enough to detect a systematic change in climate over the relevant timescale. That is, the systematic error should be less than the magnitude of relevant climatic change.
- *Coverage:* Global coverage is not required; however, substantial aerial coverage is necessary to detect small secular trends against background ''noise'' due to meteorological variability.
- *Specificity:* High information content—but not necessarily large quantities of data—are required to permit selective tests of model predictions using ''fingerprint methods.''
- *Homogeneity:* Strict temporal continuity is not required; rather, we require homogeneity of measured observables—for example, choice of spectral channel.

a. Accuracy

Accuracy is a measure of systematic error—the nonrandom deviation of a measurement from the true value. Accuracy is independent of random error that is expressed as precision, sensitivity, or signal-to-noise ratio. Accuracy is assessed by testing the measurement against fundamental standards. Precision is assessed by repetitively observing a stable quantity. Precision is often measured as a routine part of instrument observation, whereas accuracy is not so easily determined. A general heuristic for assessing accuracy is to measure the same quantity with multiple instruments that use different physical measurement methods so that their systematic errors may be assumed to be uncorrelated. Assuming random error has been eliminated, the critically analyzed disagreement between measurements serves as a diagnostic test of accuracy.

Sensor accuracy is fundamental to climate observation; it allows measurements taken at different times, by different instruments and investigators, to be directly compared. The role of sensor accuracy may be appreciated by considering the record of surface air temperature. Suppose that the sensors used to measure surface air temperature in the century preceding 1950 had an accuracy of ± 1°C. Specifically, suppose the transfer standard used to calibrate the individual thermometers had a calibration that drifted on a timescale of decades and was never tested against fundamental standards. Such a record would still have much utility for climate studies, for example, it would still be possible to detect trends in the amplitude of the diurnal cycle. It would not, however, be possible to resolve the long-term trends in surface temperature, no matter what quantity of data was available. No ex post facto analysis can remove the fundamental effect of sensor inaccuracy. The utility of the instrumental record rests on the homogeneity of individual observations, the quality of the documentation, and the accuracy of the sensors with respect to fundamental standards.

The lesson from the history of ground-based observations is simple. If meteorological observations from space are to serve for measurement of climate variability on decade-to-century timescales, then the observations should meet the standards of accuracy and homogeneity (section 2d) we require of the ground-based record. The ability of space-based sensors to gather very large quantities of data does not mitigate the requirements on data quality.

b. Coverage and orbital geometry

The detection of long timescale climate variability requires measurement of signals whose amplitude is small compared to the amplitude of intrinsic variability on shorter spatial and temporal scales. Signal detection is achieved by averaging. Although averaging is intended to remove synoptic-scale variability, data that resolves the synoptic scales may still be necessary because undersampled synoptic variability may create spurious aliases that obscure the lower-frequency variability that we aim to detect. Observations from platforms in low earth orbit necessarily undersample synoptic, particularly diurnal, variability (Salby 1989; Salby and Callaghan 1997).

Diurnal aliasing has hindered analysis of data from polar-orbiting meteorological satellites. Equator-crossing times of sun-synchronous satellites have drifted, introducing spurious secular trends in temperature substantial larger than the climate variations of interest. In addition, equator-crossing times have differed between ’satellites creating measurement inhomogeneity (section 2d). While investigators have corrected for these effects in climatological datasets (Christy et al. 1998), the residual uncertainty in this correction hinders the interpretation of the results (Hurrell and Trenberth 1997, 1998).

Methods to mitigate the problem of diurnal aliasing include the following.

- Monitor observables with small diurnal cycles; for example, use infrared rather than visible radiance and/or observe the free troposphere rather than the surface.
- Ensure that polar-orbiting satellites contain fuel to permit maintenance of a given equator-crossing time (Karl et al. 1995).
- Use inclined orbits that precess so as to sample the diurnal variability (Salby 1989).
- Make climate observations from geosynchronous orbit (Salby 1989).

c. *Specificity*

A dataset with high information content may permit more effective use of ‘‘fingerprint’’ methods, for example, optimal estimation, for the detection and attribution of climatic change.

Despite the advantages of high information content, it is not optimal for a climate-observing system to measure as many variables as possible. One reason is that specificity generally requires measurements with high sensitivity, and, as shown in section 4 there may be strong trade-offs between sensitivity and accuracy.

The utility of the global radiosonde records illustrates that high data volumes are not required in a good climate record; they demonstrate the power of a good sampling strategy, accuracy, and observational homogeneity.

Another trade-off exists between specificity and system cost. High information content requires high data rates that in turn strongly drive the cost of earth-observing systems both at the level of instrument design and throughout the spacecraft and ground-based support facilities. High system cost is in conflict with the requirement of climate data continuity since it is implausible that a high cost system can be sustained over the required decade-to-century timescales.

d. *Homogeneity*

Detection of variability on decade-to-century timescales requires a data record of equal or greater length, but it need not be continuous. Random gaps in a record do not prevent its use for the analysis of variability at any timescale, though they do decrease our statistical confidence in the results. While temporal continuity is not required, the utility of a climate dataset does depend on consistent choice of observed parameters. We call this *measurement homogeneity*. Some climate observations from space, for example, are more seriously compromised by changes in observed spectral channels or the diurnal phase of observation than by temporal gaps in the data.

Gaps are a serious problem only in the case where stable but low accuracy sensors are used so that data overlap is required to transfer the relative calibration standard; for example, the microwave sounding units (MSU) series (Christy et al. 1998).

The necessity for measurement homogeneity has profound implications for the design of climate-observing systems. In order to obtain a homogeneous record future observers will be bound to continue observing the same parameters that we chose today. In the absence of other constraints, an observing system design should therefore seek to maximize the opportunity for measurement homogeneity by choosing a set of meteorological observables that is as general as possible; for example, by choosing high-resolution radiance measurements and or by choosing precessing orbits that allow sampling of the diurnal cycle.

3. Radiance as record of climate

Passive remote sensing instruments typically measure radiance. Atmospheric state variables such as temperature or humidity may be ‘‘retrieved’’ from measured radiances by inversion methods. In principle, data archived for climatic measurements could be either radiances or retrieved variables. Because the climatic variables of primary interest are the retrieved quantities, it is natural to assume that they form the centerpiece of a climate-observing system, and that system requirements such as accuracy should be stated in terms of retrieved quantities. In general, this is the approach taken for the Earth Observing System (EOS; Asrar and Greenstone 1995). Specifying requirements for the derived variables is appropriate if the goal is studies of atmospheric processes, but may not be if the goal is the creation of a long-term climate record.

a. *Limitations of retrieved quantities as the basis for a climate record*

Retrieval products have several characteristics that limit their utility as records of long-term climate change. First, many practical retrieval methods depend on knowledge of the current climate to constrain the solution (Rodgers 1990). Some modern retrieval schemes use the output of meteorological analysis/forecast systems as constraints on the retrieved solution making the retrieved quantity dependent on a myriad of other observations and on the forecast model physics. Second, because the upwelling radiance depends on many variables, and because the radiative transfer equation is not uniquely invertable, it is not usually possible to derive an inversion method for one variable that does not depend on the values of other variables. Surface temperature retrievals, for example, are influenced by the presence of subvisual cirrus clouds so that a spurious trend in retrieved surface temperature may be caused by a trend in cloud amount. Finally, retrieval algorithms are necessarily extremely complex and are thus hard to document accurately. For example, the software required to implement the retrievals for the Atmospheric Infrared Sounder (AIRS) will require more than 50 000 lines of code (L. Strow 1998, personal communication). The detection and attribution of climate change on decade-to-century timescales requires that data be archived in a manner that is comprehensible and trustworthy for future investigators. The focus on continuous incremental improvement that is rightly central to operational retrieval systems is naturally at odds with the more conservative approach required to archive climate data for future use.

b. Radiance climatology: Radiances versus retrieved variables

Rather than using retrieved variables we may use radiance as a fundamental climate record. Radiance observations may be used directly as a basis for detecting both the climate forcings and the response to forcing (Goody et al. 1996; Kiehl 1983). Haskins et al. (1997, 1999) have shown that key covariance statistics of climate models can be directly compared with radiance observations.

The use of radiance as a climate record constitutes a frank acceptance of the fact that we cannot measure from space all of the observables that are the usual basis of terrestrial meteorology, and that we cannot directly compare space-based and terrestrial observations. By omitting the need for retrievals, the direct use of radiances in climate analysis provides a more mathematically direct comparison between theory and observation. Mathematical directness offers the potential for more statistically powerful tools for the detection and attribution of climate change. Such tools may permit more precise rejection of the effect of extraneous variables. There is a strong analogy to be drawn from recent advances in meteorological data assimilation that have shown that direct assimilation of observed radiances into the model, rather than use of the retrieved quantities, can improve the analysis accuracy. For example, Derber and Wu (1998) document improvements in the National Centers for Environmental Prediction (NCEP) analyses using direct assimilation of Television Infrared Observational Satellite (TIROS) Operational Vertical Sounder (TOVS) radiances (now a part of the operational system). They argue that direct assimilation offers advantages because both the ''first guess'' field and its error covariances are automatically consistent as would not be possible in a conventional retrieve-then-analyze system.

Because the raw instrument data are always archived, it may appear that the choice of analysis method—radiances or retrievals—may be made ex post facto, and need not influence the design of the observing system. While this is theoretically correct, it neglects key practical issues. First, there are many ways in which the choice of analysis method strongly influences the observing mission design. For example, since temperature retrievals in the infrared are limited to an accuracy of ~1 K no matter what the accuracy of the measured radiances, a mission whose goal is to produce retrieved temperatures will not specify an instrument radiance accuracy that is significantly better than ~1 K (e.g., the EOS AIRS instrument; Aumann and Overoye 1996). Second, if the focus of a mission is to produce derived quantities, then it is unlikely that the archive of raw radiances will be sufficiently quality controlled and documented to serve the needs of future investigators. In contrast, a mission design aimed at the archiving of high quality radiances will ensure a consistent focus on quality control. Such an archive allows later investigators to make their own choice of analysis method, thus facilitating the construction of a homogeneous data record as defined in section 2d.

4. Calibration of infrared radiometers

a. Accuracy requirements in the thermal infrared

Section 2a described the general accuracy requirements for climate observation; we focus here on the accurate measurement of IR radiance.

The accuracy required to detect decade-scale climatic change in IR radiance may be estimated using the reasoning of Section 2a. To detect the climate's response to large-scale forcings such as variations in solar irradiance or CO_2, we must measure changes in average surface temperature of order 0.1 K over decadal timescales, as determined by the lower bounds on the predicted response to the given forcings. In the mid-IR, between about 5 and 20 μm where the key fingerprints of climate response are located, a 0.1 K change in surface temperature produces spectrally varying changes in brightness temperature of ~0–0.3 K on a mean background temperature of 270 K (Goody et al. 1996; Kiehl 1983). Thus we require a brightness temperature accuracy of about $\pm$0.1 K on a 270 K background, which at 15 μm is a radiometric accuracy of 1.5 parts in 10^3.

An accuracy of ~1 part in 10^3 in the mid-IR is higher than is achieved by existing space-based radiometers. Consider the EOS AIRS instrument, the most technically complex infrared sounder developed to date: AIRS has an accuracy design requirement of 1.4 K at 15 μm

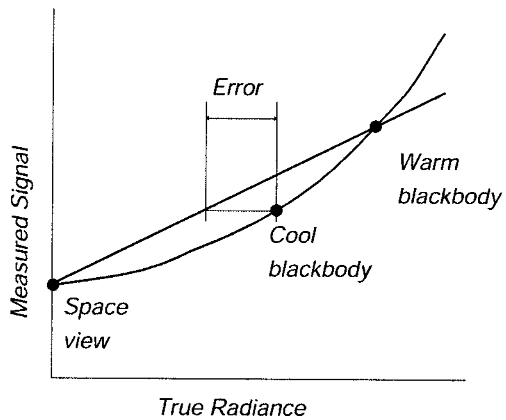

FIG. 1. Schematic illustration of instrument calibration using three radiance standards. For a linear instrument, use of three standards permits estimation of the sensor's end-to-end radiance error.

and 0.5 K at 4.2 μm (Aumann and Overoye 1996). The EOS Clouds and Earth's Radiant Energy System (CERES) sensor has an accuracy goal of 0.5% in its broadband channel that covers the thermal and solar bands (Lee et al. 1998).

Given that current infrared sensors do not have the accuracy required to directly detect long timescale variations in infrared radiance, it is natural to ask whether the required accuracy is realizable. If it is realizable, how difficult is it to achieve if the instrument is designed from first principles for the task? In order to suggest that substantial improvements in accuracy may be readily achieved, and to explain why current sensors do not have such accuracy, it is necessary to describe the instrumental physics that limits accuracy and to illustrate the trade-offs in observing system design that emerge from the instrument physics.

b. A calibration methodology for climate sensors

A good climate dataset must not only be accurate. It must convincingly demonstrate to future investigators that a given level of accuracy was in fact achieved. The key to meeting this dual challenge is to construct a calibration system that is overdetermined so that radiometric errors may be critically analyzed in orbit. One must design instruments that reveal—rather than conceal—their errors.

For a linear instrument, overdetermination of the calibration requires three measurements, as shown schematically in Fig. 1. We will show in section 4c that the principle of overdetermination can be applied to many components of a radiometer in order to improve understanding of systematic error. The proposed approach to sensor design amounts to the construction of a self-contained experiment that measures the consistency and stability of a radiometric calibration system.

Overdetermination of measurement is the traditional pragmatic focus of precision metrology. Confidence that systematic error has been controlled demands an objective method of assessing it—typically by the construction of alternative metrology systems using, where possible, different physical principles and different detailed designs.

This approach to sensor deviates from that used on most spacecraft sensors where the canonical philosophy is to exhaustively calibrate and characterize a sensor on the ground in order to estimate, and correct for, the various factors that could cause drift in the calibration during prolonged exposure to the space environment. When clear demonstration of accuracy is the goal, this method has several pitfalls. First, elaborate error analysis of this kind tends to focus on random errors because they are the easiest to assess. Yet for climate data, where the goal is accurate measurement of long-term averages, estimating and reducing systematic error is the key. Second, substantial complexity and expense are required to correctly execute the required error analysis and correction. Because of this complexity, it is difficult to make an unequivocal case that such efforts have yielded a robust estimate of systematic error. Experience with precision metrology shows that where practical, it is both simpler, and far more convincing to attack systematic error with redundant measurements.

Validation through a program of testing in orbit differs fundamentally from the strategy of validation by ''ground truth'' that is the norm in earth observation. For radiance measurements of the accuracy required for climate observation, validation by comparison with conventional observations is impractical. The problem of inadequate overlap of observing system footprints, which limits our ability to test conventional earth-observing systems, makes ground-based validation of a high-accuracy sensor implausible. A focus on ground truth validation is appropriate for testing temperature sounders for which the primary uncertainty is the retrieval method that may be tested by comparison of retrievals with observed soundings. It is wholly inappropriate for testing the radiance measurements themselves.

The fundamental design trade-offs discussed below demonstrate that instrument accuracy cannot be an afterthought. High accuracy cannot be achieved after the fact by building an instrument with high sensitivity and then calibrating and characterizing on the ground. Accuracy must be a central design goal.

c. Sources of error

For the purpose of analyzing radiometric error, a calibrated spectrometer may be considered as three subsystems: a spectrometer and detector(s), calibration sources, and a scan or pointing mirror that permits the spectrometer to view either the target scene or the calibration sources. For each subsystem we review the leading sources of error, the means of reducing the errors, and the prospects for estimating the error by in-orbit measurements.

1) Spectrometer and detectors

In an ideal spectrometer, the signal in a given channel is the convolution of the observed radiance with the channel line shape (CLS). Errors arise through uncertainty in the CLS or instrument nonlinearity. Nonlinearity may arise in any part of the analog signal chain from detectors through to analog-to-digital conversion. It is typically dominated by the detectors. It may be reduced by measuring the nonlinearity and correcting for it (Abrams et al. 1994; Aumann and Overoye 1996), or by choosing detector technologies that are inherently linear. Empirical nonlinearity correction requires the introduction of free parameters, the coefficients of the nonlinear detector sensitivity, into the radiometric calibration. While this reduces the residuals from the calibration curve fitting, it inevitably reduces confidence in error estimation. Consider the simple case illustrated in Fig. 1. End-to-end radiometric error, due to blackbodies, pointing mirror, etc., may be effectively estimated in orbit by use of three blackbodies to overdetermine the calibration. A one-parameter nonlinearity correction allows a perfect fit through three calibration points. However, it may hide radiometric errors and so eliminate the possibility of simple end-to-end error estimation.

Three kinds of IR spectrometers have been used in space-based sensors: discrete filters, diffraction gratings with detector arrays, or Fourier transform spectrometers (FTSs). Where photon-noise-limited detectors are available, as is true in the near-IR, filter or grating instruments offer higher sensitivity than is possible with FTS (Beer 1992). In addition, they can be less sensitive to detector nonlinearity than are FTSs, because each detector sees less signal than is the case in an FTS. However, for filter and grating instruments, the line shapes for each channel are independently determined by the details of alignment and construction. In a high-resolution instrument this may amount to thousands of independent parameters, each with associated uncertainties. For an FTS, the line shapes for all channels are determined by a few independent parameters.

Radiometric calibration using blackbodies provides the normalized calibration for each channel, but cannot determine the CLS. A space-borne spectrometer must either have means of calibrating the CLS in orbit, or it must rely on preflight spectral calibration combined with engineering estimates of the magnitude of the expected drift in the CLS during launch and subsequent lifetime in orbit. In an FTS, a fit to the known properties of the outgoing radiance can determine the few CLS parameters in orbit. For grating or filter instruments, spectral calibration is more difficult and tends to drift with time. (Note that such instruments are usually calibrated using an FTS.) When IR spectrometers serve as temperature sounders for operational weather prediction, high spectral accuracy is not required as the assumed CLS can be adjusted along with other parameters to improve the fit of retrieved temperatures to rawinsonde observations (Aumann and Pagano 1994). Climate observation, however, demands spectral accuracy consistent with the radiometric calibration, as will be described in section 4e(1).

2) Calibration system

Infrared spectrometers use blackbodies as calibration sources. The radiance emitted by an isothermal blackbody is the product of its emissivity and the Planck function evaluated at the blackbody temperature. Ideally, the emitted radiance will depend only on thermometry and on physical constants. Errors arise from three sources: uncertainty in emissivity, temperature inhomogeneity, or errors in thermometry.

In general, thermometry is good enough that errors can be reduced below ~0.05 K, so that they make a small contribution to the overall radiometric error budget. Thermometric error can be simply and effectively estimated in orbit by the use of redundant transducers that use different measurement technologies so that the drift in their calibration with age or radiation exposure may be assumed uncorrelated.

The reflectivity of a cavity blackbody—the compliment of its emissivity—is proportional to the surface reflectivity and inversely proportional to the square of the aspect ratio; that is, $r_{cavity} \propto r_{surf}(l/d)^{-2}$ where d is the aperture diameter and l is the cavity length (Chandos and Chandos 1974). It is possible to use blackbodies with a reflectivity greater than the desired system accuracy. For example, blackbodies with $r_{cavity} = 2 \times 10^{-2}$ could be used to calibrate a spectrometer to an accuracy of 1×10^{-2} if the emissivity was estimated and included in the calibration. This approach makes the calibration dependent on r_{surf}, which may change in orbit. A more conservative approach, suitable for climate measurement, is to make the blackbody reflectivity sufficiently low, by increasing (l/d) so that we may assume perfect emissivity ($r_{cavity} = 0$) even if aging causes r_{surf} to increase severalfold.

Errors caused by thermal gradients tend to grow linearly with the size of a blackbody cavity. Thus, as will be discussed in section 4d, there are trade-offs between the instrument sensitivity—which grows with blackbody size—and accuracy.

3) Pointing mirror and input optics

Ideally, the pointing mirror and spectrometer input optics act as a perfect switch allowing the spectrometer to view various targets with equal efficiency. In practice errors arise due to stray light and pointing mirror polarization.

Radiation that reaches the detector from outside the intended target is considered ''stray light.'' Errors are introduced by stray light during calibration because radiation from uncontrolled sources outside the blackbody apertures contributes erroneously to the apparent blackbodies radiance. Stray light can be reduced by various optical design choices, such as the use of baffles. In general, reduction in stray light is achieved at the expense of system sensitivity.

Spectrometers typically have a polarization dependent sensitivity. The pointing mirror will necessarily be weakly polarizing. Together, the pointing mirror and spectrometer form a polarizer/analyzer combination that modulates the system transmission as the mirror rotates to observe different targets, thus invalidating the blackbody-to-target comparison. In the AIRS sounder, for example, pointing mirror polarization errors are ~0.9 K equivalent temperature at 10 μm, which—it is estimated—will be reduced to ~0.2 K by a first-order correction based on preflight measurements and engineering estimates of the stability of the mirror optical properties in orbit. Application of the calibration philosophy described in section 4b suggests that polarization errors may be more convincingly removed by building a system that measures them in orbit. This was the approach chosen for the Arrhenius mission (Anderson et al. 1996), which used a 45-deg range of space view to permit calibration of mirror polarization.

d. Accuracy versus sensitivity

A direct trade-off between accuracy and sensitivity is evident in the device physics of infrared detectors. In the mid-IR, the detector of choice is typically a photoconductive HgCdTe semiconductor (PC-HgCdTe). In these detectors, increased sensitivity is achieved by increasing the lifetime of charge carriers so that a larger fraction of the carriers produced by the incident radiation are detected. Increased carrier lifetime necessarily increases the efficiency of electron–hole recombination—proportional to the square of the carrier density—which in turn increases the detector's nonlinearity. Thus, the most sensitive PC-HgCdTe detectors are the least linear. This accuracy versus sensitivity trade-off is a common design problem for real FTS systems (H. Buijs 1999, personal communication). In contrast, older technologies such as pyroelectric detectors are both much less sensitive and much more linear.

Accuracy and sensitivity are also in conflict in the optical design where increased sensitivity is achieved by increasing in an instrument's ''entendu''—its area $\times$ solid-angle product. The size of the blackbodies (for a given emissivity) increases as the square root of the entendu. Larger blackbodies tend to be less isothermal and thus less accurate [section 4c(2)]. In addition, there are trade-offs between entendu and stray light performance [section 4c(3)]. Finally, there is a trade-off with respect to spectrometer choice. FTSs are less sensitive than grating instruments in the near-IR, but may be simpler to calibrate accurately [section 4c(1)].

e. Spectral resolution and accuracy

While several causes place instrument accuracy and sensitivity in conflict, the converse is true for accuracy and spectral resolution. Spectrally resolving instruments may be calibrated more accurately than broadband instruments (Goody and Haskins 1998). Stated simply, in-orbit calibration of high-resolution instruments can determine and correct for changes in the spectrally varying instrument responsivity while broadband instruments can only determine the integrated responsivity.

Suppose we wish to measure radiance over a broad spectral interval with high accuracy. A direct approach is to construct an instrument with a CLS, $r(\nu)$, that matches the spectral region of interest. Independent of what kind of spectrally selective system is used the instrument will include blackbodies that allow accurate calibration of the integrated channel responsivity while in orbit. If r is the dimensionless factor scaling measured instrument radiance to the true radiance then calibration of the channel is equivalent to measurement of its integrated width $\Delta\nu$,

$$\int_0^\infty r(\nu)\, d\nu = \Delta\nu. \tag{1}$$

The true mean radiance over the spectral band weighted by r is then given by

$$\tilde{I} = \frac{1}{\Delta\nu}\int_0^\infty rI\, d\nu, \tag{2}$$

where $I(\nu)$ is the incident radiance. We assume that the calibration system *perfectly* determines the integrated responsivity. If the channel width is sufficiently narrow that the Planck function does not vary significantly across it, then determination of the calibration constant $\Delta\nu$ in (1) is achieved by observation of two blackbodies with I factored out of the integral in (2).

There is no way to monitor the shape of $r(\nu)$ while in orbit, so our estimate of it will be in error. We may define $r = r_0 + r'$ where r_0 is an estimate of $r(\nu)$ and r' is its error. Similarly for the radiance, we have a theoretically predicted radiance I_0, and a deviation I' that is the quantity we wish to determine experimentally. The measured $\tilde{I}$ may now be written as

$$\tilde{I}\Delta\nu = \int_0^\infty rI\, d\nu = \int_0^\infty r_0 I_0\, d\nu + \int_0^\infty r' I_0\, d\nu + \int_0^\infty r_0 I'\, d\nu + \int_0^\infty r' I'\, d\nu. \tag{3}$$

The four terms on the lhs of (3) may be interpreted as follows. The product of the theoretical radiance and repsonsivity ($r_0 I_0$) is a constant with no error. The $r_0 I'$

term is what we wish to measure experimentally, the difference between true and predicted radiance weighted by the theoretical instrument response function. The $r'I_0$ and $r'I'$ terms represent errors in our determination of r_0I' due respectively to covariance of uncertainty in the CLS with the theoretical radiance and the covariance of uncertainty in CLS with uncertainty in radiance.

Now consider the use of a spectrally resolving instrument to measure the radiance over the same broad interval $\Delta\nu$. The spectrometer has an ensemble of channels within $\Delta\nu$, each with responsivity $r_n(\nu)$, each of which are individually normalized when the instrument is calibrated using the blackbodies. We may now synthesize a broadband channel $r_n(\nu)$ with the original width,

$$r_\Sigma(\nu) = \sum r_n(\nu) \quad \text{with} \int_0^\infty r_\Sigma(\nu)\, d\nu = \Delta\nu.$$

In general, the error committed using r_Σ can be much less than the error using a broadband r because resolving the spectral interval $\Delta\nu$ allows the calibration process to measure and correct for the variations in intrinsic instrument responsivity. If a grating or multiple-filter spectrometer is used to establish the $r_n(\nu)$, the relationship between the channel responsivities is arbitrary, and so the broadband responsivity $r_\Sigma(\nu)$, may have spectral structure at the scale of the instrument resolution. This can cause substantial error if the radiance, I, has structure at the same scale. An FTS resolves this difficulty, because the sum of adjacent channel responsivities automatically produces a smooth broadband responsivity.

Our treatment has deviated slightly the treatment of Goody and Haskins (1998), who assumed that the measurement must determine the mean radiance over a spectral width $\Delta\nu$; for example, they assumed that $r(\nu)$ must be a perfect ''top hat'' function,

$$\bar{I} = \int_{\Delta\nu} I\, d\nu.$$

To illustrate these issues consider the Earth Radiation Budget Experiment (ERBE) longwave channel; Green and Avis (1996) have show that $\tilde{I}/\bar{I}$ can vary by 10% for expected variations in scene radiance I. Goody and Haskins calculate that an ideal 2.5 cm^{-1} resolution spectrometer with the same $r(\nu)$ as ERBE would have an $\tilde{I}/\bar{I}$ variability of only $\sim 10^{-4}$, because the resolving instrument can measure $r(\nu)$ and then synthesize an $r_\Sigma(\nu)$ with a top hat–shaped responsivity. We note that the ERBE longwave channel played little role in the standard ERBE data product—partially on account of the difficulty in unfiltering—and thus that the uncertainty described here are not representative of the ERBE instrument.

In summary, even if the goal is only to make an accurate measurement of broadband radiance, it may still be advantageous to use a resolving instrument.

CONSISTENT SPECTRAL AND RADIOMETRIC ACCURACY

If the goal is both accuracy and high spectral resolution, then radiometric and spectral errors must be considered together. Consider a single narrow spectral channel, defined by an $r(\nu)$ that is to be accurately measured as part of a climate record. If the radiometric calibration has an offset error α, and a normalized gain error β, so that $I_{obs} = a + (1 - \beta)I_{true}$, then the first-order error in channel radiance, $\tilde{I}$, may be expressed as

$$(\alpha + \beta\tilde{I}) + \frac{1}{\Delta\nu}\int_0^\infty r'I_0\, d\nu.$$

Here the first term is the radiometric error and the second the line-shape error [Eq. (3)]; higher-order terms have been omitted. Consistent spectral and radiometric accuracy obtains when the error terms are roughly equal. If we sum adjacent channels to synthesize a wider channel, as described in the preceding section, then the line-shape errors may be greatly reduced. But if we wish to minimize the error in a single channel then both line-shape and radiometric errors must be simultaneously minimized.

5. Demonstration of highly accurate radiometry

In order to illustrate some of the design trade-offs described in section 4, we describe laboratory calibration experiments performed on the Interferometer for Emission and Solar Absorption (INTESA), an infrared spectrometer that has flown on a National Aeronautics and Space Administration (NASA) ER-2 high-altitude aircraft. INTESA is an FTS that covers a spectral range from 3 to 50 μm with a resolution of 0.7 cm^{-1}. The radiometric calibration system includes three blackbodies, one of which is designed to operate at cryogenic temperatures to provide an accurate zero reference (Keith et al. 2000, manuscript submitted to *Appl. Opt.*).

INTESA shares many design features with a spectrometer intended for climate observation from space that was proposed as the centerpiece of the Arrhenius mission (Anderson et al. 1996; Goody et al. 1998). The mission is motivated by the approach to climate observation advocated here. It is focused on high-accuracy spectrally resolved measurement of infrared radiation in combination with a low cost approach to mission design.

INTESA's design illustrates the trade-offs between sensitivity and accuracy discussed in section 4. It uses a three-optic image relay system that facilitates the achievement of high radiometric accuracy by minimizing the size of the beam at the calibration blackbodies at the cost of signal-to-noise performance. INTESA uses both pyroelectric and semiconductor detectors. Their performance illustrates the trade-off between detector sensitivity and linearity.

Laboratory results

Systematic errors in the radiometric calibration of INTESA were investigated using the instrument's three-blackbody calibration system supplemented with an external laboratory blackbody designed to exceed the performance of the internal blackbodies. The core of the laboratory blackbody is an aluminum cylinder ~120 mm long by 30 mm diameter, that is coated internally with ''enhanced black'' (Lockheed–Martin, Denver, Colorado). Its on-axis emissivity is estimated to be greater than 0.9995. Temperature inhomogeneity within the core was <0.1 K, and thermometric error was <0.05 K. Experiments were performed in a dry nitrogen atmosphere.

End-to-end instrumental errors are assessed by comparing the measured temperature of a blackbody with its radiance as measured by the calibrated instrument. The instrument's calibration is in turn derived from the observation of at least two other blackbodies using the methods of Revercomb et al. (1988). When more than three blackbodies are used, errors can be readily assessed by plotting the raw signal in a given spectral band *versus* the blackbody radiance in that band as predicted by the Plank function and its measured temperature, with a data point for each blackbody. Figure 2 shows the laboratory data in this format. Figure 3 shows the equivalent temperature error as a function of wavenumber.

An overall radiometric error of ~0.1 K was achieved across the spectral region from 250 to 1400 cm^{-1}. Expressed as radiance, our error was less than 10^{-3} in the mid-IR. In order to achieve this accuracy over such a wide dynamic range it was necessary to use the pyroelectric detector. Lower accuracy was obtained with the high-sensitivity PC-HgCdTe detector.

The laboratory accuracy achieved here is comparable to that achieved by the University of Wisconsin Atmospheric Emitted Radiance Interferometer (AERI) instrument (Revercomb et al. 1993) in calibration tests against a National Institute of Standards and Technology (NIST) blackbody. The AERI-NIST test (Kannenberg 1998) demonstrated radiometric consistency of 0.03 K, with the National Institute of Standards and Technology (NIST) blackbody at 303 K. In comparison, the INTESA calibration spans a ~10 times larger temperature range and has errors about 3 times larger.

Both of these laboratory results were achieved on instruments that incorporate the many design compromises required in an operational system. They do not represent an upper bound on achievable accuracy. Based on knowledge of the leading systematic errors in INTESA we estimate that a laboratory instrument designed solely for radiometric accuracy could reasonably achieve an order-of-magnitude higher accuracy than demonstrated here. These results suggest high accuracy may be achieved in operational instruments if it is a central design goal, and if one does not simultaneously demand maximum sensitivity.

6. Conclusions

If meteorological observations from space are to serve as a long-term climate record, they must meet the standards of accuracy and measurement homogeneity required of terrestrial observations. Few existing space-based observing systems meet these standards.

Meeting the requirements of a climate record represents a serious challenge for space-based meteorological observation. Central to the difficulty is the strong divergence between the requirements for climate observation versus those for meteorology and process studies. This divergence is evident in both system and instrument design. Consider system design. The inhomogeneity of observed parameters and of analysis systems designed for weather prediction is an inherent consequence of the continuous change and improvement that is—correctly—the focus of operational weather prediction (Karl et al. 1995). While operational analysis systems offer a means of systematically combining inhomogeneous observations into a coherent climate record, the resulting record is then heavily dependent on the analysis algorithm and the forecast model physics (Trenberth and Guillemot 1998). Analysis system discontinuity may be resolved by reanalysis of the raw data, however when Santer et al. (1999) compared reanalysis datasets from NCEP and the European Centre for Medium-Range Weather Forecasts (ECMWF) with the observational record they found that substantial temporal discontinuities remain, limiting the utility of the reanalyses as climate records. These difficulties demonstrate that post facto analysis cannot compensate for intrinsic inhomogeneity in the observational record.

Now consider instrument design. Both weather prediction processes studies require observing systems that deliver high spatial and temporal resolution. These systems necessarily require sensors with high sensitivity and angular resolution. Yet, as we have shown, in the design of practical sensors these requirements may conflict. Despite their central importance these fundamental design trade-offs are often overlooked in debate about how to make space-based observing systems more useful for climate studies. For example, analysis of the accuracy requirements for climate data suggest that attempts to improve the accuracy of meteorological radiometers by additional preflight engineering efforts and postflight ''ground truth'' observational campaigns are unlikely to yield the convincing demonstration of accuracy needed for a long-duration climate record.

Evidence of these difficulties may be seen in the existing record of meteorological observations from space that suffers from serious limitations with respect to accuracy and homogeneity. For example, note the continuing difficulty in interpreting the ~20-yr record from the microwave sounding units (MSU; Christy 1995; Christy et al. 1998; Hurrell and Trenberth 1997, 1998; Wentz and Schabel 1998). It appears unlikely that these

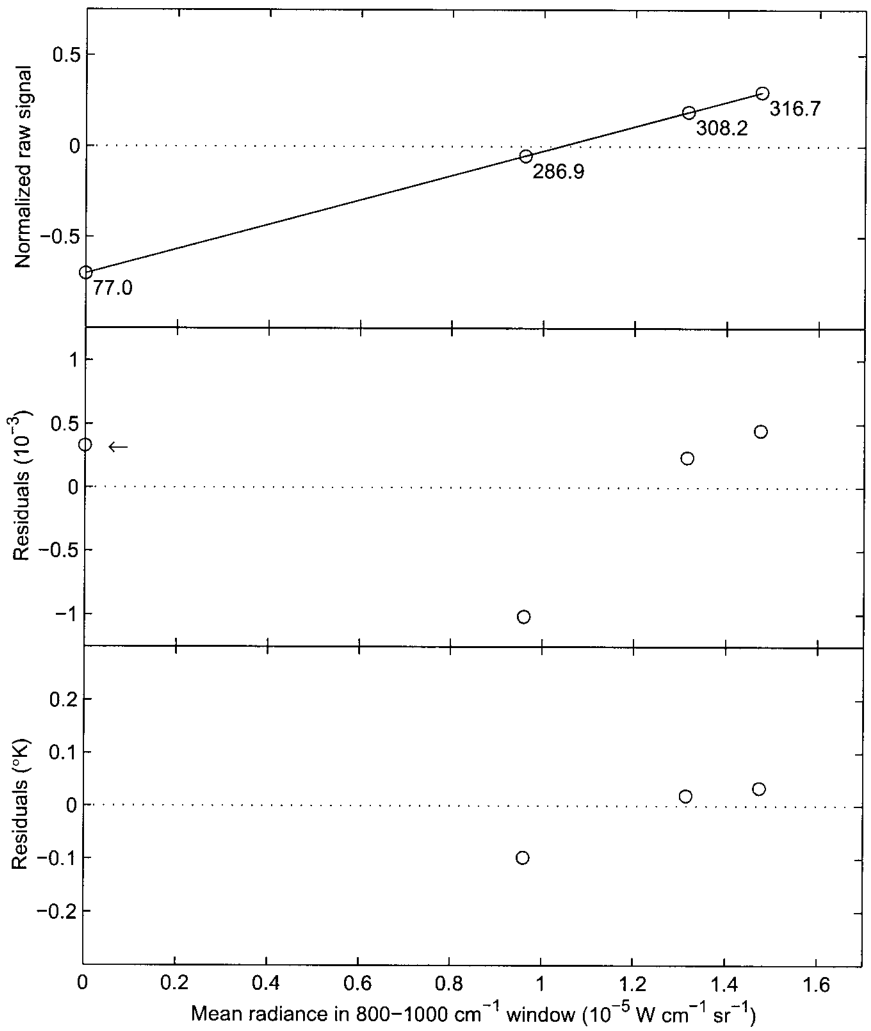

FIG. 2. Laboratory data demonstrating high accuracy radiance calibration. All data are averaged over the 800–1000-cm^{-1} spectral window. The *x* axis of each panel shows the mean radiance in the spectral window of the Planck function evaluated at the measured blackbody temperature, except for the cryogenic blackbody for which a temperature of 77 K is assumed. (top) Plots the raw signal for each blackbody with each point labeled by temperature. The signals are phase corrected using the complex phase angle of the warmest blackbody to correct all others; the value plotted is the real part of the result in with the minimum-to-maximum range of the signal normalized to unity. (middle) Shows the residuals from a linear fit to the data, equivalent to the residual radiance error at each point normalized to the full-scale radiance. The arrow indicates the left-most point. (bottom) Shows the residuals in equivalent temperature units. The 77 K point is omitted because its equivalent temperature error is too large.

difficulties will be resolved by new sensors planned or under development.

Recognition that a root difficulty with space-based climate observation lies in the conflicting demands of climate and meteorology rather than in the inherent difficulty of climate observation suggests a solution. While climate observation should not be divorced from operational meteorology, the design of observing systems must take full account of the conflicts between the two. Solutions based purely on augmentation of existing meteorological sensors are not likely to be a practical or cost-effective route to climate observation.

Use of high accuracy sensors obviates the requirement for strict satellite-to-satellite temporal overlap that is necessary if stable, but insufficiently accurate, sensors are used. Depending on sensor overlap is a risky strategy for generating a climate record as the necessity to avoid gaps will substantially increase costs. In any case, for a decade-to-century–scale record, gaps are inevitable.

These considerations suggest an alternative approach to climate observation, an approach that is focused on measuring a small number of key observables with high accuracy. A climate dataset does not require high spatial resolution, daily global coverage, high signal-to-noise

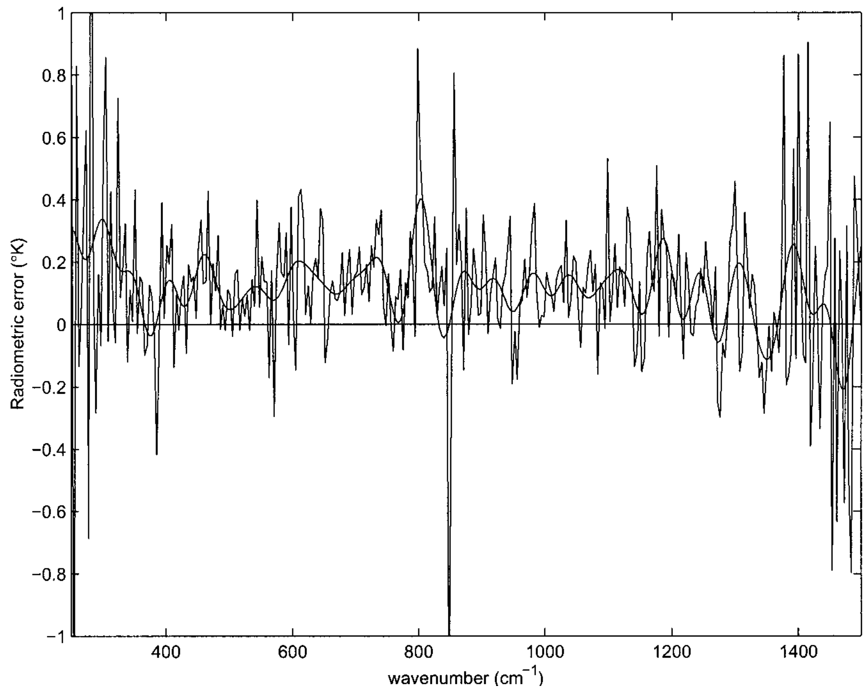

FIG. 3. Laboratory data showing radiometric error as a function of wavenumber. The instrument was calibrated using blackbodies at 316 and 77 K. Calibration and phase correction was performed using the method of Revercomb et al. (1988). Data shows the error when viewing a third blackbody at 287 K, obtained by subtracting the measured blackbody temperature from the equivalent temperature of the measured radiance. The jagged line shows the data at the raw resolution of $\sim$2 cm^{-1}, the smooth line show data at a resolution of $\sim$25 cm^{-1}. The mean equivalent temperature offset across the 300–500 cm^{-1} band is 0.12 K. The fact that the error is roughly constant across the band suggests that the leading residual error is due to thermometry rather than to a radiance offset or a nonlinearity.

ratios for individual measurements, or simultaneous measurements of several quantities. These are the factors that drive up the costs of earth observation. Two examples of this approach, one based on high accuracy high-resolution measurement of infrared radiance, and the other based on high accuracy measurement of atmospheric refractivity profiles are described by Goody et al. (1998). Climate observation can be inexpensive if we attack the problem with simple sensors that generate far less data than the meteorological sensors to which we are accustomed.

Acknowledgments. We thank Richard Goody for his devotion to the goal of simple affordable climate observation, and John Dykema for support of the experimental work. The paper benefited from a review by Bruce Wielicki. This work was supported under NASA Grant NAG-1-1849. We dedicate this paper to our late friend and collaborator Robert Haskins.

REFERENCES

Abrams, M. C., G. C. Toon, and R. A. Schindler, 1994: Practical example of the correction of Fourier-Transform spectra for detector nonlinearity. *Appl. Opt.,* **33,** 6307–6314.

Anderson, J. G., and Coauthors, 1996: Arrhenius: A small satellite for climate research. A proposal submitted to NASA-HQ December 1996, No. ESSP-II-0030, 160 pp. [Available from the authors or online at http://www.arp.harvard.edu/sci/rad/arrhenius.]

Asrar, G., and R. Greenstone, 1995: *MTPE EOS Reference Handbook.* EOS Project Science Office, NASA Goddard Space Flight Center, 287 pp.

Aumann, H. H., and R. J. Pagano, 1994: Atmospheric infrared sounder on the Earth Observing System. *Opt. Eng.,* **33,** 776–784.

——, and K. Overoye, 1996: The Atmospheric Infrared Sounder (AIRS) on the Earth Observing System: In-orbit radiometric spectral calibration. *Proc. SPIE,* **2744,** 712–721.

Beer, R., 1992: *Remote Sensing by Fourier Transform Spectrometry.* Wiley, 153 pp.

Chandos, R. J., and R. E. Chandos, 1974: Radiometric properties of isothermal diffuse wall cavity sources. *Appl. Opt.,* **13,** 2142–2151.

Christy, J. R., 1995: Temperature above the surface layer. *Climatic Change,* **31,** 455–474.

——, R. W. Spencer, and E. S. Lobl, 1998: Analysis of the merging procedure for the MSU daily temperature time series. *J. Climate,* **11,** 2016–2041.

Derber, J. C., and W. S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.,* **126,** 2287–2299.

Goody, R., and R. Haskins, 1998: Calibration of radiances from space. *J. Climate,* **11,** 754–758.

——, J. Anderson, and G. North, 1998: Testing climate models: An approach. *Bull. Amer. Meteor. Soc.,* **79,** 2541–2549.

——, R. Haskins, W. Abdou, and L. Chen, 1996: Detection of climate forcing using emission spectra. *Earth Obs. Remote Sens.,* **5,** 22–33.

Green, R. N., and L. M. Avis, 1996: Validation of ERBS scanner radiances. *J. Atmos. Oceanic Technol.,* **13,** 851–862.

Hanel, R. A., B. Schlachman, D. Rodgers, and D. Vanous 1971: *Nimbus 4* Michelson Interferometer. *Appl. Opt.,* **10,** 1376–1382.

Haskins, R. D., R. M. Goody, and L. Chen, 1997: A statistical method for testing a general circulation model with spectrally resolved satellite data. *J. Geophys. Res.,* **102,** 16 563–16 581.

——, ——, and ——, 1999: Radiance covariance and climate models. *J. Climate,* **12,** 1409–1422.

Hurrell, J. W., and K. E. Trenberth, 1997: Spurious trends in satellite MSU temperatures from merging different satellite records. *Nature,* **386,** 164–167.

——, and ——, 1998: Difficulties in obtaining reliable temperature trends: Reconciling the surface and satellite microwave sounding unit records. *J. Climate,* **11,** 945–967.

Iacono, M. J., and S. A. Clough, 1996: Application of infrared interferometer spectrometer clear sky spectral radiance to investigations of climate variability. *J. Geophys. Res.,* **101,** 29 439–29 460.

Kannenberg, R., 1998: IR Instrument Comparison Workshop at the Rosenstiel School of Marine and Atmospheric Science (RSMAS). *Earth Obs.,* **10** (3), 51–54.

Karl, T. R., and Coauthors, 1995: Critical issues for long-term climate monitoring. *Climatic Change,* **31,** 185–221.

Kiehl, J. T., 1983: Satellite detection of efects due to increased atmospheric carbon dioxide. *Science,* **222,** 504–506.

Lee, R. B., and Coauthors, 1998: Prelaunch calibrations of the Clouds and the Earth's Radiant Energy System (CERES) tropical rainfall measuring mission and Earth Observing System morning (EOS-AM1) spacecraft thermistor bolometer sensors. *IEEE Trans. Geosci. Remote Sens.,* **36,** 1173–1185.

Revercomb, H. E., H. Buijs, H. B. Howell, D. D. Laporte, W. L. Smith, and L. A. Sromovsky, 1988: Radiometric calibration of Ir Fourier-Transform spectrometers—Solution to a problem with the high-resolution interferometer sounder. *Appl. Opt.,* **27,** 3210–3218.

——, F. A. Best, R. G. Dedecker, T. P. Dirkx, R. A. Herbsleb, R. O. Knuteson, J. F. Short, and W. L. Smith, 1993: Atmospheric Emitted Radiance Interferometer (AERI) for ARM. Preprints, *Fourth Symp. on Global Change Studies,* Anaheim, CA, Amer. Meteor. Soc., 46–49.

Rodgers, C. D., 1990: Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.,* **95,** 5587–5595.

Salby, M. L., 1989: Climate monitoring from space: Asynoptic sampling considerations. *J. Climate,* **2,** 1091–105.

——, and P. Callaghan, 1997: Sampling error in climate properties derived from satellite measurements: Consequences of undersampled diurnal variability. *J. Climate,* **10,** 18–36.

Santer, B., J. J. Hnilo, J. S. Boyle, C. Doutriaux, M. Fiorino, D. E. Parker, K. E. Taylor, and T. M. L. Wigley, 1999: Uncertainties in ''observational'' estimates of temperature change in the free atmosphere. *J. Geophys. Res.,* **104,** 6305–6333.

Trenberth, K. E., and C. J. Guillemot, 1998: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalyses. *Climate Dyn.,* **14,** 213–231.

Wentz, F. J., and M. Schabel, 1998: Effects of orbital decay on satellite-derived lower-tropospheric temperature trends. *Nature,* **394,** 661–664.