Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Atmospheric Chemistry and Physics
Discussions

How long do satellites need to overlap? Evaluation of climate data stability from overlapping satellite records

Elizabeth C. Weatherhead[1], Jerald Harder[2], Eduardo A. Araujo-Pradere[3], Jason M. English[4], Lawrence E. Flynn[5], Stacey Frith[6], Jeffrey K Lazo[7], Peter Pilewskie[2], Mark Weber[8], Thomas N. Woods[2]

[1] University of Colorado, Boulder, Colorado, USA
[2] Laboratory for Atmosphere and Space Physics, University of Colorado, Boulder, Colorado, USA
[3] School of Science, Miami Dade College, Miami, Florida, USA
[4] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA
[5] NOAA, NESDIS, College Park, Maryland, USA
[6] NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[7] National Center for Atmospheric Research, Boulder, Colorado, USA
[8] University of Bremen FB1, Bremen, Germany

Correspondence to: Elizabeth Weatherhead (Betsy.Weatherhead@Colorado.edu)

Abstract. Sensors on satellites provide unprecedented understanding of the Earth's climate system by measuring incoming solar radiation, as well as both passive and active observations of the entire Earth with outstanding spatial and temporal coverage that would be currently impossible without satellite technology. A common challenge with satellite observations is to quantify their ability to provide well-calibrated, long-term, stable records of the parameters they measure. Ground-based intercomparisons offer some insight, while reference observations and internal calibrations give further assistance for understanding long-term stability. A valuable tool for evaluating and developing long-term records from satellites is the examination of data from overlapping satellites. Prior papers have used overlap periods to identify the offset between data from two satellites and estimate the added uncertainty to long-term records. This paper addresses the length of overlap needed to identify an offset or a drift in the offsets of data between two sensors. The results are presented for the general case of sensor overlap by using the case of overlap of the SORCE SIM and SOLSTICE solar irradiance data as an example. To achieve a 1% uncertainty in estimating the offset for these two instruments' measurement of the Mg II core (280 nm) requires approximately 5 months of overlap. For relative drift to be identified within 0.1% yr^{-1} uncertainty, the overlap for these two satellites would need to be 2.6 years. Additional overlap of satellite measurements is needed if, as is the case for solar monitoring, unexpected jumps may occur because these jumps add to the uncertainty of both offsets and drifts; the additional length of time needed to account for a single jump in the overlap data may be as large as 50% of the original overlap period in order to achieve the same desired confidence in the stability of the merged dataset. Extension of the results presented here are directly applicable to satellite Earth observations. Approaches for Earth observations may be challenged by the complexity of those observations but may also benefit from ancillary observations taken from ground-based and in situ sources. Difficult choices need to be made when monitoring approaches are considered; we outline some attempts at optimizing networks based on economic principles. The careful evaluation of monitoring overlap is important to the appropriate application of observational resources and to the usefulness of current and future observations.

Keywords: Satellite overlap, satellite monitoring, instrument intercomparison, instrument stability, climate records, trend detection, homogenization of datasets, solar spectral irradiance, ozone.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

1. Introduction

One of the fundamental requirements for construction of long-duration climate records is the ability to analytically assess the stability of over-lapping observational records so that the time series of different sensors can be combined into a single reliable record. This need is particularly valuable for satellite-borne sensors that are susceptible to a wide variety of sensitivity degradation mechanisms influenced by the space environment as well as by spacecraft, instrument, and operational considerations. Many of these influences can contribute in unexpected ways to the overall instrument stability thereby adding non-geophysical trending or structure to the combined data records compiled from multiple missions. These problems can also be exacerbated when comparing instruments with different time histories, i.e. the comparison of two sensors – one during its early-orbit phase with one that has been in space for an extended length of time. For overlapping spaced-based observations, even with reliable on-board calibration and instrument sensitivity degradation correction schemes, time-limited intercomparison campaigns are still required to objectively identify potential systematic errors in one or both instruments. A variety of techniques exist for merging datasets from different sources – including two different satellites – using statistical models, physical models, and efforts at in situ calibration (Chander et al., 2013b; Peterson et al., 1998). Each technique has great strengths and can offer not just adjustments for merging of datasets, but estimates of uncertainty in long-term stability. Some techniques focus on tying data to absolute reference standards with the intent of developing traceability to reference standards (Wielicki et al., 2013; Fox et al., 2013). The development and use of these high accuracy climate benchmark instruments has been advocated for since the early 2000's and described in the NISTIR 7047 (2004) and ASIC3 (2007) workshop reports. These high accuracy instrument systems will provide two fundamental products of great value to the climate science community: 1) reliable long-term records of basic climate forcings, response, and feedback for analysis and climate model verification, and 2) in-flight calibration standards for environmental operational satellite sensors including weather satellites that do not have a rigorous pre-flight radiometric calibration requirement or the ability to perform degradation corrections on-orbit. Weber et al. (2016) addressed the issue of requirements on stability for detecting a desired long-term trend from a multiple instrument timeseries by accounting for variations in instrumental lifetime and merging biases in a Monte Carlo simulation. In both of these cases where inter-satellite calibration transfer can occur and in the concatenation of multiple generations of the benchmark satellite instruments, verification of merging different observation systems is important for validation of expected agreement. Should an offset or drift be detected, the use of that information is for the instrument scientists to determine based on all available information; without sufficient overlap there is a limit to the magnitude of offset or drift that can be detected. This paper will address techniques that can address the goal of verifying the stability of merged data records using observations from overlapping independent satellite instruments.

A particular challenge in the design of climate observing systems is how to preserve data quality and facilitate valid comparisons of observations that extend over a series of missions measuring the same geophysical quantity. A number of in-depth techniques are used by instrument scientists to understand the fundamental (Level 1) observations, including wavelength scale corrections, detector responsivity evaluation, and field-of-view sensitivity monitoring. With the regular insertion of new technology driven by interest in reducing costs and/or improving performance also comes the need to separate the effects of changes in the Earth system from effects ascribable to changes and gaps in the observing system. Credible, ongoing programs of sensor calibration and validation, sensor characterization, data continuity, and strategies for ensuring overlap across successive sensors are thus essential (NRC, 2000a). Multiple efforts describing key challenges and/or requirements have been published (Chander et al., 2013a; Fröhlich, 2009; Willson and Hudson, 1991; Willson and Mordvinov, 2003). Adams et al. (2014) revealed up to a 6% relative drift per decade between different ozone observing satellites, confounding some attempts to detect signs of ozone recovery. Rahpoe et al. (2015) show that both drifts and biases in current satellite observing systems are often large compared to the signals of interest. Developing a stable record requires a rigorous calibration and sensor characterization program and an observation approach that ensures sufficient temporal overlap to achieve accurate cross-calibration between successive sensors. The approaches presented in this paper focus on developing useful checks on the final data products (Level 2) from multiple instruments. Measuring the small changes associated with long-term global climate change from space is particularly challenging. For example, the satellite instruments must be capable of observing atmospheric and surface temperature trends as small as 0.1°C dec.$^{-1}$, ozone changes as little as 1% dec.$^{-1}$, and variations in the Sun's output as tiny as 0.1% dec.$^{-1}$ (Ohring et al., 2005).

There is a long and valuable history of efforts to attempt in-flight calibration, particularly on multi-spectral sensors (Slater et al., 1996). The Global Space-based Inter-Calibration System (GSICS) coordinates the development of tools to intercompare different Earth observing systems. Through careful analysis of spectral signals, relative

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

stability can be assessed and small problems with individual sensors can be identified. However, long-term stability and absolute calibration still present a challenge to these internal consistency methods, leading many to look to other approaches for absolute calibration, traceability, and the ability to verify stability.

NRC (2000a) highlighted the need for precise inter-satellite calibration, recommending that there should be a 1-year overlap between successive Ozone Monitoring Profiler Suite launches to allow sensor intercomparison and guarantee long-term traceability. Analogously, a 1-year overlap in observations of both solar irradiance and spectral solar irradiance is part of the summary recommendations of Ohring et al. (2005). Randel and Thompson (2011) went a step further by exploring the utility of combining the SAGE II ozone observations with tropical measurements from the SHADOZ ozonesonde network, to study interannual variability and trends. There is a long overlap period between the two datasets (1998–2005), and comparisons show excellent agreement. Not all satellite records have the benefit of such a long-term in situ dataset for intercomparison. For ground-based ozone measurements a change from one instrument type to another, tandem operations of both instruments at a given station for at least three years are recommended in order to assess that both instruments capture reasonably well the seasonal as well as interannual variability (Staehelin et al., 2003).

To improve the precision and usefulness of multi-instrument time series for identifying drifts, it is necessary to remove biases between data sources, including those resulting from (a) calibration differences; (b) spatial and temporal sampling or resolution differences; (c) changes in data processing versions; (d) inherently different spectral sensitivities; (e) different instrument types with varying inherent vertical coordinates; and (f) changes in instrument orientation or orbital characteristics. Sufficient overlap in the measurement obtained with a different generation of a given instrument can be used to determine the magnitude of measurement differences and thereby to remove biases in the data (NRC, 2000a). Multiple efforts are ongoing internationally to assure that emerging ground-based, in situ and satellite records can be useful to climate analyses, most notably the Global Space-based Inter-Calibration System (GSICS) which is a joint effort by WMO and the Coordination Group for Meteorological Satellites (CGMS) to monitor and harmonize data quality from operational weather and environmental satellites. Harmonized datasets often require adjusting for offsets, spurious drifts and instrument or location-specific problems (Salby and Callaghan, 1997; Araujo-Pradere et al., 2011; Dudok de Wit, 2011).

In this paper, we estimate the direct impact of length of overlap between satellites to the continuity of data from two overlapping satellites. We examine three separate factors that are of direct importance to the users of merged satellite datasets: the quantified offset of the two datasets, the drift between the two datasets, and the impact of sudden jumps in the data during periods of overlap. We also examine the impacts of drifts in individual records on the long-term stability of the merged satellite records. Intercomparison of satellite records cannot, in isolation, determine which of two systems is more accurate or stable. Indeed, agreement of two observing systems can occur when both are similarly inaccurate, or similarly drifting. However, intercomparisons offer valuable, independent assessment useful for developing a long-term record. For illustrative purposes, we will use the potential overlap of two instruments observing solar radiation to address these three factors. We will discuss some factors relevant to extension of these techniques to Earth observations. In the final section of this paper, we outline methods for optimizing the set of choices which are needed to create a long-term and stable climate record under a variety of constraints.

1.1 Offsets

Since 1978, there have been near-continuous space-based observations of ozone columns and stratospheric ozone profiles from a combination of missions. Temporal overlaps between these instruments have allowed detailed intercomparisons to play a key role in assessing the precision, accuracy, and long-term drift of the instruments (WMO, 1989, 2011a, 2011b; Bodeker, 2001). However, these overlaps have been somewhat serendipitous; little commitment has been made to ensure the continuity and long-term traceability of the full ground-based, in situ and satellite ozone measurements. Gaps in the satellite record are problematic for detecting long-term trends in ozone, because satellites offer the only means for global monitoring of the ozone distribution. Ground-based observations from spectrophotometers, sondes, or lidars offer only sparse spatial coverage, although they play a pivotal role in validating satellite measurements and helping verify their accuracy. However, they are inadequate for observing, for example, the evolution of polar ozone loss in the Arctic, an issue that is emerging as a major environmental concern for the decades ahead. Continuity of space-based observations of ozone, involving planned overlap between successive instruments, is essential for detecting and interpreting long-term trends.

As it was concluded in NRC (2000b), it takes a special effort to preserve the quality of data acquired with different satellite systems and sensors, so that valid comparisons can be made over an entire set of observations. There are few examples of continuous data records based on satellite measurements where data quality is consistent across changes in sensors, even when copies of the sensor design are used. Stolarski and Frith (2006) argue the need for the evaluation of uncertainties, concluding that a longer overlap period leads to more confidence in the calculated bias, and a reduced offset uncertainty. Bourassa et al. (2014) attempted to quantify interannual variability and decadal trends by combining stratospheric ozone profile measurements from different satellite systems including using the Stratospheric Aerosol and Gas Experiment (SAGE) II satellite instrument (1984–2005) with measurements from the Optical Spectrograph and InfraRed Imager System (OSIRIS) instrument on the Odin satellite (2001–present). Sensor characterization and an effective, ongoing program of sensor calibration and validation are essential in order to separate the effects of changes in the Earth system from effects owing to changes in the observing system.

Efforts at merging satellite data in the past have focused on deriving offsets to limit relative differences before combining data from different sensors into a continuous record (e.g. Wentz and Schabel, 1998; Santer et al., 2003; Smith et al., 2008; Dudok de Wit et al., 2008; Chander et al., 2013b). Bourassa et al. (2014) showed that the uncertainty in drift from a continuous record from multiple satellites is critical to long-term monitoring of the Earth. The uncertainty of merging satellite data records is a continual challenge with a variety of approaches employed including comparison to ground-based records, statistical intercomparison of satellites by latitude, time of day and season, as well as use of physical models to look for appropriate consistencies with available data.

One of the most studied issues underscoring the importance of proper treatment of multiple satellite records involves the merging of Microwave Sounding Unit temperature records. Christy et al. (1995, 1998, 2000) accurately pointed out that trends from satellite temperature records were not in agreement with other temperature records and showed a cooling of the troposphere rather than a warming. Additional work showed that a number of corrections to the satellite record could make a direct and notable difference on the trend derived from the resulting data (NRC, 2000a, 2000b; Zou and Qian, 2016). Some of the most salient lessons from this effort were summarized by Thorne et al. (2005), who concluded, among other points that, "individual adjustments will a priori retain a non-climatic signal of unknown sign and magnitude regardless of how reasonable and physically plausible the chosen homogenization approach." In this paper, we fully respect this sentiment and look to see, to what extent, some confinement of the problem may be achieved through proper overlap of independent instruments.

1.2 Drifts

There are several fundamental factors that can contribute to a drift in satellite observations including decay of instrumentation and changes in satellite orbit. Efforts are ongoing to minimize the impact of these factors, but all corrections involve assumptions and each satellite may invoke different approaches to monitor and address stability. Future plans may well involve traceable calibration to reference standards, but this will not fully address the challenge of merging two different types of derived satellite products. The merging of satellite records, at a minimum, needs to test for potential drift between the overlapping satellite records. The physical origins of instrument drift are rooted in the inability to correct changes in the components of a spacecraft observation system. Degradation of detectors, optics and vital spacecraft subsystems such as attitude and temperature control can all combine in different ways to produce long-term drifts in the instrument. Rahpoe et al. (2015) find intercomparisons of six different ozone limb measurements to drift relative to each other at a statistically significant rate, sometimes as high as 5% per decade or more. However, most drifts were statistically insignificant due to the limited length of data records generally less than 10 years. In the case of solar viewing instruments, BenMoussa et al. (2013) discuss in detail causes and effects of degradation in a variety of different instruments that span nearly two decades and cover a broad wavelength range. They conclude that there is no single best method to correct and monitor degradation and the correction schemes for overlapping missions are likely to be very different depending on the instrument hardware selection. Earth-observing instruments are not immune to the same problems apparent in Sun-viewing instruments. An important example of this is well documented in the efforts to correct drift in the CIRES instrument (Cloud and Earth's Radiant Energy System, see Loeb et al., 2016 and references therein). In this report, long-term stability was linked to loss of optical transmission due to UV exposure and molecular contamination, very similar to the mechanisms discussed in BenMoussa et al. (2013). Fruit et al. (2002) have addressed the effects of energetic particle on glass transmission, but inhibiting and characterizing carbonization of optical surfaces remains a steadfast and unsolved problem.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

1.3 Jumps

The effect of sudden jumps can be large and can be directly improved by an increase in planned overlap of satellites, when possible; three satellites can make the ability to identify and understand a sudden jump significantly easier, but is often beyond current monitoring approaches. Hurrell and Trenberth (1997) point to the importance of two small, discrete, downward shifts in merged satellite records that dominated the trend results for tropospheric temperature records. A careful intercomparison of sudden shifts in temperature and humidity profiles from sonde records (Free et al., 2002) show that these efforts can disagree profoundly with each other, resulting in changes to observed trends by between 35% and 80%. Jumps can occur for a variety of reasons related to instrument changes. Depending on the physical source of the jumps, the effects can last from less than a few hours to multiple years. We focus on the ability to detect and understand the jumps that last more than a few months, as they may be the interruptions that can cause the most serious damage to long-term records, particularly those used in the context of climate research.

1.4 Planning for needed homogeneity

The user communities, including the climate community, continue to be clear about their requirements for useful satellite observations (WMO, 2011a; Wulfmeyer et al., 2015; Ohring et al., 2005). The requirement for a long-term, stable record is challenging to formulate and justify. A variety of advances have been offered to allow for more accurate satellite observations: on-board calibration, independent verification, and in-depth modeling of instrument performance can all assist in characterizing the accuracy, biases, and stability of the satellite measurements, many based on the fundamental measurement equations and availability of internal instrument monitoring. The challenge will still remain to understand and merge records from different satellites – each potentially using its own calibration and collection approaches – to provide a single observational record. One of the key factors that we can control is the length of overlap between existing and future satellites. Analysis of an overlap record can only give us an estimate of relative drift, but in the absence of traceable in-flight calibration, it is often one of our best checks on long-term stability of the final data products. Understanding that decisions on overlap will directly affect both the cost of monitoring and the value of the final dataset for evaluating long-term changes in climate, we propose one approach to objectively evaluate the length of overlap needed to achieve a specific stability in the merged data record.

2. Approach

The statistical analysis techniques developed by Weatherhead and collaborators (see for example, Weatherhead et al., 1998; 2000) provide a basis for addressing the length of time needed for adequate overlap based on the magnitude of the signal variance as well as residual noise autocorrelation. Space-based observations are frequently interrupted by spacecraft or instrument faults that can significantly increase the length of time required to detect a data trend. In this manuscript, we perform a case study by applying these techniques to existing SORCE SIM and SOLSTICE instrument data thereby illustrating the use of statistical methodology to estimate the length of overlap needed to achieve records of specified stability. This study serves as an example of possible analyses that would ensure adequate overlap between two sensors measuring the same climate record. The techniques discussed herein are applicable to instrument scientists pursuing improvements in on-board instrument corrections, but also for mission planning by program managers to ensure the best overlap characteristics of adjoining missions.

The detailed in-flight circumstances that produce instrument instabilities are highly specific to individual sensors so the best practice is to employ instrument telemetry and on-orbit calibration methods traceable to international standards. Such approaches can be used to develop detailed measurement equations that can account for the occurrence of degradation and correct the measured signal to produce high quality Level 1 data. The measurement equation carries its own uncertainties and, in principal, allows for the estimation of time-dependent uncertainties as a function of mission day. This measurement equation approach is advocated in the "Guide to the expression of uncertainty in measurements" (JCGM, 2008) by the Joint Committee for the Guides in Metrology (JCGM) and relies only on the known and measurable properties of the subsystems that compose the full instrument used for the observation. Instrument teams apply these corrections to produce the final Level 2 data in an effort to provide the most accurate measurements independent of outside data sources.

2.1 Introduction of SORCE SIM and SOLSTICE instruments:

For illustrative purposes, we will use two sets of data from the Solar Radiation and Climate Experiment (SORCE) satellite: concurrent data from the Solar Stellar Irradiance Comparison Experiment (SOLSTICE) and the Spectral Irradiance Monitor (SIM). SORCE was launched on 25 January 2003 and has conducted daily measurements of the spectral and total irradiance with only a few gaps in the time series; the longest gap being a 209-day period starting on 31 July 2013. This gap was caused by a reduction in charging capacity of the spacecraft batteries, and has been successfully mitigated by operating the instruments in a day-only operation mode that does not rely on keeping spacecraft subsystems operational on the nighttime portion of the orbit. The SORCE mission and instrument performance, design, operation, and calibration are described in a series of papers published in *Solar Physics* related to the design, operation, calibration, and performance of the SORCE instruments. Harder et al. (2005a) describes the scientific requirements, design, and operation modes for the instrument. Harder et al. (2005b) discusses the fundamental measurement equations and the pre-flight calibration methodology for the instrument. A third paper (Harder et al., 2010) continues the discussion of the absolute calibration of the instrument describing additional post-launch characterizations using flight spare components and comparisons with the SORCE and UARS SOLSTICE instruments and the ATLAS 3 composite (Thuillier et al., 2004). Additional in-flight comparisons with the ESA ENVISAT SCIAMACHY instrument (European Space Agency, Environmental Satellite, SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY) are discussed in Pagaran et al. (2011). Similarly, McClintock et al. (2005a, 2005b) describe the SOLSTICE instrument design and calibration. Snow et al. (2005) describe the important solar-stellar calibration process that forms the basis of the on-orbit degradation corrections.

2.2 Set up for SIM/SOLSTICE comparison:

A crucial question for solar irradiance studies, as is true for most long-term satellite monitoring efforts, arises from the requirement to understand the length of time needed for the overlap of the ongoing SORCE mission with the next generation Total and Spectral Irradiance Sensor (TSIS) scheduled for launch in August of 2017 and deployed on the International Space Station. While Earth observations often require a minimum of a one year overlap to cover the full range of expected observations, such criteria is impractical for covering a full 11-year solar cycle in a planned overlap period. Here we are applying analytical techniques to understand the length of time needed to quantify the offset between two satellite observing systems and to understand the drift between two satellite records (Weatherhead et al., 1998). While it is unclear whether the TSIS/SORCE overlap will mimic the findings from the comparison of the two SORCE instruments, this effort will examine how potential instrument anomalies and systematic errors in the degradation corrections affect the ability to determine the length of time needed to determine a trend difference in the two sensors. For this study, we are using a subset of three years of data from 18 November 2005 to 31 December 2008 (1140 days), characterizing the time period corresponding to the descent into the solar minimum condition of Solar Cycle 23 with the minimum value apparently in the January-February period of 2009. This time period was selected to approximate what would be expected from an overlap comparison campaign conducted during the descending phase of Cycle 24 projected to be in the 2019 time frame. However, the Solar Cycle 23 minimum is the longest and quietest time period of the space age (Schrijver et al., 2011; Araujo-Pradere et al., 2011; Araujo-Pradere et al., 2012), but our analysis does not rely on this situation persisting into the Solar Cycle minimum. This paper targets common observations of the irradiance in the 280 nm spectral region which includes the highly variable core of the Magnesium II lines. This region was selected because the variability of the Mg II lines is an important indicator of solar chromospheric variability and is frequently used for space weather applications (Viereck et al., 2004; Marchenko and Deland, 2014) and as a proxy for solar influence on stratospheric ozone and temperature (Hood and Zhou, 1998). It should be noted that the SIM (version 22) and SOLSTICE (version 15) used in this study are used as-reported on the publically available SORCE web page (http://lasp.colorado.edu/home/sorce/data/ssi-data/). SIM and SOLSTICE corrections are made independently of one another, but the higher resolution 0.1 nm resolution SOLSTICE data is integrated into a fixed 1-nm bin centered at 280 nm. The SIM instrument has a FWHM resolution of 1.1 nm with 6 samples per resolution element. While some offset in irradiance is expected due to spectral sampling used to generate the data products, the difference is fixed and does not drift as a function time due to the well-defined wavelength scale and spectroscopic properties of the two instruments (Harder et al., 2010).

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

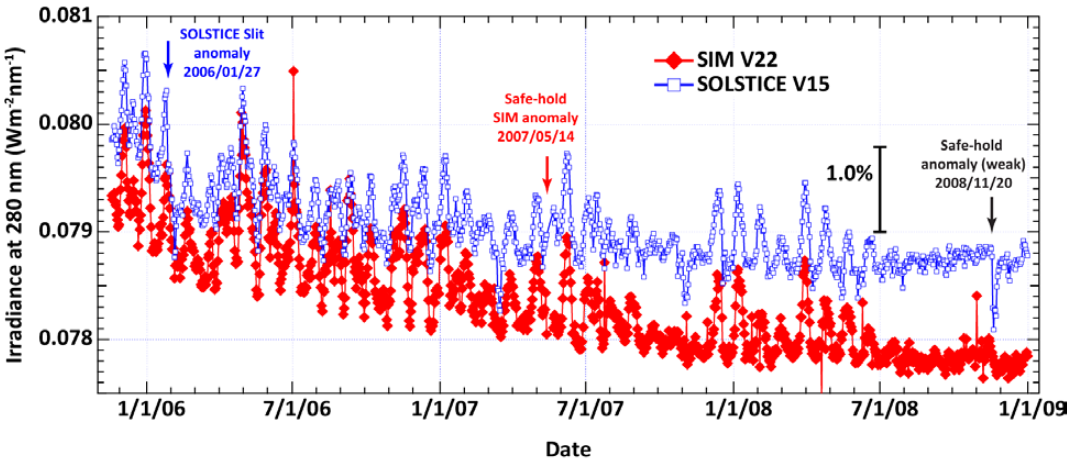

Figure 1: A 1140-day segment of concurrent SORCE SOLSTICE (version 15) and SIM (version 22) data during the descending phase of Solar Cycle 23. The 27-day variations seen in this plot are caused by solar rotational modulation of active regions dispersed on the Sun and are not due to instrument noise; note that this modulation is apparent in both datasets. This data segment contains all three of the sources of the uncertainties identified in this study: 1) Drift, as seen in the time-dependent dispersion in the data time series between SOLSTICE and SIM. 2) Offset is caused by differences in the absolute calibration between the instruments. 3) Jumps are identified in the time record and are the result of documented instruments faults. These three sources of uncertainties combine and contribute to the length of overlap needed to derive a robust climate record from satellite records.

Figure 1 shows the time series comparison of SOLSTICE version 15 and SIM version 22 used in this study. These overlapping datasets illustrate three types of inconsistencies that occur in geophysical records.

1. There is an offset of about 0.5% in the absolute calibration between the SOLSTICE and SIM. The pre-flight absolute calibration is on the order of 1-2% so within the ability to absolutely calibrate the spectrometer. However, in an overlap experiment as we are conducting here this offset could be significant.
2. There is an apparent drift in the data between the two instruments. The advancement in SOLSTICE version 15 contains a new correction that removes an annual oscillation in the data induced by a change in the size of the degradation spot 'burned-in' to the collimator mirror – see McClintock et al. (2005a) for more detail on the optical configuration. As the Earth-satellite system moves around the elliptical orbit of the Sun a different illumination occurs on the first optic thereby modulating the intensity of light that propagates through the rest of the optical system. This same effect occurs in the SIM data but has been a part of the standard degradation correction for the last versions. SOLSTICE version 15 tends to flatten the apparent long-term magnitude of the 280 nm variability relative to earlier SOLSTICE versions.
3. There are jumps in the time series related to spacecraft and instrument anomalies. Significant events are apparent in the record, such as when SOLSTICE experienced a failure of the mechanism that changes the entrance slit from the solar to the stellar mode on 27 January 2006. The slit was moved back into position for continuous solar observations but did not return to the exact same position so the optical path through the instrument changed and therefore disrupted the degradation corrections and the wavelength scale. Similarly, a spacecraft safe-hold event on 14 May 2007 caused the instruments to become very cold and significantly changed the SIM wavelength scale and perhaps the transmission properties of the instrument. This safe-hold event had little effect on the performance of the SOLSTICE.

The next sections of this paper will address the effects of these three anomalies (offsets, drifts, and jumps) in the SORCE datasets and discuss their impacts on dataset uncertainty and on the length of measurement overlap required to achieve a specific level of stability in the final dataset.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

3. Offsets

Offsets in the values measured by two different satellite observing systems are now a well-documented issue across many satellite monitoring systems. Making corrections to sequential satellites is standard practice to allow for a continuous data record that can be used in scientific studies. Details of the merging process directly influence the resultant trends and add to the level of uncertainty in the final datasets (Karl et al., 1986). In this section we consider the case where overlap is non-existent and for the case where overlap exists, we consider the length of time required to achieve a specific uncertainty in an overlapping set of data. These cases illustrate not only the need for overlap periods, but for overlap periods of sufficient duration to make a quantifiable improvement in the long-term record.

Free et al. (2002) carried out a comparison of seven different groups examining radiosonde records with the intent of identifying, quantifying and correcting observed discontinuities. The different groups identified a widely different number of discontinuities using a range of techniques. They further differed significantly on the magnitude of corrections and even, at times, the sign of the needed correction. Thus, even identified and corrected discontinuities introduce some uncertainty in the long-term stability of the record. Weatherhead et al. (1998) showed that these corrections can be quantified with advanced statistical techniques. The resulting uncertainty in long-term stability increases the number of years necessary to detect trends, independent of how large the correction is.

As an example of one of the potential problems, consider the straight-forward method for merging two sequential (non-overlapping) data records by forcing the mean level of the three years of data prior to the discontinuity be equal to the mean level of the three years of data after the discontinuity. In such a situation, those six years of data are being forced, by the algorithm, to have very little trend. Imagine a situation where there are two such discontinuities in a twenty year record; more than half of the data has been coerced to have virtually no trend, making the resultant data unreliable for many long-term monitoring uses. The case of no overlap can occur due to a variety of reasons, including the sudden loss of a satellite, or in the case of ground-based observations, when an expensive instrument needs to be moved and no duplicate instrument can be employed to allow for overlap. The end result of any offset correction will have a direct impact on the magnitude of the resulting trends.

Planning satellite operations for overlaps in monitoring missions allows for a quantification and minimization of the uncertainty in the merged dataset. To estimate the time needed for overlap requires an estimate of what the overlap time series would look like. We use monthly averages, as is a common standard in many climate-related research efforts for several reasons: monthly averages avoid the match-up issues and potential non-linearity of short-term features, such as storms, and offer enough resolution to observe long-term behavior of the match-up.[1] If, as is the case of the SOLSTICE-SIM overlap, the data and the difference between the two datasets look like Fig. 2, the paired data can be used to estimate offsets and uncertainties in the derived offsets.

[1] While it could be argued that there is nothing unique about the time-step of one month, it is a common practice in climate analyses. However, the example datasets used in this paper measure extra-terrestrial solar radiation. With the Sun's rotation of 27.2753 days, we have a natural timeframe close to a monthly average. In Appendix A we carry out the calculations in this paper with monthly averages and with averages based on the solar rotation schedule; we see no notable change in the basic conclusions adopting the more natural solar rotation schedule instead of monthly averages.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Atmospheric Chemistry and Physics
Discussions

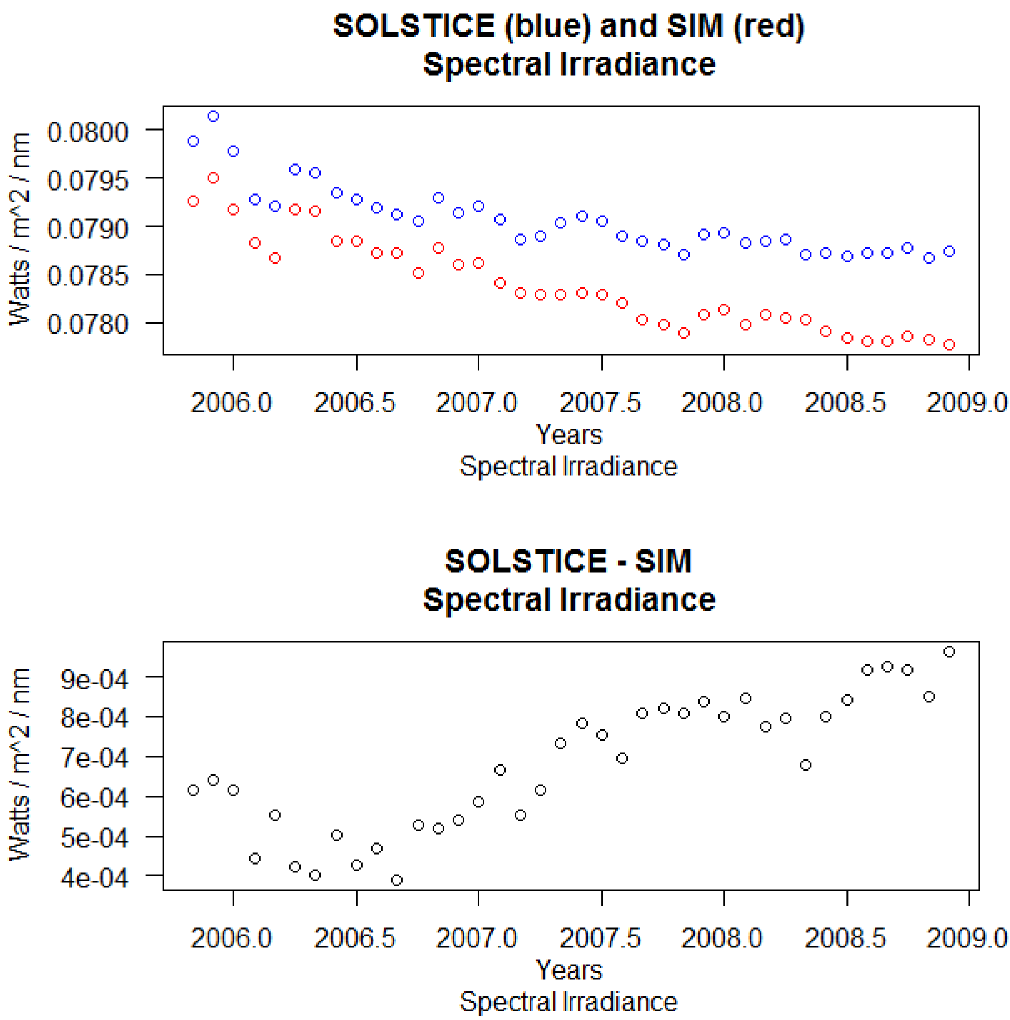

Figure 2: Monthly averaged SOLSTICE data, SIM data and SOLSTICE-SIM (watts m^{-2} nm^{-1}) as a time series for the period October 2004 through December 2008. The data show that the observed difference are small, but do not appear to be stable, thus a simple level shift to bring the two datasets in line may not fully address the match-up and stability issues of the dataset.

The overlap data depicted in Fig. 2 show a mean difference between the two datasets of $6.8*10^{-4}$ watts m^{-2} nm^{-1}, with a standard error on this mean of $2.7*10^{-5}$ watts m^{-2} nm^{-1} when the classic standard error calculation ignores autocorrelation. However, this figure does not support the assumption that the observed difference between SOLSTICE and SIM are stable and would continue beyond the observed end of the analysis period of December 2008 because of the apparent drift in the differences. For cases when a drift is not involved, we can make use of the standard formula for the standard error on the mean of the observed time series of differences when simple autocorrelation is present:

$$SE_{mean} \cong {}^{\sigma}/_{\sqrt{n}} \frac{\sqrt{1+\varphi}}{\sqrt{1-\varphi}} \qquad (1)$$

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Where σ is the observed magnitude of variability of the observed differences in monthly averages; φ is the observed autocorrelation in those differences, and *n* is the number of months of observed overlap. This estimate of Standard Error of the mean is dependent on the data behaving as an autoregressive with time-lag of one month, AR(1), with the underlying interventions behaving approximately as a Gaussian distribution. This more appropriate formula gives a standard error on the mean of $5.2*10^{-5}$ watts m^{-2} nm^{-1}, notably larger than if autocorrelation is ignored. These constraints for the formula are some of the reasons that monthly averages are used as often as they are: monthly averaged data is often better behaved and match-up problems from different instruments are minimized. We show the behavior of the underlying interventions as Gaussian and our tests for AR(1) in Appendix B.

We can invert the formula for the Standard Error on the mean in Eq. (1), and solve for *n* resulting in the number of years to estimate the mean offset between two satellites for a given accuracy as:

$$Number\ of\ Months\ to\ Estimate\ an\ Offset \cong 1.96^2\ {\sigma^2}/{Offset\ Limit^2}\frac{1+\varphi}{1-\varphi} \quad (2)$$

The above formula shows that for a given magnitude of variability and autocorrelation in satellite overlap data (σ and φ respectively), the length of overlap needed is inversely proportional to the square of the accuracy desired for the offset estimate. Thus, if we can identify the level of uncertainty we can accept in a merged record due to the overlap offset (SE_{mean}), and if we have some understanding of the behavior of overlap differences (σ and φ), either from advance estimates or from early analysis of offset data, we can appropriately identify the length of overlap needed in a manner that is respectful of the inherent cost of added months of satellite overlap. If a higher level of certainty than 95% is required, the 1.96 factor is adjusted appropriately according to normal distribution tables. For small number of months, the 1.96 will need to be adjusted for the student-t distribution which allows a larger uncertainty when a small number of points are used. With the example used in this paper and shown in Fig. 2, we observe a magnitude of variability, σ, of $1.7*10^{-4}$ watts m^{-2} nm^{-1} and autocorrelation, φ of 0.89. If we want an Offset Limit of 0.0008 watts m^{-2} nm^{-1} (which is one percent of the mean of SOLSTICE during the overlap period), then the number of months would need to be 5 months using the student-t distribution which offers 2.8 as the appropriate factor in place of 1.96. Note that to achieve the 95% confidence limit, we must use the appropriate student-t distribution, or approximately 1.96 multiplier in the large number limit, to assure we have the desired confidence on our overlap adjustment. While this is a recursive effort because the answer, number of months, is a function of the multiplier, which is itself a function of the number of months, this exercise is not overly onerous, because high precision is not appropriate and after roughly two years of data, the large number limit of 1.96, may be considered appropriate.

The impact of the offset on the use of the data is critically important to the final analysis. While a "best" merged dataset may be produced from multiple satellites, users should never ignore the added uncertainty due to merged data sources. The magnitude of the impact of the offset correction is dependent on the use of the data. Two cases are considered here for illustrative purposes. If the merged dataset will be used to estimate the impact of storms on a stable electrical grid, and the impacts have been estimated from the effects observed using the first satellite record, an uncertainty of 0.2 (%) means that the new solar storms may well be off by ±0.2% and the uncertainty in impacts need to be appropriately calculated and conveyed. If the merged datasets will be used to estimate long-term trends, then the impact of an uncertainty of ±0.2 (%) means that any trends derived will be affected by that level of uncertainty carried out through the length of dataset used for analysis, and may affect the significance of the expected trend, if care is not taken to reduce the uncertainty in the overlap adjustments.

4. Drifts

If we can quantify the level of drift we would like to be able to detect, and if we can estimate the level of variability in the overlapping data, using approaches from Weatherhead et al. (1998), we can estimate the number of months necessary to observe a drift of that magnitude in an overlapping dataset. Weatherhead et al. (1998, 2000) have shown that one can estimate the length of time to detect trends in environmental observations. This approach is applied to estimate the time to measure a differential drift with a specified uncertainty in the observations taken by two different systems. When detection is considered at the 95% confidence level as:

$$Number\ of\ Months\ to\ Estimate\ a\ Drift\ \cong\ 12 * \left[1.96\ \frac{\sigma}{|drift|} * \sqrt{\frac{1+\varphi}{1-\varphi}}\right]^{2/3} \quad (3)$$

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Where |drift| is the absolute value of the magnitude of the differential drift, σ and φ are the magnitude of variability and autocorrelation of the overlapping monthly data, respectively. We can estimate the drift we would like to have the capability of detecting; and estimate both σ and φ from existing data – either from observations or modeled experiments. It may be noted that the natural world has variability (σ) and autocorrelation (φ) that are inherent, and may change slightly over time. A distinct advantage of satellite observations can be the frequency of the observations. MacDonald (2005) has shown that the monitoring approach can have a direct impact on these values, as well: monitoring less frequently – perhaps only once or twice a month – results in higher variability and slightly lower autocorrelation in our dataset. Thus, we can increase our measurements per month to improve our detectability (shorten our number of years to detect a given trend), only up to the limit which is the natural variability of the system.

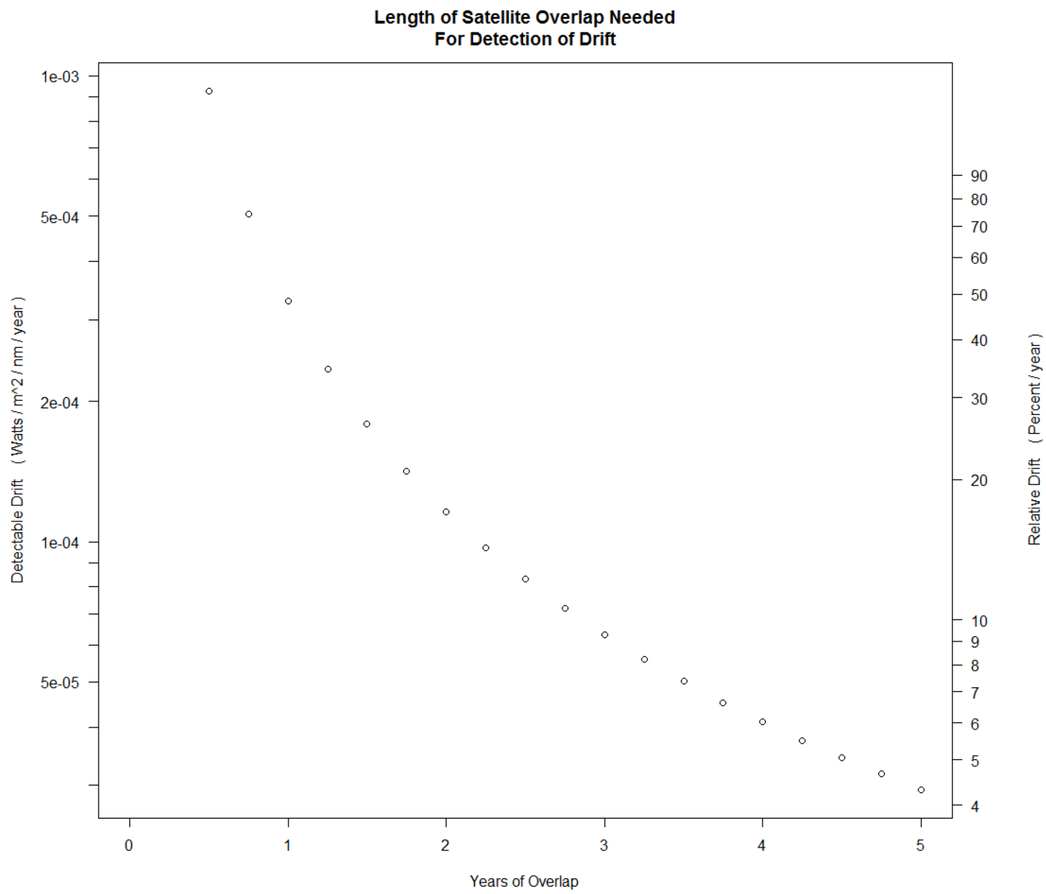

Figure 3: The ability to constrain a detectable drift is a direct function of the number of years of overlap. Using the SOLSTICE-SIM data as an example, tremendous accuracy gains are achieved for each year of monitoring for the first few years of overlap, with diminished returns after that. Respecting the cost of overlap, and making appropriate calculations with emerging overlap data, an appropriate overlap plan can be estimated to allow for scientific standards to be met. A detailed set of comments on appropriate interpretation of the time estimates along with an estimation of vertical and horizontal error bars is offered in Appendix C.

Figure 3 shows that for a given magnitude of variation and autocorrelation observed in overlap differences, we calculate the number of years of overlap needed to detect a specific level of drift. In this case, using the SOLSTICE-SIM data as an example, the magnitude of detectable linear drift drops from $1.2*10^{-4}$ to $0.6*10^{-4}$ watts m^{-2} nm^{-1}

year^{-1} by allowing the overlap to be three years, instead of two years. It is important to note that trends in overlap data, as trends in nature, can be best approximated as linear, logarithmic, or a variety of other representations, as the data and the physics of the situation imply; for Fig. 3, we assume a linear drift over the time period of the overlap. For satellite observations, a variety of changes are expected over the lifecycle of the instruments; all known and expected changes are approximated and adjusted based on current best understanding. However, particularly with new technologies, these assumptions must be checked by careful evaluation of the data, thus emphasizing the importance of an adequate overlap period to help confine potential drifts to a specified level. While pre-launch calibration may indicate drift will be less than a specific level, the ability to verify this will depend on independent intercomparisons of observations.

Although no error bars are offered in Fig. 3, it is important to remark that when estimating how long it will take to detect a drift, two statistical levels must be considered: one that identifies the meaning of "detecting a drift" and a second that identifies the likelihood of detection of that drift in the specified period of time, if that level of drift is the true, long-term drift in the overlap. Of course, it is possible to detect smaller drifts than the value obtained from any particular point (drift-overlap pair) of this figure, as it is also possible to determine a given drift a few months earlier or later than the value obtained from the point in the figure. Comments on the appropriate interpretation of the likelihood of drift detection are discussed in detail in Appendix C, where two non-standard, dimensional error bars are introduced in the figure to help the reader to understand this uncertainty.

5. Jumps

Sudden offsets or jumps in environmental observations occur for a variety of reasons (e.g. Brown, 2013). In some cases, true jumps occur in the parameter being observed. In many other situations, the observing system or assumptions used in the algorithms are responsible for the jumps and there is a desire to identify and remove these spurious jumps. A variety of approaches are used to both identify jumps (Jaxk et al., 2007; Vincent et al., 1998, 2002; Ducre'-Robitaille et al., 2003) and correct for these jumps in satellite and non-satellite observing systems (Karl and Williams, 1987; Mitchell and Jones, 2005). When information is available to identify the timing of a jump, there is considerably more confidence in the correction for the jump because physical interpretation is easier and therefore corrections can be physically based rather than statistically based. If there is no external information on timing, then one has to consider that there could be other jumps below the threshold for detection and estimates for how the instrument is behaving are more uncertain. For jumps in overlapping datasets, the correction brings into question the magnitude of any derived offset as well as the magnitude of any drift in the overlap period. For jumps in observing systems when there is no overlap, the challenge of appropriately identifying and correcting jumps is notably more difficult, again pointing to the value of redundant observing systems when possible.

The statistical removal of a spurious jump involves two steps: identifying a jump in the overlap differenced data and estimating the magnitude of the needed correction. The uncertainty associated with these sudden jump corrections adds uncertainty to the derived offsets and drifts in the overlap data, requiring additional years of overlap to achieve the same level of confidence in the offset and drift estimates. As mentioned earlier, Free et al. (2002) showed that different intelligent approaches can produce varied results in both the number and magnitude of recognized jumps in a time series. The range of estimates for the magnitude of jumps in this paper underscores the added uncertainty to the long-term stability of the dataset when any jump is identified and corrected. We consider the two cases separately of how a jump affects the offset estimate and how a jump affects the drift estimate.

While ancillary data about the instrument or the observed parameter may be used to identify the existence and timing of a jump, deriving the magnitude of correction by examining the data is dependent on the amount of variability (both magnitude and autocorrelation) in the overlap difference data. For the case of an otherwise stable offset between the two sets of satellite data, the added uncertainty on the true offset is enhanced based on the uncertainty of the jump. The impact of the jump on offset estimates is a minimum when the offset occurs in the middle of the overlap period, because maximum information is available to identify the size of the jump. For the case of a drifting offset between the two sets of satellite data, the added uncertainty on the true drift is also enhanced based on the timing of the jump. Because the impact of the jump is co-linear with the derived drift, estimating the overall drift in instrument offset is more difficult in the presence of jumps. Weatherhead et al. (1998) show that the impact of a jump on deriving an accurate estimate of drift is dependent on the timing of the jump. Thus, a longer time of overlap is required for accurately confining a drift in overlap data:

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Atmospheric Chemistry and Physics
Discussions

$$\textit{Number of Months to Estimate a Drift} \cong 12 * \left[1.96 \frac{\sigma}{|drift|} * \sqrt{\frac{1+\varphi}{1-\varphi}}\right]^{2/3} \frac{1}{[1-3\tau(1-\tau)]^{1/3}} \quad (4)$$

Where τ is the fraction of the data before the identified jump occurs. As opposed to the impact of jumps on estimating offsets, the impact of jumps on derived drifts is largest when the jump occurs in the middle of the overlap period (τ=0.5), where the amount of time is increased by a factor of 1.59. This increase is in length of time needed to estimate a drift is due to the similar temporal signature of both drifts and jumps on a long-term record as illustrated in Fig. 4. Equation 4 assumes that the drift and offset are fitted to the data simultaneously. If data are fitted sequentially to an offset and then the a drift, the derived drift will be considerably smaller than if the data were fitted to a drift and then to an offset because the two functions (drift and offset) are not orthogonal and thus the derived results for magnitude of offset and magnitude of drift are not commutative. To be explicitly clear, correction of offsets in advance of deriving drifts can artificially minimize the amount of observed drift, while ignoring offsets (perhaps because they are not easily detectable) can either add to or diminish the derived drift.

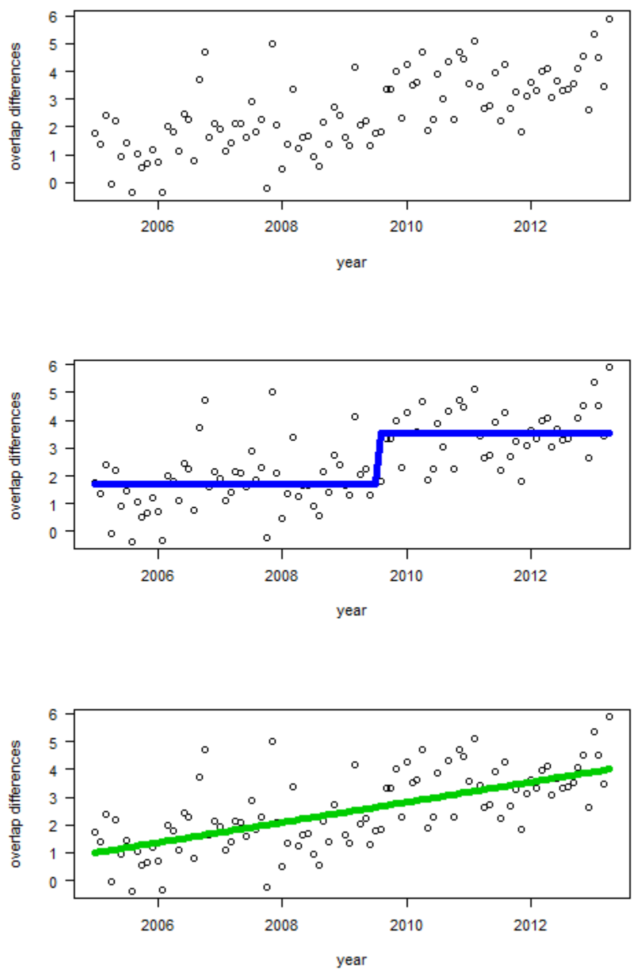

Figure 4: This dataset was created to visually show the potential impact of a spurious jump on the estimate of offsets and drifts. In all three plots, the same data are shown, with the second plot showing how the data could be modeled as an offset. The data were actually created by adding a linear trend to simulated autoregressive data, as shown in the third

plot. The confounding nature of jumps and trends cannot easily be separated; although ancillary data can be extremely helpful. Note these are synthetic data with arbitrary units.

In the case of the SORCE instruments, these jumps are mostly prompted by changes in the performance of hardware subsystems of the spectrometers. For example, a significant jump occurred in the SIM on 14 May 2007 related to the instrument becoming very cold during a safe-hold event, and upon recovery, the instrument did not return to the same state as before the safe-hold. This event produced a significant change in the hardware that controls the wavelength drive (see Harder et al., 2005a for a description of the wavelength drive mechanism). Even after the wavelength correction a residual jump in irradiance level was still observed; the most likely cause of this jump is due to a change in light transmission caused by a change in optical path through the instrument that is different than before the safe-hold event.

For the particular case of Fig. 1, the jump appears to be adequately corrected, but other wavelengths show a discernable discontinuity. A similar observation can be made about the SOLSTICE slit anomaly in 2006 (see Fig. 1). Most disconcerting about these events is the possibility that the jumps can be disguised as a change in stability and produces results similar to what appears in the discussion on Fig. 4.

6. Impacts of Uncertainty in Drifts

The primary purpose of overlap in instruments is to understand how two different instruments are responding to the parameter each is intended to measure. Uncertainty in the long-term stability of data products directly affects the level of confidence one can have in the long-term, merged datasets. The impact of drifts on the long-term records can be illustrated with a simple Monte Carlo experiment illustrated in Fig. 5. Consider two satellites launched with seven year lifetimes. The second satellite (blue trace) overlaps the first satellite (green trace) by either one or two years. Each new satellite launches with an uncertainty estimate on the pre-flight absolute calibration of its measurements and some level of unidentified drift in the instrumental record, acknowledging that all identified drifting factors have already been corrected to the best abilities of the instrument team.

The overlap period can be a valuable check on the potential drift of satellite measurements, with the longer the overlap period, the more constraint that can be identified as unexplained relative drift. If the information on relative drift obtained from overlap, together with valuable instrument analyses, can help constrain the absolute drift on long-term data, then we can perform a Monte Carlo simulation of two situations using the example data from the SOLSTICE-SIM analysis presented earlier in this paper. The first simulated satellite might launch with an estimated uncertainty of the solar Mg-II data of 1%represented in Fig. 5a as the spread observed in year 0. Because only one year of overlap was used, the drift may be as large as 0.000328 watts m^{-2} nm^{-1} $year^{-1}$ (one standard deviation), roughly 48% $year^{-1}$; for illustrative purposes 100 simulations were randomly generated to allow for that size drift (one sigma in a normal distribution). The green lines indicate the uncertainty in the long-term (seven year) lifespan of this first simulated satellite. After six years, a second satellite is put into orbit; but because of improvements in pre-launch calibration, the initial uncertainty on this second satellite is estimated to be 0.33%, as represented in the top plot of Fig. 5 by the spread observed in the start of the second satellite in blue. This second satellite, because it also had an overlap period of one year, may have a drift as large as 0.00328 watts m^{-2} nm^{-1} $year^{-1}$, as estimated by the uncertainty on the match-up of the two satellites. The blue lines indicate the drift that could take place during the six years of flight for this satellite. Figure 5b shows the same information as the first plot, but assumes that the drift for the two satellites may only be as large as 0.000116 watts m^{-2} nm^{-1} $year^{-1}$ (one standard deviation), roughly 17% $year^{-1}$, because the satellite overlap was for a longer (two year) period.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

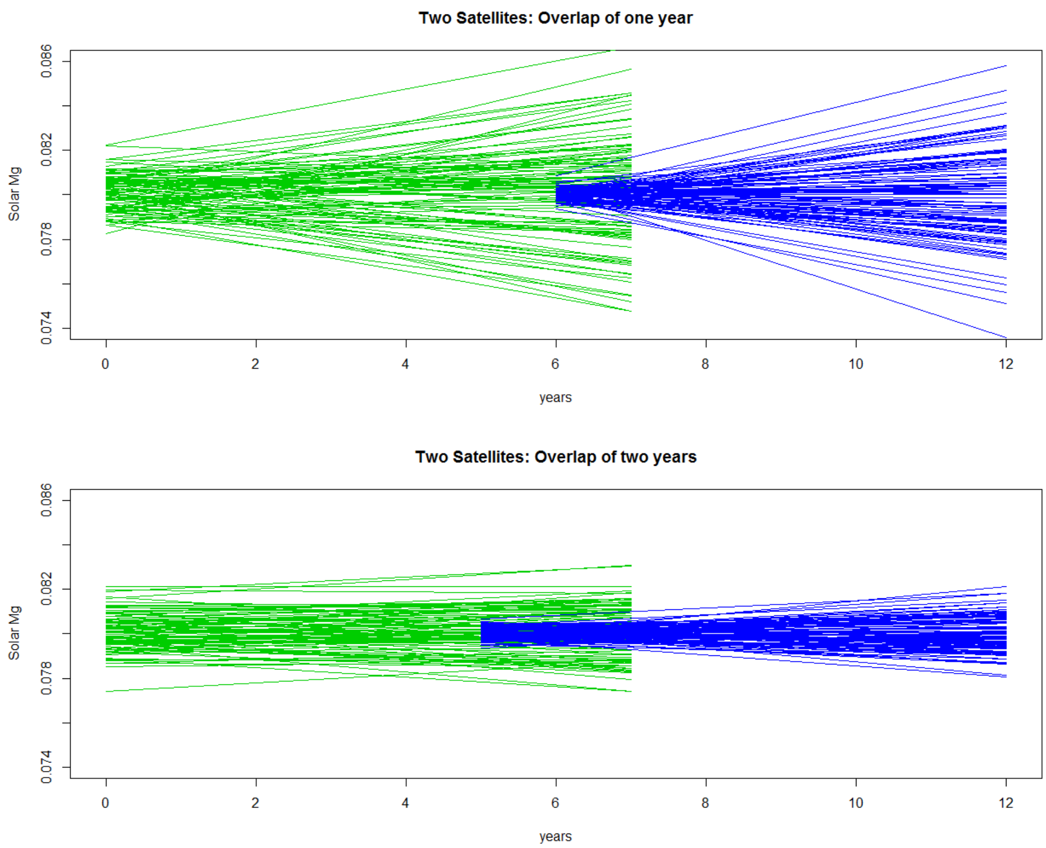

Figure 5: Two overlapping satellite observations (shown in green and blue) simulating 280 nm Mg II irradiance (watts m^{-2} nm^{-1}) with an expectation for some absolute pre-flight calibration uncertainty and some unidentified on-orbit drift. The amount of unidentified drift can be estimated in the overlap period, resulting in the possibility of a correction to the data. The uncertainty in the estimate of the drift is notably smaller with a two year overlap period in satellites compared to a one year overlap.

This method of showing the impacts of absolute uncertainty and drift, modeled after Frith et al. (2014) can be helpful in identifying the magnitude of impact of potential drift as well as how best to merge data based on the particular circumstances. These two "Frith plots" show that the long-term record will be directly impacted by the pre-flight calibration of each new satellite and our ability to constrain the range of unexplained drift in each satellite. For the long-term use of the record, the constraint on the drift can be critical to stability that may be needed for the long-term analysis of geophysical phenomena. As with any instrument intercomparison, satellite overlap by itself cannot replace calibration, but it does offer valuable information to allow instrument scientists to make judgments on stability and reliability. Frith et al. (2014) use Monte Carlo techniques to make adjustments to satellite data based on both overlap information and additional verification techniques. Ongoing studies are looking into the value of on-flight calibration (Wielicki et al., 2013; Tobin et al., 2016) as well as redundancy in observing systems (Weber et al. 2016) and the value of continued replacement of satellites to minimize the extent of unchecked drift (Frith et al., 2014 and Weber et al., 2016).

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

7. Application to Earth Observations

Extending these approaches to assure appropriate satellite overlap of Earth observations involves some challenges and advantages that are not apparent in the example of the solar observations used in Sections 3, 4, and 5. One large challenge is that Earth observations often involve spatial and temporal variations (e.g. Araujo-Pradere et al., 2004) that total solar output does not have. Two satellites making Earth observations likely provide data with differences in spectral resolution and sensitivity, spatial resolution and sampling, as well as temporal period (e.g. Chander et al., 2013b, Toohey et al., 2013). These corrections, combined with the spatial and temporal variations of Earth's parameters introduce a large set of challenges that are not easily overcome. A distinct advantage for satellite Earth observations is that a variety of in situ and ground-based observations may exist to help bridge the gap between two sets of satellite measurements, even if the overlap is of insufficient duration and coverage to provide the needed continuity.

A number of efforts have worked to define the requirements for Earth observations, including the recently completed effort by WMO (2011b), which addresses the stability needed for various parameters. The value of satellites to understanding the Earth system with a global perspective is unparalleled. To understand long-term changes, stable data records are needed but are not easily evaluated without comparisons to other observations to assess offsets and drifts. In the absence of explicit requirements for limits on drift, we suggest that the standard error of the drift, at the one sigma level, be limited to half of the trend that one is seeking to detect. For example, if a monitoring system is designed to detect a trend of 0.2 degrees per decade, the unchecked drift of the system should be less than 0.1 degrees per decade at the one sigma level. While for Earth Sciences, the projected trend is dependent on the model and assumptions used to estimate future trends – as well as the location and time of projected trends, this estimate can be used as a starting point for discussions on how well the drift should be confined. When the verification of drift cannot be held to the level of projected trends – when using all available information – there can be serious questions as to the usefulness of the monitoring system for trend identification.

The general principles outlined in this paper and illustrated using the example of solar observational satellite overlap are directly applicable to Earth observations, but require careful application to avoid oversimplification of the problems. Coordination efforts to bring the highest level of calibration and homogenization of data are ongoing internationally through GSICS (Hewison et al., 2013; Wu et al., 2009) and merged datasets are coordinated internationally by efforts including those led by the Centre National D'Etudes Spatiales. One of the most intriguing concepts which will need to make use of careful comparison analytics is the potential for traceable absolute calibration from satellite intercomparison. These ideas have been formulated in the vision of CLARREO and TRUTHS, but may be tested in other reconfigurations (Wielicki et al., 2013; Best et al., 2008; Fox et al., 2013; Tobin et al., 2016). Efforts are already underway to estimate uncertainties due to the matching of independent satellites with these reference sources (Feldman et al., 2011; Lukashin et al., 2013). Even with perfect calibration of the reference radiometers, challenges with matching of observations and instrument characterization will require comparisons over a long period of time to assure accurate transfer capabilities. In this case, the techniques outlined in this paper could be applied, so long as underlying assumptions are met and that the transfer of standards is in line with GUM standards. (JCGM, 2008).

8. Optimization and economic benefit

Decisions to improve a single observing system can rarely be made without consideration of the potential impact on the support of other approaches to improve monitoring. Figure 3 shows that the accuracy for detection of drift improves as a function of the number of years overlap. Marginal improvements in detection of drift decrease as the number of years increase, making the optimal choice of overlap years difficult to identify. If a specific stability criterion is the objective, the minimum required overlap can be directly determined to meet that criterion. Assuming that the total costs increase as overlap time increases (this value can often be in excess of $1-2M per year), from an economic perspective the minimum required overlap time should also be chosen as the most cost-effective.

In addition to overlap time, a number of choices may be proposed to improve observational climate records including, as already mentioned, improved pre-calibration of satellite systems, intra-satellite calibration mechanisms, redundancy in observing systems, and campaign verification of observations. In considering any such broader set of approaches to achieve a specific stability criterion, the least cost option or combination of options should be chosen. However stability criteria are often not available or substantiated. Under the constraint of fixed budgets, and without specific criteria, the maximum overlap period affordable would be optimal. If there are multiple approaches being considered, the combination of improvement approaches to achieve the greatest stability in the data within the exogenously determined budget would be optimal. Given the complementary information in various approaches to improve both absolute and relative calibration, linear optimization approaches may need to be developed to identify the best mix to achieve optimal calibration.

Without specific stability criteria or budget constraints, economic criteria suggest choosing the overlap period or combination of approaches to achieving data stability that provides the maximum net societal benefit (i.e., total benefits minus total costs).[2] Identifying the societally optimal choice implies choosing the overlap (or possible mix of calibration methods) where marginal cost equals marginal benefits. Marginal costs are the change in program costs to achieve one unit of improvement (e.g. watts m^{-2} nm^{-1} $year^{-1}$ in detectable drift as per Fig. 3). From an economic perspective, the marginal benefits of one unit of improvement in the data quality should be measured in terms of potential changes in societal outcomes from the use of improved information (a value of information or VOI approach (Laxminarayan R. and M.K. Macauley, eds. 2012)). While there has been some work on the value of information from satellites (Donaldson and Storeygard, 2016; Cooke et al., 2014; Macauley, 2006), there have not been many applications of economic analysis to determine optimal observational systems. Morss et al. (2005) provide a nice simplified summary of relevant economic theory and a case study based on primarily hypothetical valuation estimates.

Given the limited number of applied studies on societal benefits of satellite data standard approaches would need to be adapted to further understand the value of stable records and develop the decision making tools to optimize observation systems. Weatherhead et al. (2015) have gathered community input to help identify key science questions that need to be addressed, while Feldman et al. (2015) have outlined some tools for assuring observing systems can meet these approaches. The general field of climate observing system simulation experiments C-OSSEs is gaining serious attention from both scientists and science managers internationally.

8.1. Application of statistical approaches

This paper provides guidelines for what may be considered idealized situations. Approximations have been made about the linear nature of drifts, and timescales of jumps that are likely approximations to considerably more complicated instrumental response. Individual judgment is needed to apply the results from this paper as instrument characteristics often change over time and PIs will often have additional information that will guide their decisions about the quality and stability of instruments. Ground-based instruments would likely add further information to help evaluate stability. Notably, the beginning and end of most satellite missions are the periods where most challenges occur in the instruments and may alter the guidelines presented here. As an example, instruments behaving very badly, perhaps are not of sufficient quality to contribute useful information and less overlap would be needed. Much of this uncertainty points to the value of redundancy of sensors and the value of complementary observing approaches, despite their potentially high cost.

9. Conclusion

We acknowledge, as many colleagues before us (e.g. WMO, 2011a; Wulfmeyer et al., 2015; Ohring et al., 2005; Wielicki et al., 2013; NAS), the importance of a continuous satellite record to understand solar and planetary behavior. In this paper we focus on the development of a relatively stable data record, making full use of available satellite data, as opposed to calibration efforts to allow a traceable record of absolute accuracy. We examine three cases for the merging of satellite data: identifying and quantifying an offset, estimating drifts between two satellite records, and understanding the impacts of sudden changes in the data records on both offset and drift estimates. For studies making direct use of the satellite data, either to develop a continuous record or verify a continuous record, the most direct control available in an observing strategy is to control the length of overlap in the satellite records. We identify the impact of length of time of overlap on all three of these cases and illustrate these approaches with two satellites used to observe solar output. In reality, most satellites carry multiple sensors; final choices on overlap

[2] Appendix D provides a more technical explanation of the optimal choices under difference decision situations from an economic perspective.

periods may involve comparing results from each of these sensors. Solutions may include powering down some observations earlier than others, or prioritizing the results from different sensors.

The uncertainty due to the merging of satellite records, is unavoidable, but quantification of this uncertainty is possible. In the case of identifying or verifying the offset in two satellite records, the uncertainty is inversely proportional to the square root of the number of months of overlap. For a given tolerance of uncertainty in long-term relative stability, the number of months of overlap can be implemented to achieve that tolerance if a reasonable estimate can be made concerning the variability in the overlap data. If no estimate of overlap variability is available, the behavior of the first few months of overlap can be evaluated to estimate the length of time needed for identifying overlap times necessary to achieve the prescribed tolerance. To identify the relative drift between two satellites to a specified uncertainty, the specified uncertainty in drift can be achieved due to drift inversely proportional to number of months of overlap to the 2/3 power. The impact of abrupt disruptions in the overlap period on offsets can require up to 50% more overlap to be able to identify the offset and drift with the same level of tolerance.

These algorithms are appropriate for a direct evaluation of data using only the data. In some cases, particularly with Earth observations, added challenges and benefits may exist. For instance, with Earth observations, additional in situ and ground-based observations may be available to reduce uncertainty in satellite overlap. However, an additional challenge to Earth observations is the need for direct temporal and spatial overlap, which can be challenging or impossible as satellite observation approaches are considered within challenging budgetary constraints. All of the techniques outlined here can be applied to identifying the level of overlap needed with reference calibration satellites, such as proposed by CLARREO. In this case, the level of uncertainty in offsets and drifts will be determined not by the length of overlap, but in the quality of the match-up between the reference and operational satellites. Under various constraints, choices of overlap can be optimized to help assure climate records that are appropriate for advancing our understanding of the Earth system. The value of this paper is the ability to estimate, either prior to satellite launch or after satellite launch, the amount of time needed to achieve or verify tolerance for a stable merged satellite record using objective criteria.

Code availability: Jerry Harder will set up a place for this once the paper has been accepted; Eduardo Araujo-Pradere and Betsy Weatherhead will annotate the code so that it is understandable to R users.

Data availability: The data are publicly available.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Appendix A: Comments on the usefulness of monthly data.

The use of monthly averaged data has been common in climate studies for many years, despite obvious deficiencies in this somewhat arbitrary choice. One deficiency is in weighting a daily value in February more highly than a daily value from any other month simply because February has fewer number of days than, for instance, May. A second deficiency is the lack of match-up from the monthly timeframe to the natural word: the summer solstice is not in the center of the June, but off-center meaning that the June average would contain more information on pre-solstice conditions than post. Even non-scientific users of climate data are used to using reports such as “Climatic Normals, World Weather Records, and Monthly Climatic Data for the World” for useful information. WMO’s Guide to Climatological Practices even suggests, “Caution is needed when data are in sub-monthly resolution...” and makes considerable effort to coordinate climate data in a standardized manner (WMO/TD 341, 1989).

These issues are admittedly not likely of great importance, but for the reader who appreciates a great deal of caution and respect with the use of data we offer this simple examination of the impact of monthly averages in the context of the algorithms presented in this paper. For the example dataset used in this work, a more natural timescale is the Carrington rotation rate of 27.2753 days. For the small study presented in this appendix, we examine how the results of this paper would have differed if we had used the data averaged in 27-day periods as opposed to using the data averaged by month.

The change in values from monthly values to solar rotation values, the mean, standard deviation, standard error on the mean and autocorrelation change very little.

	Mean (watts m^{-2} nm^{-1})	Standard Deviation (watts m^{-2} nm^{-1})	Standard Error (watts m^{-2} nm^{-1})	Auto-correlation	Time periods (months or solar rotation cycles) to identify an offset of 0.0008 watts m^{-2} nm^{-1}	Years to identify a drift of 0.0001 watts m^{-2} nm^{-1} $year^{-1}$	Additional years to accommodate a jump
SIM							
Monthly	0.078	$4.78*10^{-4}$	$4.37*10^{-4}$	0.94	44	14.3	7
Solar	0.078	4.70*10-4	$4.31*10^{-4}$	0.94	47	14.6	7
SOLSTICE							
Monthly	0.079	$3.54*10^{-4}$	$2.38*10^{-4}$	0.89	13	9.5	5
Solar	0.079	$3.41*10^{-4}$	$2.27*10^{-4}$	0.90	13	9.5	5
SOLSTICE-SIM							
Monthly	$6.8*10^{-4}$	$1.67*10^{-4}$	$1.12*10^{-4}$	0.89	5	5.8	3
Solar	$6.8*10^{-4}$	$1.69*10^{-4}$	$1.14*10^{-4}$	0.90	6	6.0	3

Table A1. Fundamental descriptive values and calculations of overlap periods were calculated using monthly averaged data and data that were averaged on a 27-day time period, which is a more natural timeframe for these calculations. We note that little impact is observed from this small change in averaging period.

We note that the differences observed are remarkably small. Differences likely would have been larger if we had used data of a shorter duration (e.g. two years of data instead of six). One note is that when the solar rotation period is used for averaging, 42 data points are derived, as opposed to the 39 data points derived from monthly averages. This “larger number” of data points is accompanied by slightly lower standard deviation and nearly constant auto-correlation and directly feeds into the standard error calculation.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Atmospheric Chemistry and Physics
Discussions

Appendix B: Comments on the applicability of estimation of number of years of overlap.

For many statistical analyses commonly carried out in climate research, data are assumed to be near-Gaussian and independent (each value is independent of the others). For environmental data, monthly averaged data are often assumed to be auto-correlated with lag 1 month in such a manner that an AR(1) model can adequately describe the behavior of the data once seasonal aspects are removed. As a reminder, the number of years needed to detect an offset is estimated as:

$$Number\ of\ Months\ to\ Estimate\ an\ Offset \cong 1.96^2\, {\sigma^2}/{Offset\ Limit^2}\, \frac{1+\varphi}{1-\varphi} \qquad (2)$$

And the number of years to detect a drift is estimated as:

$$Number\ of\ Months\ to\ Estimate\ a\ Drift\ \cong\ 12 * \left[1.96\, \frac{\sigma}{|trend|} * \sqrt{\frac{1+\varphi}{1-\varphi}}\right]^{2/3} \qquad (3)$$

With σ and φ as the monthly standard deviation and autocorrelation as described in the body of the paper. Because the estimate of the number of years is dependent on these assumptions, we explicitly test the data used as an example in this paper for illustrative purposes.

The autocorrelation, φ, is the most difficult parameter to estimate accurately in a time series, particularly when φ is large. In the case of large autocorrelation, the time series can differ from the long-term mean for many months; if the estimate of phi is made from a small number of points, the sample estimate of phi can be off, but the standard deviation and mean can also be far from representative:

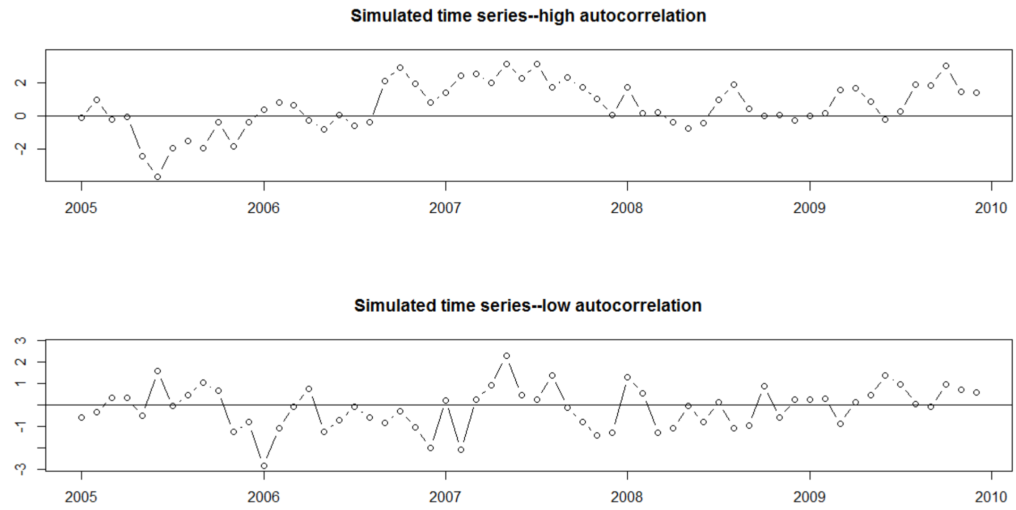

Figure B1. Two simulated time series are shown with high autocorrelation (top plot) and low autocorrelation (bottom plot). In situations of high autocorrelation (0.7 in the top plot), a time series can deviate from the long-term mean (0 in both of these simulated time series) for many months. If a short time period is used to estimate phi, likely phi will be under-estimated and the error on the sample mean may be farther from the true population mean compared to a situation with low autocorrelation (0.2 in the bottom plot). For the data used in this paper, the autocorrelation is estimated at 0.1 once drifts are accounted for and therefore the overlap period of six years is more than adequate to derive a good estimate for the long-term value of phi.

To test for AR(1) behavior in the SOLSTICE-SIM monthly overlap data, we calculate a partial autocorrelation function out to fifteen terms.

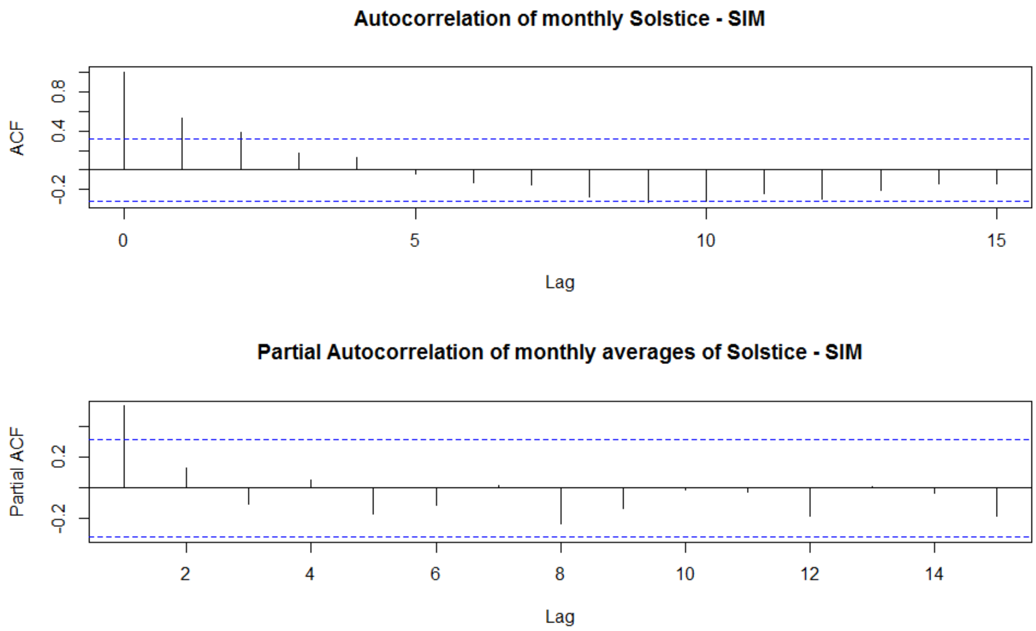

Figure B2. The Autocorrelation Function results and Partial Autocorrelation Function results support the temporal behavior of an AR(1) statistical model and therefore support the use of both Eq. (2) and Eq. (3) in the body of the paper.

The standard deviation, σ, is assumed to represent the spread of Gaussian or normal distribution. The standard deviation calculation can be carried out on any distribution and can be both informative and useful for many distributions. For the "number of years" estimate to be appropriate, the assumption is that the standard deviation represents the spread in a Gaussian distribution. For the AR(1) case, the test for Gaussian behavior is performed on the underlying interventions in the AR(1) process, which is similar to, but not identical to the residuals observed in the overlapped data. To test for Gaussian behavior, we compare our data to a standard Gaussian distribution in a QQplot (e.g. Hamilton, 1994; Box et al., 2015):

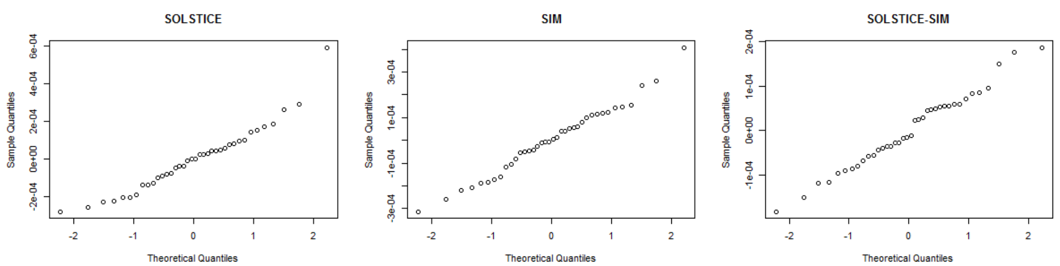

Figure B3. These three Q-Q plots compare the monthly averaged SOLSTICE, SIM and SOLSTICE-SIM differences to theoretical Gaussian distributions. The roughly linear relationship shows that the three datasets do behave in a close to Gaussian nature and thus the use of Eq. (2) and Eq. (3) are supported for the analyses presented in this paper.

Appendix C: Comments on the interpretation of time estimates.

Any estimate of how long it will take to correctly identify a drift must be taken with some level of understanding of how this estimate is made and what can be expected from using these estimates. Figure 3 offers estimates for a range of times needed to estimate specific drifts, assuming no jumps occur in the record. As a reminder, this plot was created assuming the type of overlap seen in the SOLSTICE-SIM overlap period; specifically, the calculations assume the amount of variability and autocorrelation observed in the differences (shown visually in the third plot of Fig. 2). However, different observing systems are likely to have different levels of agreement.

When deriving trends on existing data, only a single level of certainty is required: for example, what does it mean to detect a drift? Often the community has focused on detection at a 95% confidence level or a 99% confidence level. However, when estimating how long it will take to detect a drift, two statistical levels are required: one that identifies what is meant by "detecting a drift" and one that identifies the likelihood that a drift will be detected in the specified period of time, if that level of drift is the true, long-term drift in the overlap. For the first, we consider detecting a drift to mean identifying a drift that, with 95% likelihood, is not zero, although other levels may be considered. For the second, we consider the likelihood of detecting the drift (at the 95% confidence level) to be 50%. To be clear, we may detect the drift, if it is real, a few months earlier or a few months later than the estimated time.

There are no error bars in Fig. 3. We'd like to begin the discussion of appropriate error bars in this section. As stated in the previous paragraph, the data in Fig. 3 represent estimates of how long it will take to detect a specific level of drift. If we focus on a single point, for instance the two year point that indicates a drift of $1.2*10^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$ could be detected, it is possible that a slightly smaller drift could be detected in that two year of overlap, if the variability happens to result in the signal-to-noise for the overlap period is slightly more favorable. Similarly, if the actual, underlying drift is actually five times as large ($6*10^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$) it is highly likely the drift would be detectable within the two years. So the "error bars" on this one point would be slightly below the current point and would extend infinitely upward, indicating that much larger drifts could be detected in the two year period.

Extending our discussion of error bars in Fig. 3, we can similarly think in terms of horizontal error bars. Again, focusing on the one point in Fig. 3 indicating that a drift of $1.2*10^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$ could be detected in two years, we can imagine that the detection of this drift, if it is the true underlying drift, may be detectable a few months shy of two years or may take a few months more than two years. As stated above, the two years is a 50% likelihood of detection. It is highly likely that such a drift could not be detected in a few months of monitoring, but would very likely be detected in ten years of monitoring. So, again, we have error bars that are non-standard in that they extend to the left in the plot and continue indefinitely to the right.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

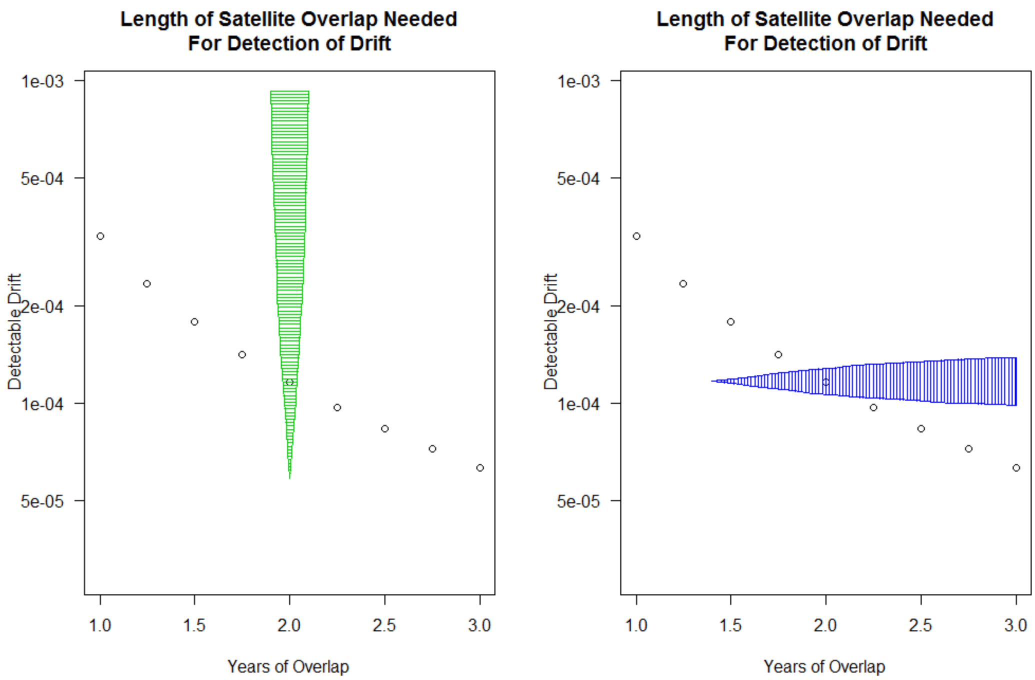

Figure C1. Estimates of how long it will take to detect a drift can be interpreted as the likely time needed. Depending on variability present, the even small drifts can be detected (although with less than 50% likelihood of detection), probability indicated by width of the green area. For a given drift level, there is a chance that the drift can be detected in less than the number of years indicated, although that likelihood is less than 50% for time less than the time indicated, probability indicated by the blue area.

If we want to express this uncertainty of likelihood of detection in a visual manner, we could employ two dimensional error bars, similar to violin plots which are often employed to express variable information. Figure C1 shows the likelihood of detecting a particular drift with two years of overlap. For drifts considerably smaller than $1.2*10^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$, the likelihood of detection with a two year overlap is represented by the width of the green area. Larger drifts can be directed with higher likelihood. Figure C1 shows the likelihood of a true drift of $1.2*10^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$ being detected in less than two years. The height of the blue bar indicates the likelihood, with the likelihood being 50% at two years and considerably higher likelihood of detection as more years of overlap are presented. There is also a small, but less than 50% likelihood that the true drift might be detected in less than two years; again, the height of the blue bars indicates the likelihood of detection.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Atmospheric Chemistry and Physics Discussions
Open Access
EGU

Appendix D: Mathematical structure for optimizing overlap decision choices.

Treating the value of information derived from satellite data as a public good (e.g. weather forecasts and climate services have non-rival and non-excludable characteristics which define public goods in economic theory), total societal benefits is the sum of the benefits realized by all users of the information. Net benefits (NB) is the difference between total benefits (TB) and total costs (TC).

$$\mathbf{NB}=\mathbf{TB}\left\{IVC\left[q\left(o\right)\right]\right\}\text{-}\mathbf{TC}\left[o\right] \qquad \text{(D1)}$$

For purposes of the current discussion we take total costs to simply be a function of the temporal overlap in satellite observations (*o*). Total costs are an increasing function of *o* (i.e., the total costs increase the more the overlap period). On the other hand, total benefits are a more complicated function of the entire process of information creation, communication, use, and decision making (labeled IVC for the Information Value Chain). The benefits of the IVC process are considered to be a function of the quality of the information, *q*, which itself is a function of the temporal overlap in satellite observations (*o*). This is a highly non-linear process and not even necessarily increasing in *q(o)* as may commonly be assumed.

If the objective of "optimizing" satellite observation overlaps is to achieve a specific quality standard, economists would approach this as a cost-effectiveness issue. In this case, the objective is to minimize the costs to achieve the exogenously determined standard. There is no consideration of benefits in this case, *o* is set to achieve $\overline{q}$ so $\mathbf{TB}\left\{IVC\left[\overline{q}\left(o\right)\right]\right\}$. The outcome is not necessarily societally optimal – the standard could be too strict in which case net benefits could be negative or the standard could be too lax in which case it would be possible to improve societal outcomes by increasing the observational overlap period and increase net benefits.[3]

If instead of a fixed standard $\overline{q}$, the objective is to maximize net societal benefit then *o* is chosen to maximize NB. Maximal societal benefits can be identified as a function of the temporal overlap by taking the first derivative of the NB formula with respect to *o*.[4]

$$\partial\mathbf{NB}/\partial o=\partial\,\mathbf{TB}\left\{IVC\left[q\left(o\right)\right]\right\}/\partial o\text{-}\partial\,\mathbf{TC}/\partial o=\frac{\partial\mathbf{TB}}{\partial IVC}\frac{\partial\mathbf{IVC}}{\partial q}\frac{\partial\mathbf{q}}{\partial o}\text{-}\partial\,\mathbf{TC}/\partial o=\mathbf{MB}_o-\mathbf{MC}_o \qquad \text{(D2)}$$

As the terms on the right indicate, the maximum societal benefits are achieved when the marginal benefits (MB_o) of a change in the overlap period are equal to the marginal costs (MC_o) of the overlap.[5] The marginal costs of the overlap period, $\partial\mathbf{TC}/\partial o$, are likely to include additional costs of data collection, assimilation, storage, and analysis and may be a fairly linear function of the length of the overlap period. The marginal benefits of increasing the overlap period though are represented as a more complex relationship between the overlap period and quality of information ($\partial\mathbf{q}/\partial o$), how changes in information quality manifest through in the information value chain ($\partial\mathbf{IVC}/\partial q$), and how changes in the quality of information provided to decision makers may manifest themselves in potential outcomes ($\partial\mathbf{TB}/\partial IVC$). While economists have extensive experience and applications in monetizing the value of potential changes in societal outcomes (e.g. lives saved, reduced damages, improved crop yields in agricultures, etc.), it is generally much more difficult to (1) identify all of the potential stakeholders and potential outcomes and (2) validly and reliably characterize and quantify the information value chain or how it changes.

---

[3] Rather than taking the standard as a given, the question could be how to set the standard to achieve a societally "optimal" outcome in terms of maximizing societal benefits. This is essentially equivalent to the unconstrained optimization of NB.

[4] For now we don't discuss second order conditions for maximization (see Morss et al., 2005 for further clarification on second order conditions for net benefit maximization which relate specifically to the shapes of the cost and benefits functions.

[5] In economics notation it is common to use the subscript to indicate the relevant factor under discussion – in this case MB_o refers to the marginal benefits of *o*, the observational overlap period.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

To dependably quantify societal benefits from satellite observations requires understanding the complex relationship between individual instruments, data streams, modeling, communication, decision making, and potential and actual societal outcomes. This requires understanding stakeholders and processes of information creation, transformation, transmission, and use along the entire information value chain through multiple stakeholders with a variety of objectives, resources, and constraints. Cooke et al. (2014) develop an illustrative example of such an analysis using the social costs of carbon (SCC) as a measure of societal benefits and a hypothetical decision framework (i.e., a tipping point that would lead to global climate impact mitigation efforts).

Team list: E. Weatherhead, E. Araujo-Pradere, J. Harder, L. Flynn, S. Frith, J. Lazo, P. Pilewskie, T. Woods, M. Weber, J. M. English

Author contribution: Jerald Harder conceived of the idea for this paper and the general techniques to be used; he also provided the data for analysis. Elizabeth Weatherhead carried out the calculations, including adapting statistical approaches to the problem and oversaw the writing. Eduardo Araujo worked on the calculations and identified stability requirements for solar and environmental data. Larry Flynn supplied input on GSICS and complementary methods of verifying stability in satellite records as well as cross verification of results, including those presented in Table 1. Stacey Frith and Mark Weber provided valuable input on merging of satellite records and applicability of results to developing long-term records. Jeff Lazo, in consultation with the other co-authors, provided the optimization methodology and economic analysis. Thomas Woods and Peter Pilewskie, as respective Principal Investigators for the NASA SORCE and the upcoming TSIS missions advised on the requirements and needs for mission overlap for the solar spectral irradiance climate record. Jason English supplied input on applicability to climate studies.

Competing interests: There are no known competing interests for this work. The authors declare that they have no conflict of interest.

Disclaimer: None

Acknowledgements: We would like to acknowledge the support of NASA for the production of the data used in this project. NASA contract NAS5-97045, and Miami-Dade College for their research support. M. Weber acknowledges the support of the DFG Research Unit SHARP (Stratospheric Change and its Role for Climate Prediction).

References

Adams, C., Bourassa, A. E., Sofieva, V., Froidevaux, L., McLinden, C. A., Hubert, D., Lambert, J.-C., Sioris, C. E., and Degenstein, D. A.: Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using MLS, GOMOS, and ozonesondes, Atmos. Meas. Tech., 7, 1, 49-64, 2014.

Araujo-Pradere, E. A., Fuller-Rowell, T. J., and Bilitza, D.: Ionospheric variability for quiet and perturbed conditions, Adv. Space Res., 34(9), 1914-1921, doi: 10.1016/j.asr.2004.06.007, 2004.

Araujo-Pradere, E. A., Buresova, D., and Fuller-Rowell, T. J.: Initial results of the evaluation of IRI hmF2 performance for minima 22-23 and 23-24, Adv. Space Res., doi: 10.1016/j.asr.2012.02.010, 2012.

Araujo-Pradere, E. A., Redmon, R., Fedrizzi, M., Viereck, R., and Fuller-Rowell, T. J.: Some characteristics of the ionospheric behavior during solar cycle 23/24 minimum, Solar Phys., The Sun–Earth Connection near Solar Minimum, edited by M. M. Bisi, B. Emery, and B. J. Thompson, doi: 10.1007/s11207-011-9728-3, Springer, 2011.

ASIC3: Achieving Satellite Instrument Calibration for Climate Change (ASIC3). Ohring, G. (Ed.), Available at: http://www.star.nesdis.noaa.gov/star/documents/ASIC3-071218-webversfinal.pdf, 2007.

BenMoussa, A., and forty additional authors: On-orbit degradation of solar instruments, Solar Phys., 288, 389-434, doi: 10.1007/s11207-013-0290-z, 2013.

Best, F., Adler, A. D. P., Ellington, S. D., Thielman, D. J., and Revercomb, H. E.: On-orbit absolute calibration of temperature with application to the CLARREO mission, Proc. SPIE, doi: 10.1117/12.795457, 2008.

Bodeker, G. E., Scott, J. C., Kreher, K., and McKenzie, R. L.: Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978–1998, J. Geophys. Res. Atmos., 106(D19), 23029-23042, 2001.

Bourassa, A. E., Degenstein, D. A., Randel, W. J., Zawodny, J. M., Kyrölä, E., McLinden, C. A., Sioris, C. E., and Roth, C. Z.: Trends in stratospheric ozone derived from merged SAGE II and Odin-OSIRIS satellite observations, Atmos. Chem. Phys., 14, 6983-6994, doi: 10.5194/acp-14-6983, 2014.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M.: Time series analysis: forecasting and control, 4th edition, John Wiley & Sons, 2015.

Brown, S.: Maintaining the long-term calibration of the Jason-2/OSTM advanced microwave radiometer through intersatellite calibration, IEEE T. Geosci. and Remote Sensing, 51, 1531-1543, 2013.

Chander, G., Helder, D. L., Aaron, D., Mishra, N., and Shrestha, A. K.: Assessment of spectral, misregistration and spatial uncertainties inherent in the cross-calibration study, IEEE T. Geosci. and Remote Sensing, 51, 1282-1296, 2013a.

Chander, G., Hewison, T. J., Fox, N., Wu, X., Xiong, X., and Blackwell, W. J.: Overview of intercalibration of satellite instruments, IEEE T. Geosci. and Remote Sensing, 51, 1056-1080, 2013b.

Christy, J. R.: Temperature above the surface layer in long-term climate monitoring by the global climate observing system edited by T. Karl, Vol. 31, Issue 2, pp 455-474, Springer, Netherlands, 1995.

Christy, J. R., Spencer, R. W., and Lobl, E. S.: Analysis of the merging procedure for the MSU daily temperature time series, J. Climate, 11, 8, 2016-2041, 1998.

Christy, J. R., Spencer, R. W., and Braswell, W. D.: MSU tropospheric temperatures: Dataset construction and radiosonde comparisons, J. Atmos. and Ocean. Tech., 17, 9, 1153-1170, 2000.

Cooke, R., Wielicki, B. A., Young, D. F., and Mlynczak, M. G.: Value of information for climate observing systems, Environ. Syst. Decis., 34, 98, doi: 10.1007/s10669-013-9451-8, 2014.

Donaldson, D., and Storeygard, A.: The view from above: applications of satellite data in economics, J. Econ. Perspect, 30(4), 171-98, doi: 10.1257/jep.30.4.171, 2016.

Ducre'-Robitaille, J.-F., Vincent, L. A., and Boulet, G.: Comparison of techniques for detection of discontinuities in temperature series, Int. J. of Climatol., doi: 10.1002/joc.924, 2003.

Dudok de Wit, T.: A method for filling gaps in solar irradiance and solar proxy data, Astron. Astrophys, 533, A29, 2011.

Dudok de Wit, T., Kretzschmar, M., Aboudarham, J., Amblard, P.-O., Auchère, F., and Lilensten, J.: Which solar EUV indices are best for reconstructing the solar EUV irradiance?, Adv. Space Res., 42, 5, 903-911, 2008.

Feldman, D. R., Algieri, C. A., Ong, J. R., and Collins, W. D.: CLARREO shortwave observing system simulation experiments of the twenty-first century: simulator design and implementation, J. Geophys. Res., 116 (D10), doi: 10.1029/2010JD015350, 2011.

Feldman, D. R., Collins, W. D., and Paige, J. L.: Pan-spectral observing system simulation experiments of shortwave reflectance and longwave radiance for climate model evaluation, Geoscientific Model Development, 8, 1943-1954, doi: 10.5194/gmd-8-1943, 2015.

Fox, N., Green, P., Brindley, H., Russell, J., Smith, D., Lobb, D., Cutter, M., and Barnes, A.: TRUTHS (Traceable Radiometry Underpinning Terrestrial and Helio Studies): A mission to achieve climate quality data, Proc. 'ESA Living Planet Symposium 2013', Edinburgh, ESA SP-722, December 2013.

Free, M., Durre, I., Aguilar, E., Seidel, D., Peterson, T. C., Eskridge, R. E., Luers, J. K., Parker, D., Gordon, M., Lanzante, J., Klein, S., Christy, J., Schroeder, S., Soden, B., McMillin, L. M., and Weatherhead, E.: Creating Climate Reference Datasets: CARDS Workshop on adjusting radiosonde temperature data for climate monitoring, Bull. Am. Meteorol. Soc., doi: http://dx.doi.org/10.1175/1520-0477(2002)083<0891:CCRDCW>2.3.CO;2, 2002.

Frith, S. M., Kramarova, N. A., Stolarski, R. S., McPeters, R. D., Bhartia, P. K., and Labow, G. J.: Recent changes in total column ozone based on the SBUV Version 8.6 merged ozone data set, J. Geophys. Res. Atmos., 119(16), 9735-9751, 2014.

Fruit, M., Gusarov, A., and Doyle, D.: Testing and qualification of optical glasses for use in a space radiation environment; the advantages and pitfalls of using a parametric approach, Proc. SPIE Int. Soc. Opt. Eng., 4823, 132, 2002.

Fröhlich C.: Evidence of a long-term trend in total solar irradiance, Astron. Astrophys., 501, L27-L30, doi: 10.1051/0004-6361/200912318, 2009.

Hamilton, J. D.: Time series analysis, vol. 2. Princeton: Princeton University Press, 1994.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Harder J. W., Lawrence, G., Fontenla, J., Rottman, G. J., and Woods, T. N.: The Spectral Irradiance Monitor: scientific requirements, instrument design, and operation modes, Solar Phys., 230, 141-167, 2005a.

Harder J. W., Fontenla, J., Lawrence, G., Woods, T., and Rottman, G.: The Spectral Irradiance Monitor: Measurement (SIMM) equations and calibration, Solar Phys., 230, 169-204, 2005b.

Harder, J. W., Thuillier, G., Richard, E. C., Brown, S. W., Lykke, K. R., Snow, M., McClintock, W. E., Fontenla, J. M., Woods, T. N., and Pilewskie, P.: The SORCE SIM Solar Spectrum: comparison with recent observations, Solar Phys., 263, 3-24, doi: 10.1007s11207-010-9555-y, 2010.

Hewison, T. J., Wu, X., Yu, F., Tahara, Y., Hu, X., Kim, D., and Koenig, M.: GSICS Inter-calibration of infrared channels of geostationary imagers using Metop/IASI, IEEE T. Geosci. and Remote Sensing, 15 (3), 2013.

Hood, L. L., and Zhou, S.: Stratospheric effects of 27-day solar ultraviolet variations: An analysis of UARS MLS ozone and temperature data, J. Geophys. Res., 103(D3), 3629-3638, doi: 10.1029/97JD02849, 1998.

Hurrell, J., and Trenberth, K.: Spurious trends in satellite MSU temperatures from merging different satellite records, Nature, 386, doi: 10.1038/386164a0, 1997.

Jaxk, R., Chen, J. L., Wang, X. L., Lund, R., and Lu, Q. Q.: A review and comparison of changepoint detection techniques for climate data, J. Appl. Meteorol. Climat., doi: http://dx.doi.org/10.1175/JAM2493.1, 2007.

JCGM – Joint Committee for the Guides in Metrology, JCGM 101:, available at: http://www.bipm.org/en/publications/guides/gum.html, 2008.

Karl, T. R., and Williams, C. N.: An approach to adjusting climatological time series for discontinuous inhomogeneities, J. Appl. Meteorol. Climat., doi: http://dx.doi.org/10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2, 1987.

Karl, T. R., Williams Jr., C. N., Young, P. J., and Wendland, W. M.: A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States, J. Climate Appl. Meteorol., 25, 145-160, 1986.

Laxminarayan, R., and Macauley, M. K. (Eds.): The value of information: methodological frontiers and new applications in environment and health. Springer Netherlands. 304 pages, doi: 10.1007/978-94-007-4839-2, 2012.

Loeb, N. G., Nanalo-Smith, N., Su, W., Shankar, M., and Thomas, S.: CERES top-of-Atmosphere Earth radiation budget climate data record: Accounting for in-orbit changes in instrument calibration, Remote Sens., 8, 182, doi: 10.3390/rs8030182, 2016.

Lukashin, C., Wielicki, B. A., Young, D. F., Thome, K., Jin, Z., and Sun, W.: Uncertainty estimates for imager reference inter-calibration with CLARREO reflected solar spectrometer, IEEE T. Geosci. and Remote Sensing, 2013.

Macauley, M. K.: The value of information: measuring the contribution ofspace-derived earth science data to resource management, Space Policy, 22(4), 274-282, 2006.

MacDonald, A. E.: A global profiling system for improved weather and climate prediction, Bull. Am. Met. Soc., 86, 1747, doi: 10.1175/BAMS-86-12-1747, 2005.

Marchenko, S. V., and DeLand, M. T.: Solar spectral irradiance changes during cycle 24, The Astrophys. J., 789, 2, 117, doi: 10.1088/0004-637X/789/2/117, 2014.

McClintock W. E., Rottman G. J., and Woods T. N.: Solar-Stellar Irradiance Comparison Experiment II (SOLSTICE II): Instrument concept and design, Solar Phys., 230, 225-258, 2005a.

McClintock W. E., Snow, M., and Woods, T. N.: Solar-Stellar Irradiance Comparison Experiment II (SOLSTICE II): Pre-launch and on-orbit calibrations, Solar Phys., 230, 259-294, 2005b.

Mitchell, T. D., and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, Int. J. of Climatol., doi 10.1002/joc.1181, 2005.

Morss, R., Miller, K. A., and Vasil, M. S.: A systematic economic approach to evaluating public investment in observations for weather forecasting, Mon. Weather Rev., 133, 374-388, 2005.

National Research Council, Issues in the integration of research and operational satellite Ssystems for climate research: Part I. science and design. Committee on Earth Studies, Space Studies Board, NRC, 2000a.

National Research Council, Reconciling observations of global temperature change, National Academy Press, Washington, D.C., 2000b.

NISTIR 7047, Satellite Instrument Calibration for Measuring Global Climate Change. Ohring, G., Wielicki, B., Spencer, R., Emery, W., Datla, R. (Eds). Available at: https://www.nist.gov/sites/default/files/documents/pml/div685/pub/nistir7047.pdf, 2004.

Ohring, G., Wielicki, B., Spencer, R., Emery, B., and Datla, R.: Satellite instrument calibration for measuring global climate change: report of a workshop, Bull. Am. Meteorol. Soc., 86, 1303-13, doi: 10.1175/BAMS-86-9-1303, 2005.

Pagaran J., Harder, J. W., Webber, M., Floyd, L. E., and Burrows, J. P.: Intercomparison of SCIAMACHY and SIM vis-IR irradiance over several solar rotational timescales, Astron. Astrophys., 528, A67, 2011.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Forland, E. J., Hanssen Bauer, I., Alexandersson, H., Jones, P., and Parker, D.: Homogeneity adjustments of in situ atmospheric climate data: a review, Int. J. Climatol., 18, 1493-1517, 1998.

Rahpoe, N., Weber, M., Rozanov, A. V., Weigel, K., Bovensmann, H., Burrows, J. P., and Laeng, A.: Relative drifts and biases between six ozone limb satellite measurements from the last decade, Atmos. Meas. Tech., 8, 10, 4369-4381, 2015.

Randel, W. J., and Thompson, A. M.: Interannual variability and trends in tropical ozone derived from SAGE II satellite data and SHADOZ ozonesondes, J. Geophys. Res., 116, D07303, doi: 10.1029/2010JD015195, 2011.

Salby, M., and Callaghan, P.: Sampling error in climate properties derived from satellite measurements: consequences of undersampled diurnal variability, J. Climate, 10, 18-36, 1997.

Santer, B. D., Wigley, T. M. L., Meehl, G. A., Wehner, M. F., Mears, C., Schabel, M., Wentz, F. J., Ammann, C., Arblaster, J., Bettge, T., Washington, W. M., Taylor, K. E., Boyle, J. S., Brüggemann, W., and Doutriaux, C.: Influence of satellite data uncertainties on the detection of externally forced climate change, Science, doi: 10.1126/science.1082393, 2003.

Schrijver, C. J., Livingston, W. C., Woods, T. N., and Mewaldt, R. A.: The minimal solar activity in 2008–2009 and its implications for long term climate modeling, Geophys. Res. Lett., 38, 6, 2011.

Slater, P. N., Biggar, S. F., Thome, K. J., Gellman, D. I., and Spyak, P. R.: Vicarious radiometric calibrations of EOS sensors, J. Atmos. and Ocean. Tech., 13, 2, 349-359, 1996.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006), J. Climate, 21, 10, 2283-2296, 2008.

Snow, M., McClintock, W. E., Rottman, G., and Woods, T. N.: Solar-Stellar Irradiance Comparison Experiment II (SOLSTICE II): Examination of the Solar-Stellar Comparison Technique, Solar Phys., 230, 295-324, 2005.

Staehelin, J., Kerr, J., Evans, R., and Vanicek, K.: Comparison of total ozone measurements of Dobson and Brewer spectrophotometers and recommended transfer functions, Tech. Rep., WMO, World Meteorological Organization Global Atmosphere Watch (WMO-GAW) Report 149. http://library.wmo.int/pmb_ged/wmo-td_1147.pdf, 2003.

Stolarski, R., and Frith, S.: Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data record: the importance of instrument drift uncertainty, Atmos. Chem. Phys., 6, 4057-4065. doi: 10.5194/acp-6-4057, 2006.

Thorne, P. W., Parker, D. E., Christy, J. R., and Mears, C. A.: Uncertainties in climate trends: lessons from upper-air temperature records, Bull. Am. Meteorol. Soc., 86, 10, 1437-1442, 1398-1399, 2005.

Thuillier G., Floyd, L., Woods, T. N., Cebula, R., Hilsenrath, E., Hers, M., and Labs, D.: Solar irradiance reference spectra for two solar active levels, Adv. Space Res., 34, 256-261, 2004.

Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-1156, 2017
Manuscript under review for journal Atmos. Chem. Phys.
Published: 20 January 2017

Atmospheric Chemistry and Physics
Discussions

Tobin, D., Holz, R., Nagle, F., and Revercomb, H.: Characterization of the climate absolute radiance and refractivity observatory (CLARREO) ability to serve as an infrared satellite intercalibration reference, J. Geophys. Res., 121 (8), 2016.

Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., and Brohede, S.: Characterizing sampling biases in the trace gas climatologies of the SPARC Data Initiative, J. Geophys. Res. Atmos., 118, 20, 2013.

Viereck, R. A., Floyd, L. E., Crane, P. C., Woods, T. N., Knapp, B. G., Rottman, G., Weber, M., Puga, L. C., and DeLand, M. T.: A composite Mg II index spanning from 1978 to 2003, Space Weather, 2, S10005, doi: 10.1029/2004SW000084, 2004.

Vincent, L.: A Technique for the identification of inhomogeneities in Canadian temperature series, J. Climate, doi: http://dx.doi.org/10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2, 1998.

Vincent, L. A., Zhang, X., Bonsal, B. R., and Hogg, W. D.: Homogenization of daily temperatures over Canada, J. Climate, 15, 1322-1334, 2002.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Jackman, C. H., Bishop, L., Hollandsworth Frith, S. M., DeLuisi, J., Keller, T., Oltmans, S. J., Fleming, E. L., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Herman, J., McPeters, R., Nagatani, R. M., and Frederick, J. E.: Detecting the recovery of total column ozone, J. Geophys. Res. Atmos., 105, D17, doi: 10.1029/2000JD900063, 2000.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X. L., Choi, D., Cheang, W. K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E.: Factors affecting the detection of trends: statistical considerations and applications to environmental data, J. Geophys. Res., 103 D14, 1998.

Weatherhead, E. C., Wielicki, B., and Ramaswamy, V.: Climate observing system simulation experiments. AGU Fall Meeting, San Francisco, California, Dec. 14-18, 2015.

Weber, M., Rahpoe, N., and Burrows, J.: Stability requirements on long-term (satellite) ozone observations and their implications for trend detection, QOS, Edinbrugh, September, 2016.

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

Wielicki, B. A., Young, D. F., Mlynczak, M. G., Thome, K. J., Leroy, S., Corliss, J., and Anderson, J. G.: Achieving climate change absolute accuracy in orbit, Bull. Am. Meteorol. Soc., 94, 10, 1519-1539, 2013.

Willson, R. C., and Mordvinov, A. V.: Secular total solar irradiance trend during solar cycles 21–23, Geophys. Res. Lett., 30(5), 1199, doi: 10.1029/2002GL016038, 2003.

Willson, R. C., and Hudson, H. S.: The sun's luminosity over a complete solar cycle, Nature, 351, 42-44, 1991.

WMO Guide to Climatological Practices WMO-No. 100, ISBN 978-92-63-10100-6. Available at http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/WMO_100_en.pdf, 2011a.

WMO Guide to Climatological Practices, Third edition, Technical specification for the evolution and future hosting of the WMO Database of Observational user requirements and observing system capabilities, 2011b.

WMO/TD-No. 341, Calculation of Monthly and Annual 30-years standard normals (prepared by a meeting of experts, Washington, March 1989) [WCDP-No. 10], 1989.

Wu, X., Hewison, T., and Tahara, Y.: GSICS GEO-LEO intercalibration: baseline algorithm and early results, SPIE Proc., 7456, doi: 10.1117/12.825460, 2009.

Wulfmeyer, V., Hardesty, M., Turner, D., Behrendt, A., Cadeddu, M., Di Girolamo, P., Schlüssel, P., van Baelen, J., and Zus, F.: A review of the remote sensing of lower-tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles, Rev. Geophys., doi: 10.1002/2014RG000476, 2015.

Zou, C. Z., and Qian, H.: Stratospheric temperature climate data record from merged SSU and AMSU-A Observations, J. Atmos. and Ocean. Tech., 2016.