International Journal of Remote Sensing
Vol. 31, No. 4, 20 February 2010, 867–880

Review Article

Uncertainty analysis of remote sensing optical sensor data: guiding principles to achieve metrological consistency

R. V. DATLA†, R. KESSEL*†, A. W. SMITH†, R. N. KACKER†
and D. B. POLLOCK‡

†National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

‡Department of Electrical and Computer Engineering, University of Alabama, 301 Sparkman Drive, OB 400, Huntsville, AL 35899, USA

(*Received 31 July 2008; in final form 16 March 2009*)

Climate change monitoring requires decades-long time-series radiometric measurements using multiple optical sensors in multiple platforms covering the globe. The problem of achieving traceability to SI units for these measurements is discussed. A major challenge is to determine the result of a measurement and its associated uncertainty using various calibration and validation processes. These processes are plagued by systematic (non-statistical) uncertainties that are not well understood. In particular, different, but in principle equivalent, SI traceable measurements may differ by more than would be expected from the uncertainties associated with the individual measurements. We propose a methodology based on the International Organization for Standardization (ISO) Guide to the Expression of Uncertainty in Measurement (GUM) for the analysis of uncertainties in such measurements along with consistency checking. This allows the measurement result and its associated uncertainty to evolve as new knowledge is gained from additional experiments, and it promotes greater caution in drawing conclusions in view of the sparse measurements. We use data from ongoing total solar irradiance measurements from various instruments in orbit to illustrate the principles.

1. Introduction

A major challenge for the international remote sensing community is to provide data that can be used to monitor climate change and to quantify changes whether natural or human induced. In 2005 the Group on Earth Observations (GEO), a voluntary partnership of governments and international organizations, established the Global Earth Observation System of Systems (GEOSS 2007). The main aim of GEOSS is to realize comprehensive and accurate Earth observations for GEO. The international remote sensing community is stated to play an important role in providing radiometric data to satisfy the requirements for climate change monitoring. Plans are under way to connect and share data from many satellite platforms around the world. In such an endeavour, data accuracy assessment tools are to be uniformly adopted and

*Corresponding author. Email: ruediger.kessel@nist.gov

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160902897882

implemented. As such, the traceability of all remote sensing measurements to internationally accepted standards and units (SI) becomes very important. The SI units are adopted and recommended by the General Conference of Weights and Measures (CGPM) and maintained across the world through the National Metrology Institutes (NMIs). SI traceability is defined as the property of a measurement result related to stated references in SI units through an unbroken chain of comparisons all having stated uncertainties (BIPM 2004) and 'it is essential that they (the measuring instruments) should be periodically calibrated against more accurate standards'. However, in the arena of remote sensing data, there is no uniformity in uncertainty analysis, especially in establishing an unbroken chain of comparisons for SI traceability so far. The major reason is that the satellite sensors upon launching into orbit cannot be recalibrated because of a lack of proven onboard or on-orbit SI traceable standards of the required accuracy for calibrations. Furthermore, it is impractical to recalibrate periodically any of the instruments against standards on Earth once they are in orbit. Often procedures adopted to resolve the calibration issues require assumptions in formulating the measurement equations that are fraught with unknown uncertainties. In this paper, we address the following questions: what is the methodology for evaluating the common reference value (CRV) of all measurements and its uncertainty when different sensors in different platforms provide the data? What metrological principles would best be applied in assigning a consistent value and associated measurement uncertainty to the CRV? What are the guidelines for expressing uncertainty in time series measurements that span decades over multiple sensor lifetimes and/or different sensors in different satellite platforms?

In §2, the background literature and definitions are reviewed, focusing on the problems in achieving the SI traceability for on-orbit sensor radiometric data. In §3, a methodology is proposed for the evaluation of uncertainties and their propagation based on the metrological principles and the International Organization for Standardization (ISO) Guide to the Expression of Uncertainty in Measurement (GUM). In §§4 and 5 the methodology is applied to the problem of arriving at the CRV and an estimate of its uncertainty for solar constant measurement data from different sensors.

2. Background and definitions

Radiance and reflectance are the basic quantities measured by Earth-viewing optical sensors in space and the data on various climate variables are deduced from these measurements by applicable physics models. There have been two workshops on climate change monitoring that dealt with the uncertainty requirements for these measurements and strategies to achieve those requirements (Ohring *et al.* 2002, 2007). The requirements deal with accuracy and stability as the two components of uncertainty in the sensor data. Accuracy refers to the closeness of agreement between a measured value and a true value (ISO 2008). The requirement on accuracy can be translated directly into a requirement on the standard uncertainty of the combined result. Stability is the ability of the instrument to maintain its metrological characteristics invariant with time (ISO 2008). In the short term most sensors are stable but eventually they degrade and drift significantly over the lifetime of the sensor. To monitor the sensor responsivity drift and account for the degradation that may be spectrally dependent is a daunting task, especially to meet the climate data quality requirements (measurement uncertainty).

Table 1. Language of GUM for statistical analysis of data.

Quantity	Symbols and definitions
Measurand	Y, state-of-knowledge distribution of the quantity subject to the measurement
Result of measurement	y, best estimate (expected value) of Y
Types of uncertainty evaluation	Type A (statistical) and Type B (non-statistical)
Standard uncertainty	$u(y)$
Expanded uncertainty	$U = k \times u(y)$ defines interval $[y \pm k \times u(y)]$, which includes a large fraction of values that could reasonably be attributed to the measurand. Here k is the coverage factor. Fraction included will be 95% for $k = 2$ or 99% for $k = 3$ for a normal distribution
Bias	Constant deviation, offset, or scaling in a result of measurement
Consistency of results	Absence of significant difference between results
Common Reference Value (CRV)	Result of measurement when different independent SI traceable measurements are combined (defined in this paper)

The language pertinent to this paper is based on the ISO GUM (ISO 1995) and is given in table 1. It should be noted that the uncertainties due to systematic biases are all grouped as uncertainty evaluations of Type B.

Absolute cryogenic radiometers are used in NMIs to achieve the lowest uncertainty measurements of optical power as they measure it directly in terms of the electrical watt, which is an SI unit that can be measured routinely to low uncertainties. The optical power measurements may be used to measure irradiance with the use of precision apertures that define the amount of power collected. Provided that the aperture area is known and that systematic effects such as diffraction and scattering of light are taken into account, the uncertainty in the irradiance measurement can, in principle, be as low as the underlying power measurement. Low-uncertainty radiance responsivity measurements are in turn made using special facilities that generate Lambertian (uniform) sources using spectrally tuneable lasers (Brown *et al.* 2000). However, the use of such techniques in space is an expensive proposition and so far most Earth-observing sensors that measure radiance have calibrations that are made traceable to SI units through pre-launch calibration activities and post-launch monitoring to account for any changes in sensor responsivity. One of the strong recommendations of the recent workshops (Ohring *et al.* 2002, 2007) is to have all sensor data SI traceable. This will allow uniform implementation of algorithms to account for sensor degradation utilizing intercomparisons with other satellites and independent ground and air-borne measurements. However, intercomparisons by themselves will not lead to a correction of the individual results but may expose the need to increase uncertainties unless there is a way to reconcile the observed differences between the instruments. Thus, intercomparisons with benchmark sensors in space would be the ideal choice to reduce growth of uncertainty. Benchmark sensors, such as the Atmospheric Infrared Sounder (AIRS), the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), are those that have undergone pre-launch calibrations with SI traceable standards and use on-board and in-orbit calibration strategies to further improve accuracy and maintain SI traceability.

Even benchmark sensors may not agree with each other when intercompared. Pollock *et al.* (2003) pointed out that the use of a single SI measurement for SI traceability may not be valid because of possibly unaccounted uncertainties in the methodology or in the measurement itself. They advocate the use of data from at least another independent SI measurement to validate the former SI traceable measurement. They discussed at length the problems that can arise using a single SI measurement for SI traceability using as an example the solar irradiance measurements made from independent SI traceable sensor platforms.

As part of a National Research Council (NRC) study of this problem, one recommendation of the space studies board is to perhaps do away with SI traceability and change the calibration paradigm from radiance to reflectance (Committee on Earth Studies 2000) for the solar reflective region of the spectrum. Reflectance is a relative, not an absolute, calibration and as such does not need reference to SI units. However, reflectance can be converted to SI units if solar irradiance measurements could be pinned down to the required accuracy. Therefore, reflectivity may be another choice for the calibration in this spectral region, but there is considerable doubt about the stability of the on-board diffusers used as reference standards. However, the use of the Moon as a reference standard is being actively considered as the Moon is very stable in this regard (Kieffer 1997). In any case, the intercomparisons of different sensors using this method may also disagree with each other, and the evaluation of uncertainty in the calibration, especially for generating and integrating long time-series climate data sets from different sensors, poses the same problems.

To address this issue of SI traceability, metrological principles for designing experiments and analysing the uncertainty in measurements would be of great help. The main premise of this approach is that the value of the parameter being measured is known only to the extent of the knowledge gained by the measured data. Unknown systematic uncertainties are to be investigated and they can only be eliminated or accounted for by improving the knowledge base. One way to achieve this is by using multiple measurements made through independent experiments. The CRV with an associated uncertainty can then be obtained using that data base. Here expert judgement plays a role. In other words, the CRV and its associated uncertainty will improve in accuracy as more independent experiments are performed and improved methods are used. However, all data are considered valuable and should be part of the data base unless proven otherwise. Application of this principle for the creation of climate data sets on various parameters would lead to more realistic uncertainties and contribute immensely in the design of experiments to improve the accuracy of climate measurements. The general principles of this metrological model of analysing measurements and evaluating uncertainties are discussed in the next section.

3. Uncertainty analysis of time series data

In remote sensing, especially for weather and climate monitoring, the data are obtained from satellite sensors in space as time series. The task is then to combine the results from different satellites for a point in time into one value with an associated uncertainty by propagating all knowledge about the measurement situation and the available data. For the discussion we use the total solar irradiance (TSI) data, which can be treated as a simple time series; the concept can be extended to more complex data as well.

At a particular point in time, there may be one, two, three or more satellite sensors making measurements of a common measurand Y (quantity to be measured; ISO 2008). We consolidate the results so that we have estimates for the same measurand with associated uncertainties at a number of specific points in time. Each of the n satellite sensors provides a consolidated result x_i with standard uncertainty $u(x_i)$, giving n results $x_1, \ldots, x_n$, with standard uncertainties $u(x_1), \ldots, u(x_n)$. The problem then is how to determine the best estimate of y and standard uncertainty $u(y)$ for the state-of-knowledge distribution of the CRV Y at that particular point in time.

The first step is to choose a model to combine the values. A generally accepted method to do this is the weighted mean

$$Y = \frac{\sum_{i=1}^{n} b_i \cdot X_i}{\sum_{i=1}^{n} b_i} \tag{1}$$

where Y is a quantity describing the state of knowledge about the CRV, and X_i are quantities describing the state of knowledge about the individual satellite results. Each result x_i and standard uncertainty $u(x_i)$ are regarded as the expected value and standard deviation of a state of knowledge distribution for X_i, respectively. The constant weighting factors, b_i, have no uncertainty. As we have no basis to assign different weights, we choose all weighting factors to be one ($b_i = 1$). Then equation (1) reduces to the arithmetic mean:

$$Y = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \tag{2}$$

Equation (2) gives the best estimate of the common reference value as $y = (1/n) \sum x_i$. If we regard all X_i as statistically independent then we can calculate the uncertainty associated with the best estimate of the CRV (y) following the standard GUM approach as

$$u^2(y) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} u^2(x_i) \tag{3}$$

This gives us a method for calculating y and $u(y)$ for the specific points in time.

The next question is whether equations (2) and (3) give us a reasonable uncertainty and whether we used all available information about the measurement situation. Suppose x_i and $u(x_i)$ represent our state of knowledge from each sensor, then the only information we have not used so far is that all sensors were measuring the same measurand at a specific point in time. If we assume that the sensors are perfect, we would expect all results from all sensors for a specific point in time to be equal, that is:

$$y = x_1 = \ldots = x_n \tag{4}$$

For real sensors we cannot expect to realize this ideal equality. However, unless something is amiss, we should expect the results to be consistent. In particular, we would expect the results $x_1, \ldots, x_n$ to be consistent with the best estimate of the CRV y. In this context consistency means that there is no significant difference between y and any of the results $x_1, \ldots, x_n$. Let us define a state of knowledge variable:

$$E_i = X_i - Y \tag{5}$$

where E_i is the difference between the CRV Y and the sensor result X_i for $i = 1, 2, \ldots, n$. Let us use the symbols ε_i and $u(\varepsilon_i)$ for the expected value and standard uncertainty of E_i, respectively. We can now derive a consistency criterion from the state-of-knowledge distribution of E_i in equation (5). We can say that the difference $\varepsilon_i = x_i - y$ is not significantly different from zero if its absolute value is smaller than some chosen multiple k of the standard uncertainty $u(\varepsilon_i)$ associated with the difference, that is:

$$|\varepsilon_i| \leq k \cdot u(\varepsilon_i) \tag{6}$$

where the standard uncertainty $u(\varepsilon_i)$ associated with ε_i is

$$u^2(\varepsilon_i) = \frac{(n-1)^2}{n^2} \cdot u^2(x_i) + \frac{1}{n^2} \cdot \sum_{j \neq i} u^2(x_j) \tag{7}$$

if all X_i are statistically independent.

The criterion defined by equation (6) shows whether the differences between the results $x_1, \ldots, x_n$ and the best estimate of the CRV y agree with the uncertainties $u(x_1), \ldots, u(x_n)$ and $u(y)$. Although this criterion (equation (6)) looks similar to the statistical t-test, the interpretation is very different. Instead of a confidence interval, an expanded uncertainty interval representing the state of knowledge about E_i is constructed from the expected value and the associated standard uncertainty. If E_i were normally distributed, then a coverage factor $k = 2$ would represent an interval $[\varepsilon_i - k\ u(\varepsilon_i),\ \varepsilon_i + k\ u(\varepsilon_i)]$, which contains about 95% of the possible values that are compatible with the knowledge represented by E_i. Given k, simple interval logic can be used to compare this interval with zero. A difference $\varepsilon_i = x_i - y$ is significant if the interval does not include the value zero. As a consequence, if the assumption of insignificant difference is not compatible with this interval, then the consistency check fails.

If the criterion in equation (6) is not satisfied for one or more of the results $x_1, \ldots, x_n$ then these results are inconsistent and as a consequence it is illogical to combine these results. In general, inconsistencies should be resolved by investigating the details of the measurement process. In remote sensing measurements this is particularly challenging because we have little, if any, opportunity to retrieve satellite-based sensors for further investigation. In the absence of more information, we have no other choice but to empirically enlarge the uncertainties $u(x_1), \ldots, u(x_n)$ associated with the individual results $x_1, \ldots, x_n$ if we want consistent results. The enlargement of uncertainties is tantamount to recognition that some as yet not understood effects are causing the divergence of the multiple results of the same quantity. Therefore, we modify equation (2) to account for additional deviations related to the results:

$$Y = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i + \delta X_i) \tag{8}$$

where δX_i is a state of knowledge variable representing the possible deviation of the sensor result X_i from the value of the measurand.

We also need to modify equations (5) and (7) to incorporate the additional quantities δX_i; thus,

$$E_i = (X_i + \delta X_i) - Y \tag{9}$$

The standard uncertainty $u(\varepsilon_i)$ of E_i determined from equation (9) is

$$u^2(\varepsilon_i) = \frac{1}{n^2} \cdot \left[(n-1)^2 u^2(\delta x_i) + \sum_{j \neq i} u^2(\delta x_j) + (n-1)^2 u^2(x_i) + \sum_{j \neq i} u^2(x_j) \right] \tag{10}$$

In case we observe inconsistencies between the results, either the best estimates or their associated uncertainties may prove to be not completely reliable. As we have no basis to correct any of the results $x_1, \ldots, x_n$, the expected value $E(\delta X_i)$ of each δX_i is set as zero. Similarly, we have no basis to judge the uncertainty of the results $x_1, \ldots, x_n$. We therefore assign the same standard uncertainty to all the deviations δX_i.

$$\left. \begin{array}{l} \delta x_i = E(\delta X_i) = 0 \\ u(\delta x_i) = u(\delta x) \end{array} \right|_{i=1 \ldots n} \tag{11}$$

Combining equations (10) and (11) we have:

$$u(\varepsilon_i) = \frac{1}{n} \cdot \sqrt{(n^2 - n) \cdot u^2(\delta x) + (n-1)^2 u^2(x_i) + \sum_{j \neq i} u^2(x_j)} \tag{12}$$

We can use expert judgement to find a reasonable uncertainty $u(\delta x_i)$. The minimum value for $u(\delta x)$ to obtain consistency is:

$$u(\delta x) \geq \max \left[\left. \sqrt{\frac{n}{n-1} \cdot \left[\left(\frac{x_i - y}{k} \right)^2 - \frac{1}{n^2} \cdot \left((n-1)^2 u^2(x_i) + \sum_{j \neq i} u^2(x_j) \right) \right]} \right|_{j=1 \ldots n} \right] \tag{13}$$

(Kessel *et al.* 2008). In the absence of additional knowledge we can use the lower bound uncertainty given by equation (13).

In cases where the original uncertainty of the contributing results varies significantly, we can use weighting factors based on the given uncertainties ($b_i = 1/u^2(x_i)$ in equation (1)). However, in cases such as the total solar irradiance data that are significantly inconsistent, the added $u(\delta x)$ dominates the uncertainties associated with the best estimate of the CRV y and the weighting factors will iterate to one. Therefore, we choose all weighting factors to be one from the start. In the following section, we analyse data from satellite sensors measuring the TSI to illustrate these principles.

4. Application to remote sensing data for TSI

As an example, the above-atmosphere TSI data from several satellites over the past several decades are examined and analysed. The available data are shown in figure 1 (NGDC 2007). The instruments that provided the data are active cavity radiometers that measure optical power in equivalent units of electrical power in watts and thus provide inherently SI traceable measurements. The data presented are corrected for known systematic biases such as in-orbit sensor degradation and are normalized to a sensor–solar distance of one astronomical unit (1 AU). As we can see, the difference between the data from each sensor is larger than the estimated uncertainty of each instrument. Therefore, the data indicate the presence of unknown systematic biases that should be investigated and corrected to improve the CRV. In figure 1, the

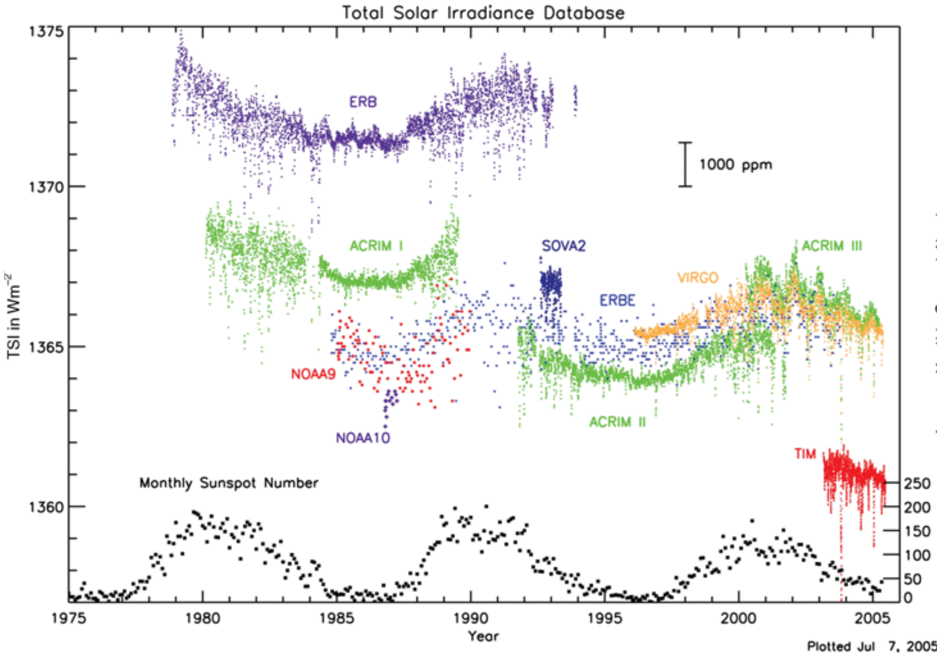

Figure 1. Top-of-atmosphere Total Solar Irradiance (TSI) data measured by various satellites from 1975 to 2005.

National Aeronautics and Space Administration (NASA) NIMBUS satellite provided the data from an on-board active cavity radiometer designated ERB (Earth Radiation Budget) for the period 1978 to 1993. The data from 1980 to 1989 shown as ACRIM 1 (Active Cavity Radiometer Irradiance Monitor) were obtained from the active cavity radiometer onboard the Solar Maximum Mission (SMM) satellite. The data labelled NOAA9 and NOAA10 are from ACRIM-type radiometers from the National Oceanic and Atmospheric Administration (NOAA) polar satellites. The data shown from 1984 to the present designated as ERBE are from the NASA Earth Radiation Budget Satellite (ERBS). The data designated ACRIM II and ACRIM III are from Upper Atmospheric Research Satellite (UARS) and ACRIMSAT missions, respectively. The data designated as VIRGO are from two types of absolute radiometers on the European Space Agency (ESA)/NASA Solar and Heliospheric Observatory (SOHO), which are called DIARAD and PMO6V. Finally, the data designated TIM are from the Solar Radiation and Climate Experiment (SORCE).

Figure 2 shows the data from different radiometers for the overlapping time period from 2003 to the middle of 2007 on an expanded scale for illustrating the principles of determining the CRV. The data are from the radiometers ACRIM III, DIADRAD, PMO6V and TIM. To compare observations of the Sun made at different times, the data were interpolated onto a common time-scale and averaged over a period of approximately 15 days. A simple upper-bound estimate of the additional uncertainty associated with sampling errors coupled with the short-term solar variability was calculated and summed in quadrature (see uncertainty bars of figure 2) with the stated uncertainties as follows. After interpolation, the standard deviation of the difference

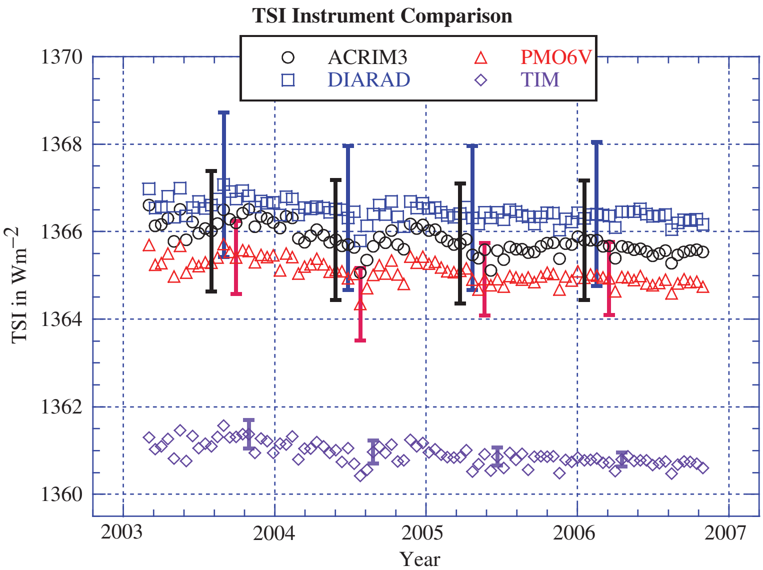

Figure 2. Available Total Solar Irradiance (TSI) data plotted for the overlapping time series of 2003 to 2007 from four satellite sensors. Uncertainty bars represent the combined standard uncertainty.

between each sensor and a reference (arbitrarily chosen as ACRIM) was computed. For TIM, PMO6 and DIARAD the standard deviation of the corresponding difference was taken as the additional uncertainty. For the ACRIM sensor the minimum of the other three standard deviations was used. The added component of uncertainty was small compared with the differences between sensors and, with the exception of the TIM sensor, was small compared to the stated instrument uncertainty.

To illustrate the calculations we use the following results from the four satellite sensors for a particular point in time in figure 2 (see figure 3).

$$x_1 = 1366.6\,\mathrm{W\,m^{-2}}, u(x_1) = 1.4\,\mathrm{W\,m^{-2}}$$
$$x_2 = 1367.0\,\mathrm{W\,m^{-2}}, u(x_2) = 1.6\,\mathrm{W\,m^{-2}}$$
$$x_3 = 1365.70\,\mathrm{W\,m^{-2}}, u(x_3) = 0.82\,\mathrm{W\,m^{-2}}$$
$$x_4 = 1361.31\,\mathrm{W\,m^{-2}}, u(x_4) = 0.21\,\mathrm{W\,m^{-2}}$$

The calculation of the arithmetic mean (equation (2)) gives $y = 1365.15\,\mathrm{W\,m^{-2}}$ with an associated standard uncertainty $u(y) = 0.57\,\mathrm{W\,m^{-2}}$. The calculation of the consistency check (equations (5) and (7)) leads to:

$$|\varepsilon_1| = 1.4\,\mathrm{Wm^{-2}}, U(\varepsilon_1) = 2.3\,\mathrm{Wm^{-2}}, k_{p95} = 2, \text{check passed}$$
$$|\varepsilon_2| = 1.8\,\mathrm{Wm^{-2}}, U(\varepsilon_2) = 2.5\,\mathrm{Wm^{-2}}, k_{p95} = 2, \text{check passed}$$
$$|\varepsilon_3| = 0.5\,\mathrm{Wm^{-2}}, U(\varepsilon_3) = 1.6\,\mathrm{Wm^{-2}}, k_{p95} = 2, \text{check passed}$$
$$|\varepsilon_4| = 3.8\,\mathrm{Wm^{-2}}, U(\varepsilon_4) = 1.2\,\mathrm{Wm^{-2}}, k_{p95} = 2, \text{check failed}$$

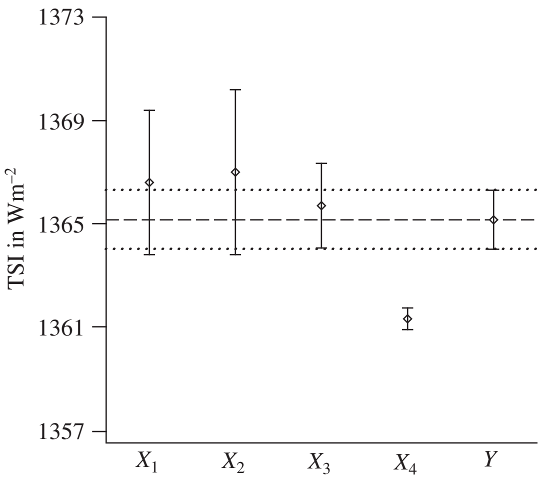

Figure 3. Total Solar Irradiance (TSI) for each satellite sensor data X_i and mean value Y without adding uncertainty. Uncertainty bars represent the expanded uncertainty ($k = 2$).

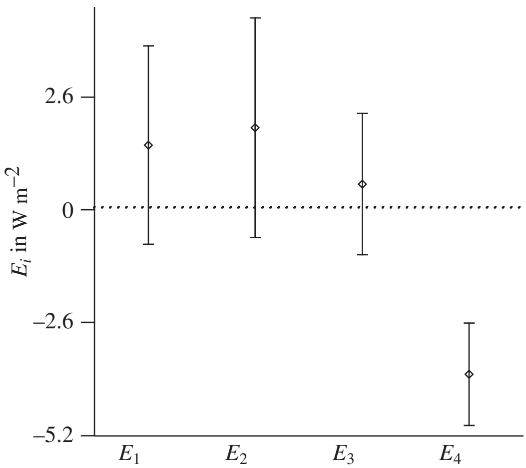

Figure 4. Difference between satellite sensor to sensor data and mean value $E_i = X_i - Y$ based on the original values without adding uncertainty. Uncertainty bars represent the expanded uncertainty ($k = 2$).

The consistency check can easily be evaluated based on figure 4. It failed for at least one contributing result and therefore δX_i is added to each result X_i with an expected value zero and standard uncertainty $u(\delta x)$.

The lower bound can be calculated with equation (13) to give $u(\delta x) = 2.1$ W m^{-2}. Based on the whole time series, the uncertainty for the possible deviation of the sensors was set to $u(\delta x) = 2.2$ W m^{-2}, which is the smallest value with two significant digits for which all values of the time series became consistent.

Including the new value for $u(\delta x)$ in equation (2) leads to a larger uncertainty for the contributing value from the sensors (see figure 5) and a final result $y = 1365.2$ W m^{-2} and $u(y) = 1.2$ W m^{-2}.

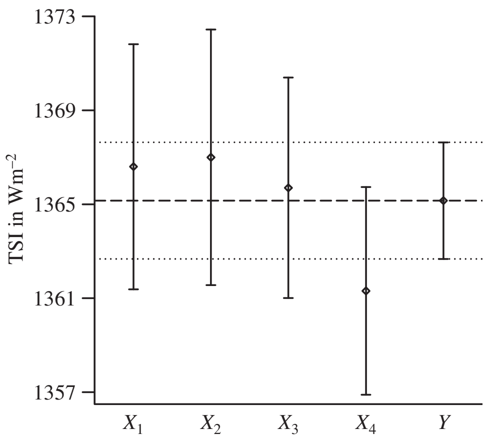

Figure 5. Total Solar Irradiance (TSI) for each satellite sensor data X_i and mean value Y including additional uncertainty $u(\delta x)$ to ensure consistency. Uncertainty bars represent the expanded uncertainty ($k = 2$).

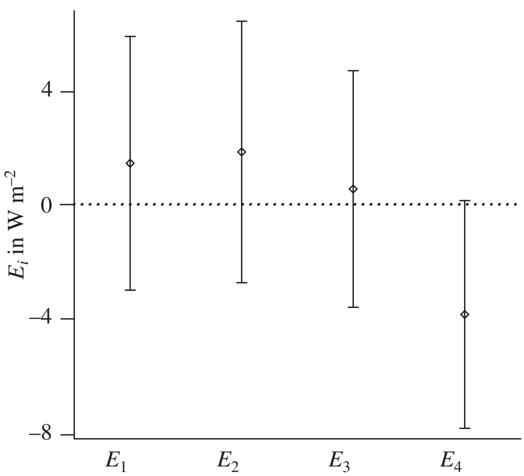

Figure 6. Difference between satellite sensor data and mean value $E_i = (X_i + \delta X) - Y$ including additional uncertainty to ensure consistency. Uncertainty bars represent the expanded uncertainty ($k = 2$).

The recalculation of the consistency check leads to (see figure 6):

$$|\varepsilon_1| = 1.4\,\mathrm{W\,m^{-2}}, U(\varepsilon_1) = 4.4\,\mathrm{W\,m^{-2}}, k_{\mathrm{p95}} = 2, \text{check passed}$$
$$|\varepsilon_2| = 1.8\,\mathrm{W\,m^{-2}}, U(\varepsilon_2) = 4.6\,\mathrm{W\,m^{-2}}, k_{\mathrm{p95}} = 2, \text{check passed}$$
$$|\varepsilon_3| = 0.5\,\mathrm{W\,m^{-2}}, U(\varepsilon_3) = 4.1\,\mathrm{W\,m^{-2}}, k_{\mathrm{p95}} = 2, \text{check passed}$$
$$|\varepsilon_4| = 3.8\,\mathrm{W\,m^{-2}}, U(\varepsilon_4) = 4.0\,\mathrm{W\,m^{-2}}, k_{\mathrm{p95}} = 2, \text{check passed}$$

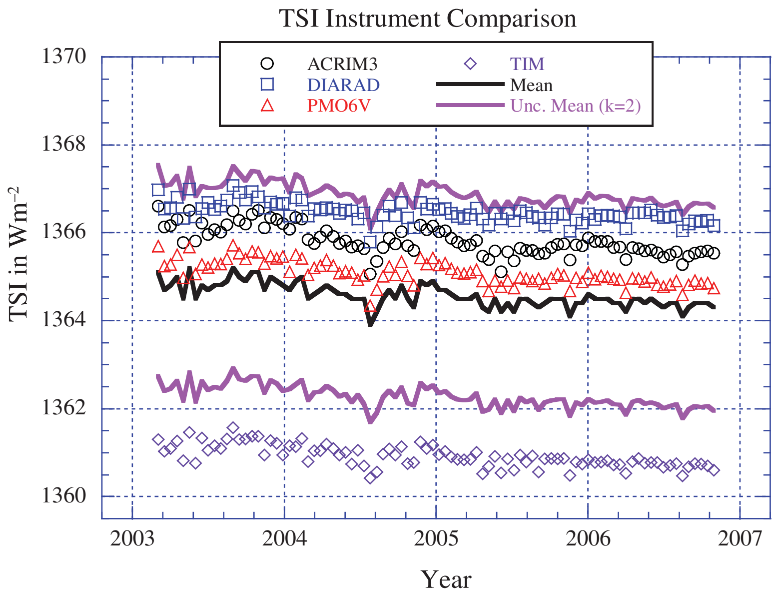

Figure 7. The CRV for the time series after adding an additional uncertainty for each sensor to make the data consistent. Symbols are the data collected for a given instrument. The solid lines represent the arithmetic mean and its expanded uncertainty ($k = 2$).

The new uncertainty of the arithmetic mean is consistent with all contributing results. The calculation was repeated for all data points of the time series. Figure 7 shows the results after adding an additional uncertainty to the data at each point in time. After adding an additional δX_i component, the uncertainties for all data points are very similar, varying from 1.24 to 1.26 W m^{-2}. In this case it is possible to use the same value for the uncertainty of the whole data series. We chose the largest uncertainty value in the series rounded to two digits as the final value for all data points and this leads to $u(y_j) = 1.3$ W m^{-2} and $U(y_j) = 2.5$ W m^{-2}, $k_{p95} = 2$ for $j = 1, 2, \ldots, m$ if m is the number of results in the data series. The maximum difference between the individually calculated uncertainty and the chosen value for the whole data series is about 4.8% of the value.

The resulting uncertainty bands surrounding the mean are significantly larger than would be the case without the added uncertainty component. From 2003 to 2007 the solar irradiance shows a steady decrease from 1365 to 1364 W m^{-2} with a standard uncertainty of 1.3 W m^{-2} in the absolute value at any given time. This barely meets the climate change science programme accuracy goal of 1.5 W m^{-2} (Ohring *et al.* 2002). Future satellite missions should aim to achieve much higher accuracies as that is the best way to ensure that small changes in solar output can be tracked over the long term. It also provides an independent means to validate the corrections that account for instrument drift, which is thought to be well understood.

5. Discussion

To look into the possible causes for the discrepancy between various TSI sensors, in 2006 NASA held a workshop at the National Institute of Standards and Technology

(NIST) and brought all the principal scientists of the TSI measurements together for discussions on the topic. The workshop report identified possible causes such as unaccounted diffraction losses and problems of measuring aperture diameters accurately that are to be investigated further (Butler *et al.* 2007). The report identified plans to intercompare replicas of the radiometers with NIST SI traceable standards in the irradiance mode to quantify unknown biases in the calibration of these radiometers. Such an undertaking will reveal unknown biases that can be corrected in the sensor data. Then the analysis presented in this paper can be repeated to determine the new CRV time series and associated uncertainty using the more realistic assumption that biases are not negligible. Further investigation into any remaining discrepancies and iteration of this process should move the measurements closer to the climate change accuracy requirements.

The sensor data analysed in this paper were for overlapping periods of time series to determine the CRV and its uncertainty. This methodology is applicable to the intercalibration of satellite sensors based on the benchmark SI traceable CRV time series for the overlapping sensor data. For the use of the Simultaneous Nadir Overpass (SNO) technique (Cao *et al.* 2005), this methodology can be adopted to generate common reference values and uncertainties between non-SI traceable sensors wherever possible and then transfer calibration from an SI traceable overpass to all the sensors in the intercomparison. For example, NASA research satellites such as MODIS and SeaWiFS provide benchmark CRV time series that can transfer calibration to NOAA and EUMETSAT operational satellites that are by themselves intercalibrated by the SNO technique. The main guiding principle is that all sensor data are to be analysed with equal weight unless proven scientifically otherwise. If scientific judgement indicates the presence of unknown systematic uncertainties, research should be undertaken to identify and correct the data and improve the CRV and its uncertainty. This can be done iteratively until the required levels of uncertainty bounds are achieved on the CRV. However, long time-series will have gaps in overlapping sensors. The extension of the methodology discussed in this paper on how to bridge the gaps if at least one sensor provides data without interruption will be the subject of a future publication.

6. Summary

Satellite time-series data require at least two independent methods to determine the CRV and its uncertainty. A CRV and the associated uncertainty can be obtained by simply combining the data from each sensor if the analysis shows few outliers. However, if the sensor data differ from one sensor to another by more than the individual sensor uncertainties, adding an uncertain influence to the sensor data using the methodology based on the ISO GUM described in this paper will be more appropriate. This approach quantifies the uncertainty more realistically and helps to investigate and seek corrections and improve the accuracy of the data. As an illustration of this methodology, the TSI data from different satellite sensor data are analysed. It is shown quantitatively that unknown systematic uncertainties contribute significantly to the overall uncertainty of the SI traceable CRV time series. Current research undertaken with NASA, NIST and other agencies to identify and correct the unknown systematic uncertainties will help to improve the accuracy of the data, and the methodology given in this paper can be repeated to evaluate corresponding CRV time series and its uncertainty as new knowledge about the possible deviations or data from future missions become available. This allows for continuous improvement of the measurements and the interpretation of the sensor data. A CRV and its uncertainty is a dynamic quantity that represents the current state of knowledge at the time it is calculated.

References

BIPM, 2004, Traceability. International Bureau of Weights and Measures/Bureau International des Poids et Mesures (BIPM). Available online at: www.bipm.org/en/convention/wmd/2004/traceability.html.

Brown, S.W., Eppeldauer, G.P. and Lykke, K.R., 2000, NIST facility for spectral irradiance and radiance responsivity calibrations with uniform sources. *Metrologia*, **37**, pp. 531–534.

Butler, J.J., Johnson, B.C., Rice, J.P., Shirley, E.L. and Barnes, R.A., 2007, Sources of difference in on-orbital total solar irradiance measurements and description of a proposed laboratory intercomparison. *Journal of Research, National Institute of Standards and Technology*, **113**, pp. 187–203.

Cao, C., Xu, H., Sullivan, J., McMillin, L., Ciren, P. and Hou, Y., 2005, Intersatellite radiance biases for the high resolution infrared radiation sounders (HIRS) onboard NOAA-15, -16, and -17 from simultaneous nadir observations. *Journal of Atmospheric and Oceanic Technology*, **22**, pp. 381–395.

Committee on Earth Studies, 2000, *Issues in the Integration of Research and Operational Satellite Systems for Climate Research: Part II. Implementation*, Appendix C, p. 78 (Washington, D.C.: Space Studies Board National Research Council).

GEOSS, 2007, *Global Earth Observation System of Systems*. Available online at: http://www.epa.gov/geoss/.

ISO, 1995, *Guide to the Expression of Uncertainty in Measurement*, 2nd edn (Geneva, Switzerland: International Organization for Standardization).

ISO, 2008, *International Vocabulary of Basic and General Terms in Metrology (VIM)*, 3rd edn (Geneva, Switzerland: International Organization for Standardization).

Kessel, R., Berglund, M. and Wellum, R., 2008, Application of consistency checking to evaluation of uncertainty in multiple replicate measurements. *Accreditation and Quality Assurance*, **13**, pp. 293–298.

Kieffer, H.H., 1997, Photometric stability of the lunar surface. *Icarus*, **130**, pp. 323–327.

NGDC, 2007, *Daily Total Solar Irradiance*. National Geophysical Data Center/ Solar-Terrestrial Physics (NGDC/STP). Available online at: http://www.ngdc.noaa.gov/stp/SOLAR/IRRADIANCE/irrad.html.

Ohring, G., Tansock, J., Emery, W., Butler, J., Flynn, L., Weng, F., St. Germain, K., Wielicki, B., Cao, C., Goldberg, M., Xiong, J., Fraser, G., Kunkee, D., Winker, D., Miller, L., Ungar, S., Tobin, D., Anderson, J.G., Pollock, D., Shipley, S., Thurgood, A., Kopp, G., Ardanuy, P. and Stone, T., 2007, Achieving satellite instrument calibration for climate change. *Eos, Transactions American Geophysical Union*, **88**, p. 136.

Ohring, G., Wielicki, B., Spencer, R., Emery, W. and Datla, R., 2002, *Satellite Instrument Calibration for Measuring Global Climate Change*. NIST Internal Report 7047 (Gaithersburg, MD: National Institute of Standards and Technology).

Pollock, D.B., Murdock, T.L. and Datla, R.U., 2003, Data uncertainty traced to SI units. Results reported in the international system of units. *International Journal of Remote Sensing*, **24**, pp. 225–235.