

1 How long do satellites need to overlap? Evaluation of climate data 2 stability from overlapping satellite records

3
4 Elizabeth C. Weatherhead¹, Jerald Harder², Eduardo A. Araujo-Pradere³,
5 Greg Bodeker⁴, Jason M. English⁵, Lawrence E. Flynn⁶, Stacey M. Frith⁷, Jeffrey K Lazo⁸, Peter
6 Pilewskie², Mark Weber⁹, Thomas N. Woods²

7
8 ¹University of Colorado, Boulder, Colorado, USA

9 ²Laboratory for Atmosphere and Space Physics, University of Colorado, Boulder, Colorado, USA

10 ³School of Science, Miami Dade College, Miami, Florida, USA

11 ⁴Bodeker Scientific, Alexandra, New Zealand

12 ⁵Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA /
13 NOAA Earth System Research Laboratory, Global Systems Division, 325 Broadway, Boulder, CO 80305

14 ⁶NOAA, NESDIS, College Park, Maryland, USA

15 ⁷Science Systems and Applications, Inc., Lanham, MD USA

16 ⁸National Center for Atmospheric Research, Boulder, Colorado, USA

17 ⁹University of Bremen FB1, Bremen, Germany

18
19 *Correspondence to:* Elizabeth Weatherhead (Betsy.Weatherhead@Colorado.edu)

20

21 **Abstract.** Sensors on satellites provide unprecedented understanding of the Earth's climate system by measuring
22 incoming solar radiation, as well as both passive and active observations of the entire Earth with outstanding spatial
23 and temporal coverage. A common challenge with satellite observations is to quantify their ability to provide well-
24 calibrated, long-term, stable records of the parameters they measure. Ground-based intercomparisons offer some
25 insight, while reference observations and internal calibrations give further assistance for understanding long-term
26 stability. A valuable tool for evaluating and developing long-term records from satellites is the examination of data
27 from overlapping satellite missions. This paper addresses how the length of overlap affects the ability to identify an
28 offset or a drift in the overlap of data between two sensors. Ozone and temperature datasets are used as examples
29 showing that overlap data can differ by latitude and can change over time. New results are presented for the general
30 case of sensor overlap by using SORCE SIM and SOLSTICE solar irradiance data as an example. To achieve a 1%
31 uncertainty in estimating the offset for these two instruments' measurement of the Mg II core (280 nm) requires
32 approximately 5 months of overlap. For relative drift to be identified within 0.1% per year uncertainty (0.00008
33 watts m⁻² nm⁻¹ yr⁻¹, the overlap for these two satellites would need to be 2.5 years. Additional overlap of satellite
34 measurements is needed if, as is the case for solar monitoring, unexpected jumps occur adding uncertainty to both
35 offsets and drifts; the additional length of time needed to account for a single jump in the overlap data may be as
36 large as 50% of the original overlap period in order to achieve the same desired confidence in the stability of the
37 merged dataset. Results presented here are directly applicable to satellite Earth observations. Approaches for Earth
38 observations offer additional challenges due to the complexity of the observations but Earth observations may also
39 benefit from ancillary observations taken from ground-based and in situ sources. Difficult choices need to be made
40 when monitoring approaches are considered; we outline some attempts at optimizing networks based on economic
41 principles. The careful evaluation of monitoring overlap is important to the appropriate application of observational
42 resources and to the usefulness of current and future observations.

43
44 **Keywords:** Satellite overlap, satellite monitoring, instrument intercomparison, instrument stability, climate records,
45 trend detection, homogenization of datasets, solar spectral irradiance, ozone.

46

47

1 1. Introduction

2 Stable, long-term time series of environmental data are critical to the ongoing investigation and understanding of the
3 environment. One of the fundamental requirements for construction of long-duration climate records is the ability to
4 analytically assess the characteristics of time series of different sensors so they can be combined into a single
5 reliable record. This need is particularly valuable for satellite-borne sensors that are susceptible to a wide variety of
6 sensitivity degradation mechanisms influenced by the space environment as well as by spacecraft, instrument, and
7 operational considerations. Many of these influences can contribute in unexpected ways to the overall instrument
8 stability thereby adding non-geophysical trending or structure to the combined data records compiled from multiple
9 missions. These problems can also be exacerbated when comparing instruments with different time histories, for
10 instance the comparison of two sensors – one during its early-orbit phase with one that has been in space for an
11 extended length of time. For overlapping spaced-based observations, even with reliable on-board calibration and
12 degradation correction schemes, time-limited intercomparison campaigns are important to objectively identify
13 potential systematic errors in one or both instruments. A variety of techniques exist for merging datasets from
14 different sources – including two different satellites – using statistical models, physical models, and efforts at in situ
15 calibration (e.g. Chander et al., 2013b; Peterson et al., 1998). Each technique has great strengths and can offer not
16 just adjustments for merging of datasets, but estimates of uncertainty in long-term stability. Weber et al. (2016)
17 addressed the issue of requirements on stability for detecting a desired long-term trend from a multiple instrument
18 time series by accounting for variations in instrumental lifetime and merging biases in a Monte Carlo simulation.
19 Another approach to addressing satellite uncertainty, based on maintaining traceability through on-board calibration
20 capabilities using absolute references, has been advocated through the CLARREO and TRUTHS programs
21 (Wielicki et al., 2013; Fox et al., 2013). For both programs, verification of merging of these new approaches will be
22 important for validation of expected agreement. For both current as well as proposed satellite systems, the
23 instrument scientists will decide how to use all available information to make corrections or assign uncertainty
24 estimates to the data should an offset or drift be detected in overlapping satellite records. Without sufficient overlap
25 there is a limit to the magnitude of offset or drift that can be detected. This paper presents techniques that can
26 address the stability of merged data records using observations from overlapping satellite instruments.

27 Measuring the small changes associated with long-term global climate change from space is both extremely
28 important and particularly challenging. For example, the satellite instruments must be capable of observing
29 atmospheric and surface temperature trends as small as $0.1^{\circ}\text{C dec.}^{-1}$, ozone changes as little as $1\% \text{ dec.}^{-1}$, and
30 variations in the Sun's output as tiny as $0.1\% \text{ dec.}^{-1}$ (Ohring et al., 2005). A particular challenge in the design of
31 climate observing systems is how to preserve data quality and facilitate appropriate evaluations of observations that
32 extend over a series of missions measuring the same geophysical quantity. A number of in-depth techniques are used
33 by instrument scientists to understand the fundamental (Level 1) observations, including wavelength scale
34 corrections, detector responsivity evaluation, and field-of-view sensitivity monitoring. With the regular insertion of
35 new technology driven by interest in reducing costs and/or improving performance also comes the need to separate
36 the effects of changes in the Earth system from effects ascribable to changes and gaps in the observing system.
37 Credible, ongoing programs of sensor calibration and validation, sensor characterization, data continuity, and
38 strategies for ensuring overlap across successive sensors are thus essential (NRC, 2000a). Multiple efforts describing
39 key challenges and/or requirements have been published (Chander et al., 2013a; Fröhlich, 2009; Willson and
40 Hudson, 1991; Willson and Mordvinov, 2003). Adams et al. (2014) revealed up to a 6% relative drift per decade
41 between different ozone observing satellites, confounding some attempts to detect signs of ozone recovery. Rahpoe
42 et al. (2015), Hubert (2016) and Tegtmeier et al. (2013) all show that both drifts and biases in current satellite
43 observing systems are often large compared to the signals of interest. The approaches presented in this paper focus
44 on developing useful checks on the final data products (Level 2) from multiple instruments.

45 In the last two decades, there has been an increasing understanding that the merging of records, and the uncertainty
46 associated with that merging, cannot be considered independently of the final use of the data. NRC (2000a)
47 highlighted the need for precise inter-satellite calibration, recommending that there should be a 1-year overlap
48 between successive Ozone Monitoring Profiler Suite missions to allow sensor intercomparison and guarantee long-
49 term traceability. Analogously, a 1-year overlap in observations of both solar irradiance and spectral solar irradiance
50 is part of the summary recommendations of Ohring et al. (2005). NRC (2000b) concluded that a special effort is
51 required to preserve the quality of data acquired with different satellite systems and sensors, so that valid
52 comparisons can be made over an entire set of observations. Randel and Thompson (2011) explored the utility of
53 combining the SAGE II ozone observations with tropical measurements from the SHADOZ ozonesonde network, to
54 study interannual variability and trends. However, not all satellite records have the benefit of such long-term in situ

1 dataset for intercomparison. Bourassa et al. (2014) quantified interannual variability and decadal trends by
2 combining stratospheric ozone profile measurements from different satellite systems including using the
3 Stratospheric Aerosol and Gas Experiment (SAGE) II satellite instrument (1984–2005) with measurements from the
4 Optical Spectrograph and InfraRed Imager System (OSIRIS) instrument on the Odin satellite (2001–present), noting
5 significant differences between the different observational sets. These studies indicate that a more robust
6 understanding of our data records is essential to meeting requirements and making appropriate use of the final data.

7 Multiple efforts are ongoing internationally to assure that emerging ground-based, in situ and satellite records can be
8 useful to climate analyses, most notably the Global Space-based Inter-Calibration System (GSICS) which is a joint
9 effort by WMO and the Coordination Group for Meteorological Satellites (CGMS) to monitor and harmonize data
10 quality from operational weather and environmental satellites. Harmonized datasets often require adjusting for
11 offsets, spurious drifts and instrument or location-specific problems (Salby and Callaghan, 1997; Araujo-Pradere et
12 al., 2011; Dudok de Wit, 2011). To improve the precision and usefulness of multi-instrument time series for
13 identifying biases, it is necessary to remove offsets between data sources, including those resulting from (a)
14 calibration differences; (b) spatial and temporal sampling or resolution differences; (c) changes in data processing
15 versions; (d) inherently different spectral sensitivities; (e) different instrument types with varying inherent vertical
16 coordinates; and (f) changes in instrument orientation or orbital characteristics or collection times; as examples see
17 Chander et al. (2013b) and Toohey et al. (2013). These potential problems are further exacerbated by temporal gaps
18 or insufficient overlap in the satellite records.

19 In this paper, we estimate the direct impact of length of overlap between satellites to the continuity of data from two
20 overlapping satellites. We examine three separate factors that are of direct importance to the users and creators of
21 merged satellite datasets: the quantified offset of the two datasets, the drift between the two datasets, and the impact
22 of sudden jumps in the data during periods of overlap. We note that intercomparison of satellite records cannot, in
23 isolation, determine which of two systems is more accurate or stable. Indeed, agreement of two observing systems
24 can occur when both are similarly inaccurate or similarly drifting and instruments can drift outside of the
25 intercomparison periods. However, intercomparisons offer valuable, independent assessment useful for developing a
26 long-term record. For illustrative purposes, we look at ozone, temperature and solar radiation satellite records and
27 discuss how these three factors can affect the long term records of these parameters. We present techniques for
28 evaluating overlapping data with the solar dataset because the data are less dependent on satellite drift, diurnal
29 match-ups and differences in the instrument-dependent field of view. We note that the usefulness of overlapping
30 data is highly dependent on the length of overlap and the ability to match overlapping data with high precision. In
31 the final section of this paper, we outline methods for optimizing the set of choices which are needed to create a
32 long-term and stable climate record under a variety of constraints, most notably economic constraints. Optimization
33 will result in better use of resources to achieve more accurate and stable merged datasets.

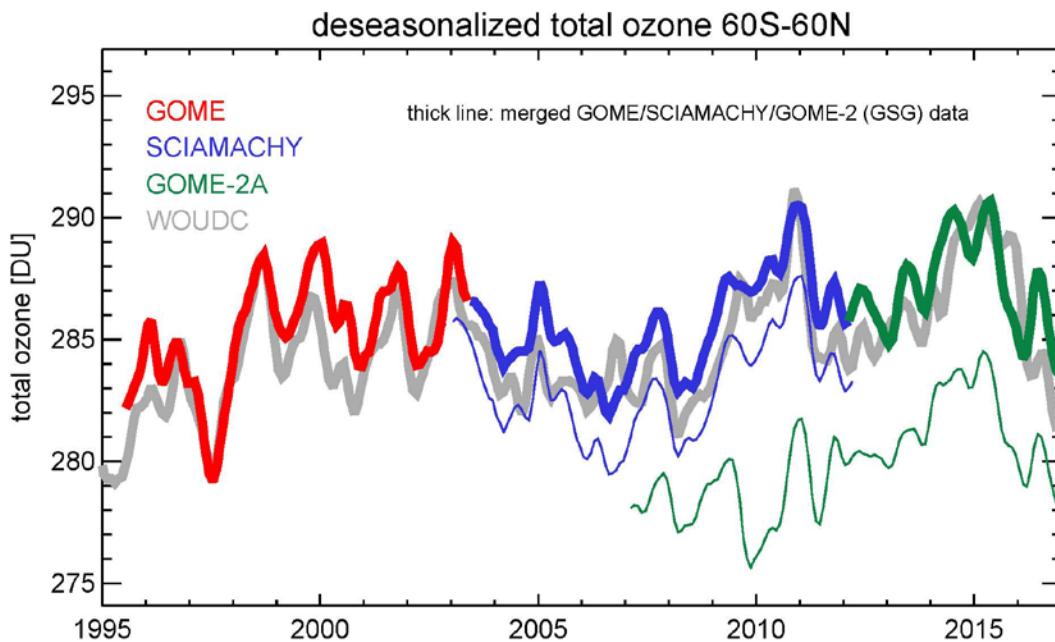
34 **1.1 Overlap of Earth Observation Satellites**

35 The value of satellite observations to understanding the variations, climatologies and changes in the Earth's
36 atmosphere has been profound. Temperature, ozone, water vapor, aerosols and carbon are now understood in ways
37 unimaginable compared to the pre-satellite era. Continued development of new technologies, including sensor
38 development, calibration capabilities and refinement of occultation techniques, has resulted in continued
39 improvements in our observing systems. However, the challenge remains to merge the observations from these
40 evolving systems into scientifically and societally useful observations.

41
42 *Ozone*
43 Some of the most studied satellite records are the internationally sponsored ozone records. (WMO, 2014 and
44 reference within, Staehelin et al, 2003). Since 1978, there have been near-continuous space-based observations of
45 ozone profiles from a combination of missions. Temporal overlaps between these instruments have allowed detailed
46 intercomparisons to play a key role in assessing the precision, accuracy, and long-term drift of the instruments
47 (WMO, 1989, 2011a, 2011b; Bodeker, 2001). However, these overlaps have been somewhat serendipitous; little
48 commitment has been made to ensure the continuity and long-term traceability of the full ground-based, in situ and
49 satellite ozone measurements. The satellite ozone records, as with the temperature records, benefit from multiple
50 observations--both satellite and in situ--over the past four decades (Eckert et al., 2014; Staehelin et al., 2003),

1 resulting in insights into the delivered accuracy and stability of satellite measurements.

2



3

4

5 **Figure 1.** The thick and multi colored line is the merged total ozone satellite record (GSG data) from GOME,
6 SCIAMACHY and GOME-2A (Weber et al., 2011, 2016a). The thin blue and green lines indicate the original data before
7 offsets were corrected to create a continuous record. The success of the analytical effort to combine these three satellite
8 records is confirmed by the good agreement of the merged dataset with the zonal mean data derived from ground based
9 Brewer and Dobson data as part of the WOUDC (World Ozone and UV Data Center, update from Fioletov et al., 2002)
10 which is presented as the thick grey line.

11

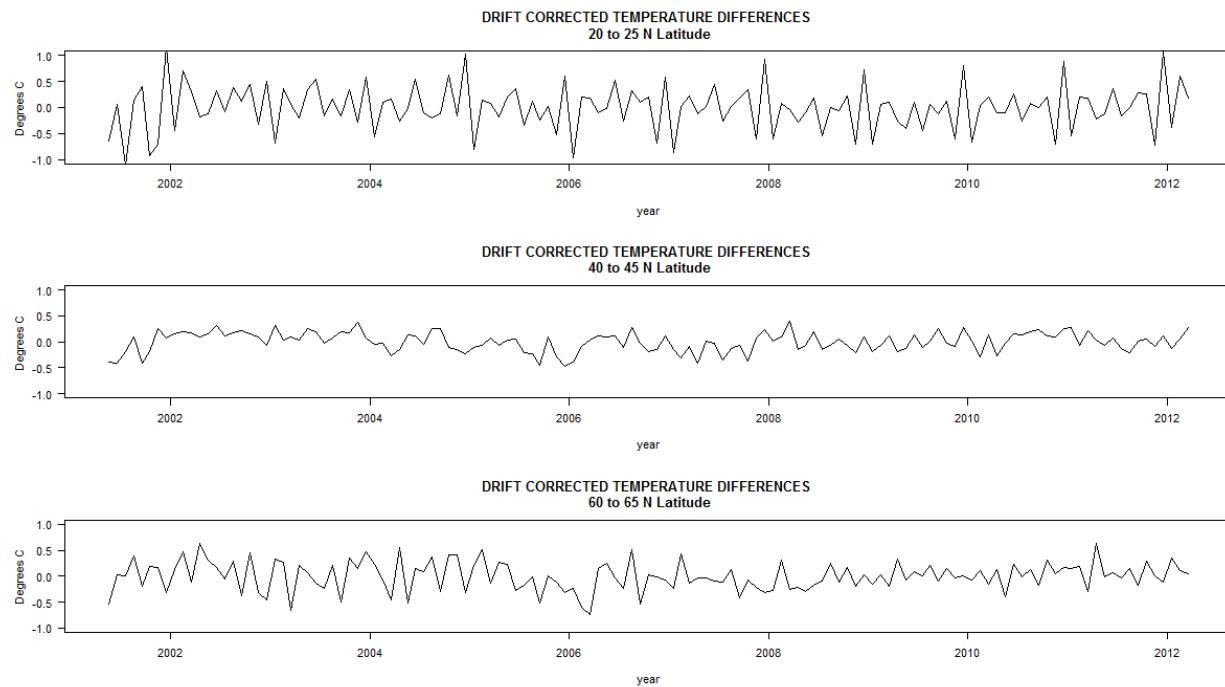
12 Figure 1 shows the result of merging the GOME, SCIAMACHY and GOME-2A total ozone time series (Weber et
13 al., 2011, 2015) into a continuous time series. In this case, the SCIAMACHY and GOME-2A observations (thin
14 blue and green lines) were successively bias adjusted to be continuous with the original GOME data. Biases (offsets)
15 were determined as a function of latitude in steps of 1 degree using monthly zonal means. Despite extensive pre-
16 calibration efforts and monitoring of instrument performance, differences are noted between data from the
17 overlapping satellites. There appears a drop of the original GOME-2 data record during the 2009-2011 period
18 relative to SCIAMACHY, which seems to be larger than the overall bias between two datasets. In this case the very
19 large overlap period from 2007 until 2012 was an advantage and no further corrections beyond the latitude
20 dependent biases were needed to adjust GOME-2. Due to this non-physical drop in the GOME-2A data, the
21 SCIAMACHY data became the preferred choice in the merged (GSG) dataset during the overlap period (2007-
22 2011). In contrast the overlap period for SCIAMACHY and GOME was less than 10 months (2002-2003).

1 Additional corrections beyond a simple bias are difficult and may require the use of external reference data,
 2 although the need for additional corrections may be indicated from satellite overlap data. The long-term stability of
 3 these data are critical for estimating ozone recovery and understanding the complex long-term factors affecting
 4 stratospheric ozone. Lessons from these overlap data serve to offer guidance for future decisions on satellite
 5 observations and overlap periods.

6 *Temperature*

7 Perhaps no other set of satellite records has been as studied as the temperature records derived from the MSU and
 8 AMSU satellites. Two distinct challenges complicating the algorithms needed to develop reliable long-term
 9 temperature records are: 1) multiple satellites, in situ and ground based measurements are available each with unique
 10 characteristics and 2) the level of agreement differs with latitude and altitude. Multiple sources of data can
 11 complicate merged data sets because different choices, even when reasonable, can lead to different long-term
 12 characteristics in the record (e.g., Thorne et al., 2005; also see Sec. 3 of this paper). However, multiple records also
 13 allow different groups to produce independent merged data sets which have long overlaps and can be directly
 14 compared. Through these comparisons we gain valuable information about the uncertainties that arise from the
 15 merging process itself, and whether the data sets are stable relative to the requirements of the analysis.

16



17
 18

19 **Figure 2** Monthly differences between a merged MSU Channel 4 –AMSU Channel 9 satellite temperature record and a
 20 second merged temperature record based on two satellite-based mid-infrared spectrometers (ACE-FTS and MIPAS), a
 21 microwave sounder (SMR), and three satellite-based radio occultation experiments (GRACE, CHAMP, and TSX)
 22 (Penckwitt et al., 2015). Shown are the differences once mean offsets and long-term drifts for each latitude band are
 23 removed through a statistical regression, in degrees C. The top plot is for 20-25N; the middle plot is for 40-45N and the
 24 bottom plot is for 60-65N. These plots show the variability in overlap is highly dependent on latitude, as is often the case
 25 with Earth observations.

26 Figure 2 shows the residuals from two independently merged data sets; the results show notable month-to-month
 27 differences between a merged dataset of MSU Channel 4 and AMSU Channel 9 monthly averaged deseasonalized
 28 data compared to a combined dataset from six satellite instruments providing high vertical resolution temperature
 29 data: MIPAS, ACE-FTS, SMR, GRACE, CHAMP and TSX. Both channels are designed to observe the lower
 30 stratosphere. A full description of this merged temperature record and comparisons with the MSU/AMSU merged
 31 record is given in Penckwitt et al. (2015). The comparisons of overlapping data, after offset and drift have been

1 removed, show several relevant features. First, the variability in the overlapping data varies significantly by latitude.
2 Second, differences in the merged data sets can be remarkably large -- over a half degree C for latitudinal averages -
3 - despite differences between data sets being minimized in the merging process and offset and drift between the
4 compared data set being removed. Third, even when the linear drift over the length of the overlap is removed, the
5 data show apparent drifts that last for several years in each latitude band. Such variations can limit the usefulness of
6 the merged records, but can also highlight issues with particular data sources that can then be addressed. The
7 latitudinal dependence of the variability may indicate regions that are better suited for analysis than others, though in
8 all cases the physical reasons for the correlated variability and potential drifts needs to be carefully examined. The
9 stability levels of satellite temperature datasets are critical for understanding the merged and complex feedbacks that
10 determine regional long-term temperature changes; understanding apparent offsets and drifts between different
11 sources of information is important, particularly when they are large compared to expected trends.

12 **1.2 Planning for needed homogeneity**

13 Detecting and understanding long-term changes require some of the most challenging stability criteria in order for
14 confidence to be placed on the final results. A number of individuals and coordinated groups have worked to define
15 the requirements for Earth observations, including the recently completed effort by WMO (2011b), which addresses
16 the stability needed for various parameters. In the absence of explicit requirements for limits on drift, we suggest
17 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.
18 For example, if a monitoring system is designed to detect a trend of 0.2 degrees per decade, the unchecked drift of
19 the system should be less than 0.1 degrees per decade at the one sigma level. While for Earth sciences, the projected
20 trend is dependent on the climate model and assumptions used to estimate future trends – as well as the location and
21 time of projected trends--this estimate can be used as a starting point for discussions on how well the drift should be
22 confined. When the verification of drift cannot be held to the level of projected trends, there can be serious questions
23 as to the usefulness of the monitoring system for trend identification.

24 The user communities, including the climate community, continue to request high stability from satellite
25 observations (WMO, 2011a; Wulfmeyer et al., 2015; Ohring et al., 2005). A variety of ideas have been offered to
26 allow for more accurate satellite observations: on-board calibration, independent verification, and in-depth modeling
27 of instrument performance can all assist in characterizing the accuracy, biases, and stability of the satellite
28 measurements, many based on the fundamental measurement equations and availability of internal instrument
29 monitoring. There is a long and valuable history of efforts to attempt in-flight calibration, particularly on multi-
30 spectral sensors (Slater et al., 1996). GSICS coordinates the development of tools to intercompare different Earth
31 observing systems. (Hewison et al., 2013; Wu et al., 2009) Through careful analysis of spectral signals, relative
32 stability can be assessed and even small problems with individual sensors can be identified. Ground-based and in
33 situ observations continue to offer some of the most useful information for constraining offsets and drifts in satellite
34 instrument as well as providing reliable, direct information on the Earth System. Both campaigns and long-term
35 monitoring efforts continue to help verify the accuracy and stability of satellite observations. Despite current efforts,
36 long-term stability and absolute calibration still present a challenge to current internal consistency methods, leading
37 many to look to other approaches for absolute calibration, traceability, and the ability to verify stability.

38 Perhaps the most innovative and needed advancements will come though future in-flight calibration approaches. The
39 development and use of these high accuracy climate benchmark instruments has been advocated for since the early
40 2000's and described in the NISTIR 7047 (2004) and ASIC3 (2007) workshop reports. These high accuracy
41 instrument systems will provide two fundamental products of great value to the climate science community: 1)
42 reliable long-term records of basic climate forcings, response, and feedback for analysis and climate model
43 verification, and 2) in-flight calibration standards for environmental operational satellite sensors including weather
44 satellites that do not have a rigorous pre-flight radiometric calibration requirement or the ability to perform
45 degradation corrections on-orbit. These ideas have been formulated in the visions of CLARREO and TRUTHS, but
46 may be tested in other reconfigurations (Wielicki et al., 2013; Best et al., 2008; Fox et al., 2013; Tobin et al., 2016).
47 Until such techniques are developed and tested, ground based and in situ observations continue to offer some of our
48 best ways of tying satellite observations to traceable standards. Efforts are already underway to estimate
49 uncertainties due to the matching of independent satellites with these reference sources (Feldman et al., 2011;
50 Lukashin et al., 2013).

51 Even with future improvements in satellite observation accuracy, the challenge will remain to understand and merge
52 records from different satellites – each potentially using its own calibration and collection approaches – to provide a

1 single observational record. One of the key factors that we can control is the length of overlap between existing and
2 future satellites. Analysis of an overlap record can only give us an estimate of relative drift, but in the absence of
3 traceable in-flight calibration, it is often one of our best checks on long-term stability of the final data products.
4 Understanding that decisions on overlap will directly affect both the cost of monitoring and the value of the final
5 dataset for evaluating long-term changes in climate, we propose approaches to objectively evaluate the length of
6 overlap needed to achieve a specific stability in the merged data record. In Section 8 we offer an approach to
7 evaluate how important overlap is compared to other choices that can help improve a long-term data record.

8 **2. Approach**

9 The statistical analysis techniques developed by Weatherhead and collaborators (see for example, Weatherhead et
10 al., 1998; 2000) provide a basis for addressing the length of time needed for adequate overlap based on the
11 magnitude of the signal variance as well as residual noise autocorrelation. In this paper, we perform a case study by
12 applying these techniques to existing SORCE SIM and SOLSTICE instrument data thereby illustrating the use of
13 statistical methodology to estimate the length of overlap needed to achieve records of specified stability. The
14 techniques discussed herein may be useful for instrument scientists pursuing improvements in on-board instrument
15 corrections, but also for mission planning by program managers to ensure the best overlap characteristics of
16 adjoining missions; the basic concepts of uncertainty from merging of datasets are directly useful to those interested
17 in using the data.

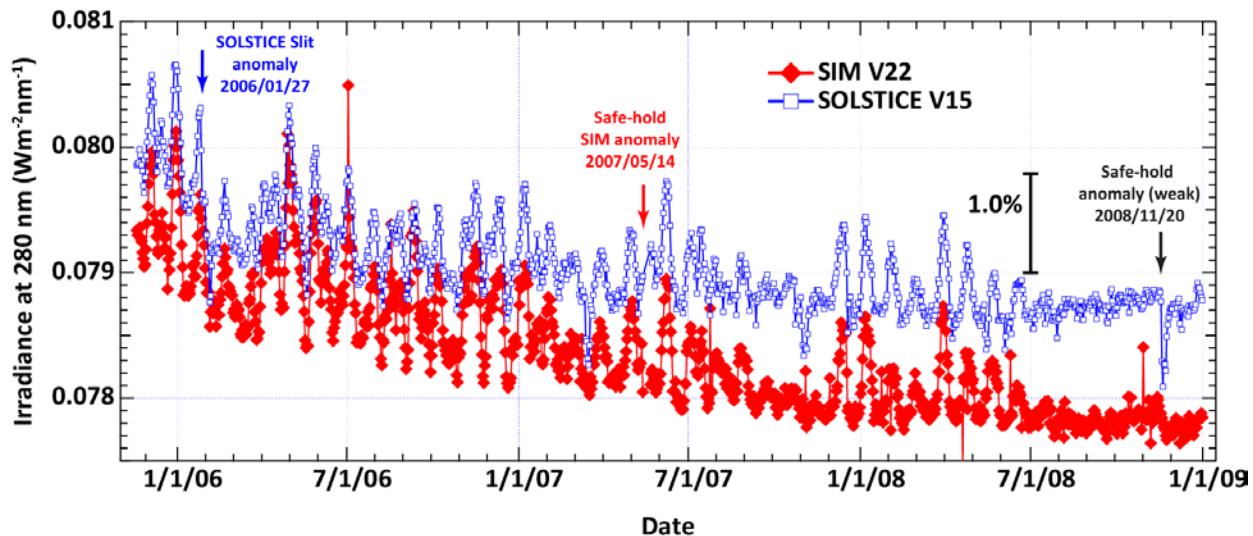
18 In order to appropriately analyze satellite observations, it is necessary to understand and appropriately incorporate
19 the available information on the pre-flight calibration of instruments and in-flight expected behavior. The detailed
20 in-flight circumstances that produce instrument instabilities are highly specific to individual sensors so the best
21 practice is to employ instrument telemetry and on-orbit calibration methods traceable to international standards.
22 Such approaches can be used to develop detailed measurement equations that can account for the occurrence of
23 degradation and correct the measured signal to produce high quality Level 1 data. The measurement equation carries
24 its own uncertainties and, in principal, allows for the estimation of time-dependent uncertainties as a function of
25 mission day. This measurement equation approach is advocated in the “Guide to the expression of uncertainty in
26 measurements” (JCGM, 2008) by the Joint Committee for the Guides in Metrology (JCGM) and relies only on the
27 known and measurable properties of the subsystems that compose the full instrument used for the observation.
28 Instrument teams apply these corrections to produce the final Level 2 data in an effort to provide the most accurate
29 measurements independent of outside data sources. For this study, we assume all relevant corrections have been
30 made to the data to account for known biases and drifts in the instrument.

31 **2.1 Introduction of SORCE SIM and SOLSTICE instruments:**

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

1 **2.2 Set up for SIM/SOLSTICE comparison:**

2 Solar irradiance is a crucial driver in the Earth's atmospheric system, influencing variability, circulation and long-
 3 term behavior of the atmosphere and having a direct role in atmospheric chemistry for the upper layers of the
 4 atmosphere. A motivation for this solar irradiance study, as is true for most other long-term satellite monitoring
 5 efforts, arises from the need to understand the length of time needed for the overlap of the currently operating
 6 SORCE mission with the next generation Total and Spectral Irradiance Sensor (TSIS). TSIS is currently scheduled
 7 for launch in the fourth quarter of 2017 for deployment on the International Space Station. While Earth observations
 8 often require a minimum of a one year overlap to cover the full range of expected observations, such arbitrary
 9 criteria ignore longer timescale phenomena including ENSO and NAO, and are impractical for covering a full 11-
 10 year solar cycle in a planned overlap period. Here we are applying analytical techniques to understand the length of
 11 time needed to quantify the offset between two satellite observing systems and to understand the drift between two
 12 satellite records (Weatherhead et al., 1998). While it is unclear whether the TSIS/SORCE overlap will mimic the
 13 findings from the comparison of the two SORCE instruments, this effort will examine how potential instrument
 14 anomalies and systematic errors in the degradation corrections affect the ability to determine the length of time
 15 needed to determine a drift difference in the two sensors. For this study, we are using a subset of three years of data
 16 from 18 November 2005 to 31 December 2008 (1140 days), characterizing the time period corresponding to the
 17 descent into the solar minimum condition of Solar Cycle 23 with the minimum value apparently in the January-
 18 February period of 2009. This time period was selected to approximate what would be expected from an overlap
 19 comparison campaign conducted during the descending phase of Cycle 24 projected to be in the 2019 time frame.
 20 However, the Solar Cycle 23 minimum is the longest and quietest time period of the space age (Schrijver et al.,
 21 2011; Araujo-Pradere et al., 2011; Araujo-Pradere et al., 2012), but our analysis does not rely on this situation
 22 persisting into the Solar Cycle minimum. This paper targets common observations of the irradiance in the 280 nm
 23 spectral region which includes the highly variable core of the Magnesium II lines. This region was selected because
 24 the variability of the Mg II lines is an important indicator of solar chromospheric variability and is frequently used
 25 for space weather applications (Viereck et al., 2004; Marchenko and Deland, 2014) and as a proxy for solar
 26 influence on stratospheric ozone and temperature (Hood and Zhou, 1998). It should be noted that the SIM (version
 27 22) and SOLSTICE (version 15) used in this study are used as-reported on the publically available SORCE web
 28 page (<http://lasp.colorado.edu/home/sorce/data/ssi-data/>). SIM and SOLSTICE corrections are made independently
 29 of one another, but the higher resolution 0.1 nm resolution SOLSTICE data are integrated into a fixed 1-nm bin
 30 centered at 280 nm. The SIM instrument has a FWHM resolution of 1.1 nm with 6 samples per resolution element.
 31 While some offset in irradiance is expected due to spectral sampling used to generate the data products, the
 32 difference is fixed and does not drift as a function time due to the well-defined wavelength scale and spectroscopic
 33 properties of the two instruments (Harder et al., 2010).



34
 35
 36
 37

Figure 3: An 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

1 active regions dispersed on the Sun and are not due to instrument noise; note that this modulation is apparent in both
2 datasets. This data segment contains all three uncertainty sources identified in this study: offsets, drifts and jumps.

3 Figure 3 shows the time series comparison of SOLSTICE version 15 and SIM version 22 used in this study. These
4 overlapping datasets illustrate three types of inconsistencies that occur in geophysical records, and have been noted
5 in both the ozone and temperature satellite datasets described above. These three sources of uncertainties combine
6 and contribute to the length of overlap needed to derive a robust climate record from satellite records.
7

- 8 1. There is an offset of about 0.5% in the pre-flight calibration between the SOLSTICE and SIM. The pre-
9 flight absolute calibration is on the order of 1-2% so within the ability to absolutely calibrate the
10 spectrometer. Note that the observed differences in Figure 3 are within the expected pre-flight calibration
11 uncertainty, but these differences are still large relative to some scientific uses for solar data. The value of
12 overlapping missions for an appropriate period of time is the ability to verify pre-flight calibration
13 estimates of uncertainty and potentially improve the long-term datasets for scientific applications.
- 14 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
16 of the degradation spot 'burned-in' to the collimator mirror – see McClintock et al. (2005a) for more detail
17 on the optical configuration. As the Earth-satellite system moves around the elliptical orbit of the Sun a
18 different illumination occurs on the first optic thereby modulating the intensity of light that propagates
19 through the rest of the optical system. This same effect occurs in the SIM data but has been a part of the
20 standard degradation correction for the last versions. SOLSTICE version 15 tends to flatten the apparent
21 long-term magnitude of the 280 nm variability relative to earlier SOLSTICE versions.
- 22 3. There are jumps in the time series related to spacecraft and instrument anomalies. Significant events are
23 identified in instrument and spacecraft housekeeping telemetry and changes in behavior before and after
24 these events can be characterized and corrected in the time series. Examples of these phenomena are seen if
25 Figure 3 where SOLSTICE experienced a failure of the mechanism that changes the entrance slit from the
26 solar to the stellar mode on 27 January 2006. The slit was moved back into position for continuous solar
27 observations but did not return to the exact same position so the optical path through the instrument
28 changed and therefore disrupted the degradation corrections and the wavelength scale. Similarly, a
29 spacecraft safe-hold event on 14 May 2007 caused the instruments to become very cold and significantly
30 changed the SIM wavelength scale and perhaps the transmission properties of the instrument. The change
31 in the SIM wavelength grid is apparent in the uncorrected data, but in Figure 3 the data are interpolated
32 onto a standard mission-length wavelength scale and does not appear as a jump in this figure. The 2007
33 safe-hold event had little effect on the performance of the SOLSTICE. The jump associated with the 2006
34 SOLSTICE slit anomaly has also been corrected and the change in character seen SOLSTICE data at this
35 time represents the best compromise over the full wavelength range of the instrument.
36

37 The next sections of this paper will address the effects of these three anomalies (offsets, drifts, and jumps) in the
38 SORCE datasets and discuss their impacts on dataset uncertainty. The primary contribution of this paper is to
39 quantitatively address the impact of the length of measurement overlap on helping verify a specific level of stability
40 in the final dataset.

41 3. Offsets

42 Efforts at merging satellite data in the past have focused on deriving offsets to limit relative differences before
43 combining data from different sensors into a continuous record (e.g. Wentz and Schabel, 1998; Santer et al., 2003;
44 Smith et al., 2008; Dudok de Wit et al., 2008; Chander et al., 2013b).

45 One of the most studied issues underscoring the importance of proper treatment of multiple satellite records involves
46 the corrections and merging of Microwave Sounding Unit temperature records. Christy et al. (1995, 1998, 2000)
47 accurately pointed out that trends from satellite temperature records were not in agreement with other temperature
48 records and showed a cooling of the troposphere rather than a warming. Additional work showed that a number of
49 corrections to the satellite record could make a direct and notable difference on the trend derived from the resulting
50 data (NRC, 2000a, 2000b; Zou and Qian, 2016). Some of the most salient lessons from this effort were summarized
51 by Thorne et al. (2005), who concluded, among other points that, "individual adjustments will a priori retain a non-
52 climatic signal of unknown sign and magnitude regardless of how reasonable and physically plausible the chosen

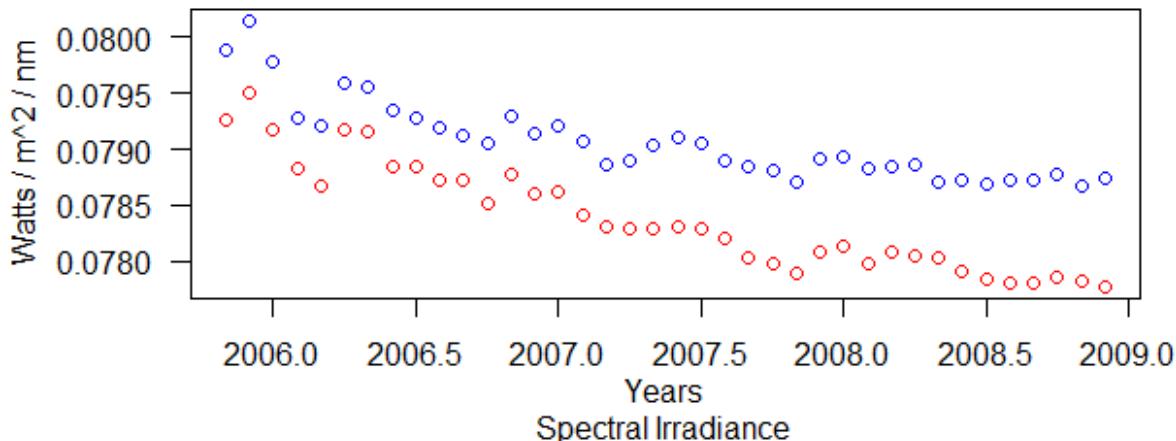
1 homogenization approach." The uncertainty of merging satellite data records is a continual challenge with a variety
2 of approaches employed including comparison to ground-based records, statistical intercomparison of satellites by
3 latitude, time of day and season, as well as use of physical models to look for appropriate consistencies with
4 available data. Details of the merging process directly influence the resultant trends and add to the level of
5 uncertainty in the final datasets (Karl et al., 1986). In this section we consider the case where overlap is non-existent
6 and for the case where overlap exists we consider the length of time required to achieve a specific uncertainty in an
7 overlapping set of data. These cases illustrate the need for overlap periods of sufficient duration to make a
8 quantifiable improvement in the long-term record.

9 We consider the straight-forward method for merging two sequential (non-overlapping) data records by requiring
10 the mean level of the three years of data prior to the discontinuity be equal to the mean level of the three years of
11 data after the discontinuity. In such a situation, those six years of data are being forced, by the algorithm, to have
12 very little trend. Imagine a situation where there are two such discontinuities in a twenty year record; more than half
13 of the data has been coerced to have virtually no trend, making the resultant data unreliable for many long-term
14 monitoring uses. The case of no overlap can occur due to a variety of reasons, including the sudden loss of a
15 satellite, or problems on launch of newer satellites. The end result of any offset correction will have a direct impact
16 on the magnitude of the resulting trends.

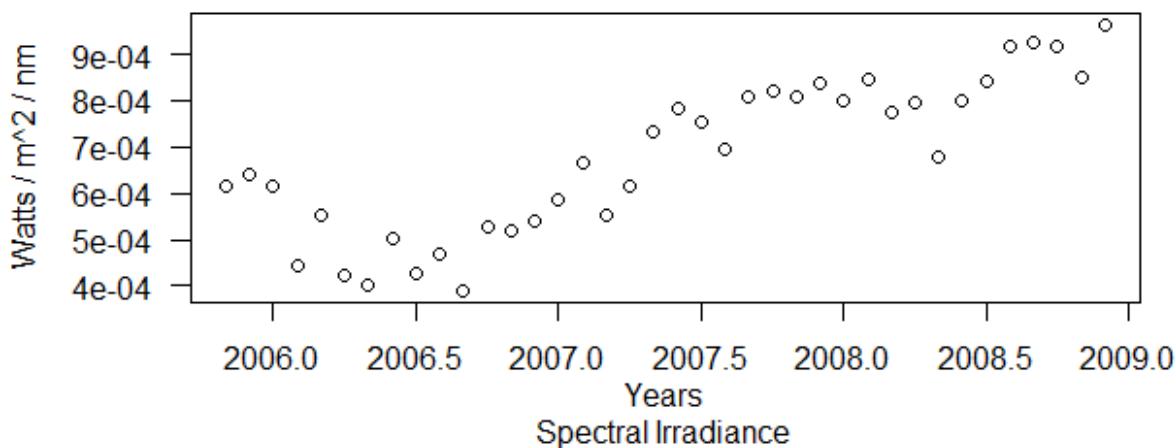
17 To estimate the time needed for overlap requires an estimate of what the overlap time series would look like. We use
18 monthly averages, a common standard in many climate-related research efforts for several reasons: monthly
19 averages avoid the match-up issues and potential non-linearity of short-term features, such as storms (Araujo-
20 Pradere, 2004), and offer enough resolution to observe long-term behavior of the match-up.¹ If, as is the case of the
21 SOLSTICE-SIM overlap, the difference between the two datasets look like Fig. 4 with significant agreement in
22 observed variability, 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.

**SOLSTICE (blue) and SIM (red)
Spectral Irradiance**



**SOLSTICE - SIM
Spectral Irradiance**



1
 2 Figure 4: Monthly averaged SOLSTICE data and SIM data (top plot) and SOLSTICE-SIM (280 nm Mg II) data (bottom
 3 plot) ($\text{watts m}^{-2} \text{ nm}^{-1}$) as a time series for the period October 2004 through December 2008. The data show that the
 4 observed differences are small, but do not appear to be stable, thus a simple level shift to bring the two datasets into
 5 agreement may not fully address the match-up and stability issues of the dataset.

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

13

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

2 Where σ is the observed magnitude of variability of the observed differences in monthly averages; ϕ is the observed
 3 autocorrelation in those differences, and n is the number of months of observed overlap. This estimate of Standard
 4 Error of the mean is dependent on the data behaving as an autoregressive with time-lag of one month, AR(1), with
 5 the underlying interventions behaving approximately as a Gaussian distribution. This more appropriate formula
 6 gives a standard error on the mean of 5.2×10^{-5} watts $m^{-2} nm^{-1}$, notably larger than if autocorrelation is ignored.
 7 Monthly averages have a broad range of uses in environmental sciences for trend detection, development of
 8 climatologies and monitoring the behavior of the Earth. Monthly averages can remove higher frequency noise and
 9 some sampling match-up problems, but they can also obscure important details and can often introduce their own
 10 biases, especially when sampling is irregular in time or space (Toohey et al., 2013; Toohey and von Clarmann,
 11 2013). We show the behavior of the underlying interventions as Gaussian and our tests for AR(1) in Appendix B.

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

$$15 \quad \text{Months to Estimate an Offset} \cong 1.96^2 \frac{\sigma^2}{\text{Offset Limit}^2} \frac{1+\varphi}{1-\varphi} \quad (2)$$

The above formula shows that for a given magnitude of variability and autocorrelation in monthly 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. The factor of 1.96 is to support a 95% confident limit on the offset with a 50% likelihood of detection; if more confidence is needed in the offset, a higher factor can be used based on classic statistical tables. 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. 4, 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. Note also that 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 the formula offers an estimate of length of time needed to limit uncertainty in an offset, and such an estimate is rarely precise to many significant digits. We conclude for the datasets we have been exposed to that after roughly two years of data collection the large number limit of 1.96, may be considered appropriate.

37 The impact of the offset on the use of the data is critically important to the final analysis. While a “best” merged
38 dataset may be produced from multiple satellites, users should never ignore the added uncertainty due to merged
39 data sources. Using the merged data without including the impacts of the merging would result in smaller standard
40 errors in computing means, variability in trends, than is actually appropriate. The magnitude of the impact of the
41 offset correction is dependent on the use of the data. Two cases are considered here for illustrative purposes. If the
42 merged dataset will be used to estimate the impact of storms on a stable electrical grid, and the impacts have been
43 estimated from the effects observed using the first satellite record, an uncertainty of 0.2% means that the new solar
44 storms may well be off by $\pm 0.2\%$ and the uncertainty in impacts need to be appropriately calculated and conveyed.
45 If the merged datasets will be used to estimate long-term trends, then the impact of an uncertainty of $\pm 0.2\%$ means
46 that any trends derived will be affected by that level of uncertainty carried out through the length of dataset used for
47 analysis, and may affect the significance of the expected trend, if care is not taken to reduce the uncertainty in the
48 overlap adjustments.

1 **4. Drifts**

2 While offsets are routinely addressed in the merging of satellite datasets, potential drifts in satellite data are also
 3 critically important to many of the final applications of climate data, most notably trend detection both of the direct
 4 parameter being observed and observations that are dependent on the observed parameter. There are several
 5 fundamental factors that can contribute to a drift in satellite observations including decay of instrumentation and
 6 changes in satellite orbit. Efforts are ongoing to minimize the impact of these factors, but all corrections involve
 7 assumptions and each satellite may invoke different approaches to monitor and address stability. The merging of
 8 satellite records, at a minimum, needs to test for potential drift between the overlapping satellite records. The
 9 amount of drift that can be detected through satellite overlap depends on the length of overlap period and on the
 10 quality of the match-up in overlapping data.

11 The impacts of undetected drift will have direct impact on the scientific results derived from the data. Bourassa et al.
 12 (2014) showed that the uncertainty in drift from a continuous record from multiple satellites is critical to long-term
 13 monitoring of the Earth. Rahpoe et al. (2015) find intercomparisons of six different ozone limb measurements to
 14 drift relative to each other at a statistically significant rate, sometimes as high as 5% per decade or more. However,
 15 most drifts were statistically insignificant due to the limited length of data records--generally less than 10 years. In
 16 the case of solar viewing instruments, BenMoussa et al. (2013) discuss in detail causes and effects of degradation in
 17 a variety of different instruments that span nearly two decades and cover a broad wavelength range. They conclude
 18 that there is no single best method to correct and monitor degradation and the correction schemes for overlapping
 19 missions are likely to be very different depending on the instrument hardware selection. An important example of
 20 this is well documented in the efforts to correct drift in the CIRES instrument (Cloud and Earth's Radiant Energy
 21 System, see Loeb et al., 2016 and references therein). In this report, long-term stability was linked to loss of optical
 22 transmission due to UV exposure and molecular contamination, very similar to the mechanisms discussed in
 23 BenMoussa et al. (2013). Fruit et al. (2002) have addressed the effects of energetic particle on glass transmission,
 24 but inhibiting and characterizing carbonization of optical surfaces remains a steadfast and unsolved problem. In each
 25 case, evaluation of how best to characterize the drift takes place. For the Solstice-SIM data overlap, we noted that
 26 the differences between the two sets of data showed lower variability than the ratio of the data, indicating an offset
 27 would be better modeled as an additive adjustment. In many cases, uncertainty on satellite records' drift can be even
 28 more significant to scientific uses of the data than the offsets from one instrument to the next. In this paper, we look
 29 to see, to what extent, some confinement of the problem may be achieved through appropriate overlap of
 30 independent instruments.

31 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
 32 in the overlapping data, using approaches from Weatherhead et al. (1998), we can estimate the length of time
 33 necessary to observe a drift of that magnitude in an overlapping dataset. Weatherhead et al. (1998, 2000) have
 34 shown that one can estimate the length of time to detect trends in environmental observations. This approach is
 35 applied to estimate the time to measure a differential drift with a specified uncertainty in the observations taken by
 36 two different systems. When detection is considered at the 95% confidence level, estimated overlap for detection is:

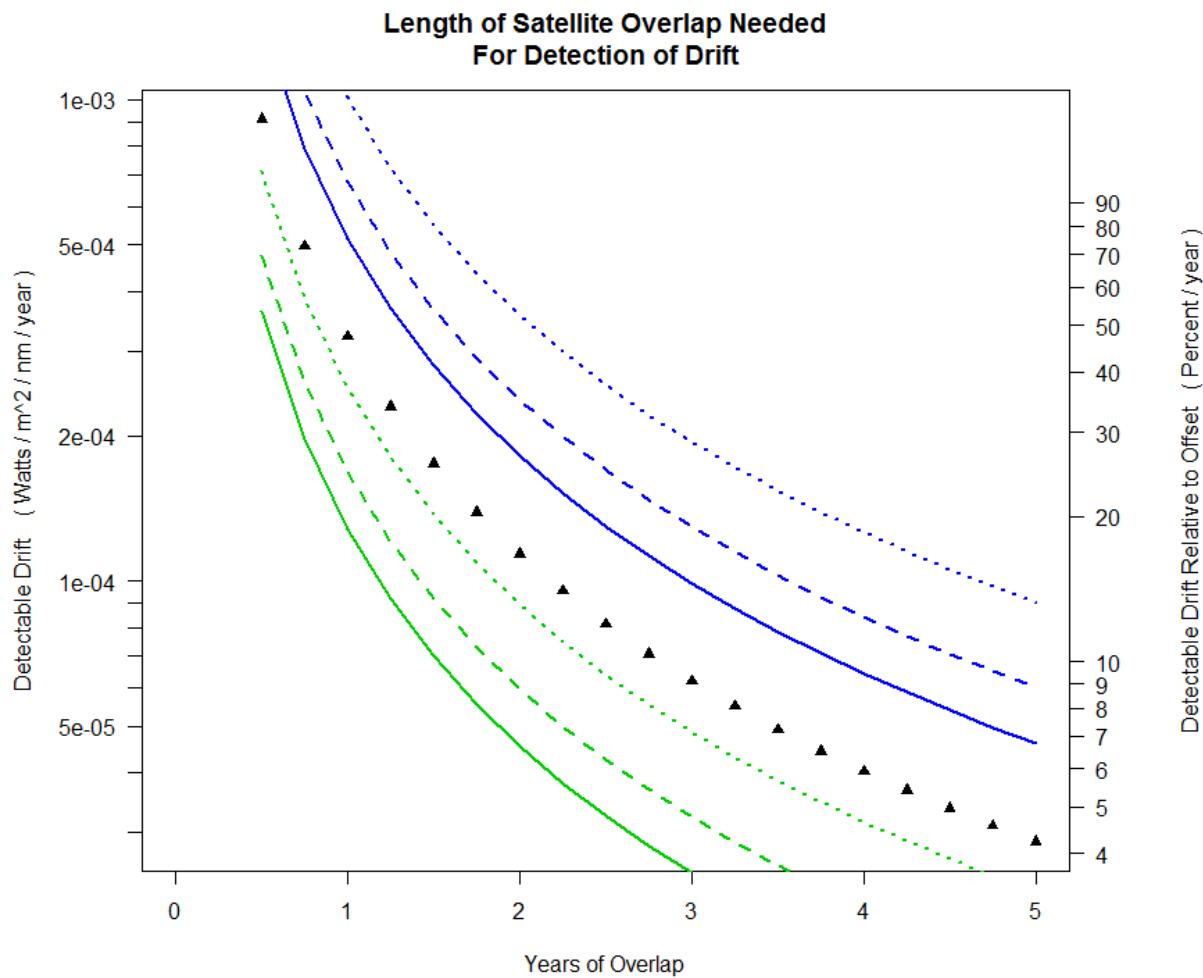
$$37 \text{ Months to Estimate a Drift} \cong 12 * \left[1.96 \frac{\sigma}{|drift|} * \sqrt{\frac{1+\varphi}{1-\varphi}} \right]^{2/3} \quad (3)$$

38 Where $|drift|$ is the absolute value of the magnitude of the differential drift, σ and φ are the magnitude of variability
 39 and autocorrelation, respectively, of the differenced monthly data once any existing trend is removed. We can
 40 identify the drift we would like to have the capability of detecting; and estimate both σ and φ from existing data –
 41 either from observations or modeled experiments. As an example, a drift of 0.1% per year of the observed SOLSTIC
 42 data, ($0.00008 \text{ watts m}^{-2} \text{ nm}^{-1} \text{ yr}^{-1}$), with the observed variability in the overlap data (sigma of $8.59*10^{-5} \text{ watts m}^{-2}$
 43 nm^{-1} and phi of 0.57) results in 2.5 years likely needed to achieve that uncertainty.

44 It may be noted that the natural world has variability (σ) and autocorrelation (φ) that are inherent, and may change
 45 slightly over time. A distinct advantage of satellite observations can be the frequency of the observations.
 46 MacDonald (2005) has shown that the monitoring approach can have a direct impact on these values, as well:
 47 monitoring less frequently – perhaps only once or twice a month – results in higher variability and slightly lower
 48 autocorrelation in our dataset. As encapsulated in Equation1, an increase in the number of measurements per month
 49 improves the detectability (shortens the number of years to detect a given trend), but only up to the limit of the
 50 system's natural variability. For each situation of overlapping of satellite missions, the results will depend on the

1 method of observation and the parameter being observed; for Earth observations, the results can also depend on
 2 location and even time of year.

3



4
 5
 6 Figure 5: The ability to constrain a detectable drift is a direct function of the number of years of overlap and the quality
 7 of the overlap data. Black triangles offer the detectable drift in the difference between the SOLSTICE and SIM satellite
 8 instruments as a function of the overlap period; calculations made use of the observed characteristics of the Solstice-SIM
 9 monthly differenced data: magnitude of variability of 8.58×10^{-5} watts $m^{-2} nm^{-1}$ and autocorrelation of 0.57. Green lines
 10 assume a smaller magnitude of variability in the overlap data (half of what is observed in SOLSTICE-SIM) and
 11 autocorrelation of 0.4 (solid), 0.6 (dashed) and 0.8 (dotted). Blue lines assume a larger magnitude of variability in the
 12 overlap data (twice what is observed in SOLSTICE-SIM) and autocorrelation 0.4 (solid), 0.6 (dashed) and 0.8 (dotted).

13 Using the SOLSTICE-SIM data as an example, tremendous accuracy gains are achieved for each year of monitoring
 14 for the first few years of overlap, with diminished returns after that. Respecting the cost of overlap, and making
 15 appropriate calculations with emerging overlap data, an appropriate overlap plan can be estimated to allow for
 16 scientific standards to be met. Figure 5 shows that for a given magnitude of variability and autocorrelation observed
 17 in overlap differences, we calculate the number of years of overlap needed to detect a specific level of drift. In this
 18 case, using the SOLSTICE-SIM data as an example, the magnitude of detectable linear drift drops from 1.1×10^{-4} to
 19 0.6×10^{-4} watts $m^{-2} nm^{-1} year^{-1}$ by allowing the overlap to be three years, instead of two years. The level of agreement
 20 of data from the two instruments results in the magnitude of variability and autocorrelation observed in the
 21 differences. For observations that are very much in agreement with each other, we can expect low variability, and
 22 thus, a relative drift can be detected earlier, as represented by the green lines in Figure 5. For a poor match between

1 the overlapping observations, we can expect longer times of overlap will be needed as represented by the blue lines
2 in Figure 5. The relative drift on the right side of Figure 5 offers information on the percentage basis of the overlap
3 data, not in the raw SOLSTIC or SIM datasets, because it is the fundamental behavior of the differences that
4 determines the information content of the overlap period. It is important to note that drifts in overlap data, as drifts
5 in nature, can be approximated as linear, logarithmic, or a variety of other representations, as the data and the
6 physics of the situation suggest; for Fig. 5, we assume an approximately linear drift over the time period of the
7 overlap. For satellite observations, a number of changes are expected over the lifecycle of the instruments; all known
8 and expected changes are approximated and adjusted based on current best understanding. However, particularly
9 with new technologies, these assumptions must be checked by careful evaluation of the data, thus emphasizing the
10 importance of an adequate overlap period to help confine potential drifts to a specified level. While pre-launch
11 calibration may indicate drift will be less than a specific level, the ability to verify this will depend on independent
12 intercomparisons of observations.

13 Although no error bars are offered in Fig. 5, it is important to remark that when estimating how long it will take to
14 detect a specified drift, two statistical levels must be considered: one that identifies the meaning of “detecting a
15 drift” and a second that identifies the likelihood of detection of that drift in the specified period of time, if that level
16 of drift is the true, long-term drift in the overlap. Of course, it is possible to detect smaller drifts than the value
17 obtained from any particular point (drift-overlap pair) of this figure, as it is also possible to determine a given drift a
18 few months earlier or later than the value obtained from the point in the figure. Comments on the appropriate
19 interpretation of the likelihood of drift detection are discussed in detail in Appendix C, where two non-standard,
20 dimensional error bars are introduced in the figure to help the reader to understand this uncertainty.

21 5. Jumps

22 Jumps are permanent or semi-permanent level shifts in the data that occur at specific points in time and are not
23 attributable to the parameter being observed; jumps could represent a change in sensitivity of an instrument or a
24 change in location or orientation of the satellite. To ignore such events results in greater uncertainty in the
25 appropriate offsets and in artificial drifts in the final data set. If the magnitude and time of the shift are known, then
26 the data can be adjusted before being analyzed for offsets and drifts. If the magnitude and timing of jumps are
27 unknown, as is more often the case, and must be derived from the data, the presence of such level shifts increase the
28 uncertainty of both the offset and drift in the overlapping data and hence lengthens the time necessary to achieve a
29 high quality final dataset. Continuous satellite overlap can make the ability to identify and understand sudden jumps
30 significantly easier, but this is often beyond current monitoring approaches.

31 It is tempting to believe that jumps can be easily identified and corrected. Testing this belief, Free et al. (2002)
32 carried out a comparison of seven different groups examining radiosonde records with the intent of identifying,
33 quantifying and correcting observed discontinuities in temperature sonde time series. The different groups identified
34 a widely different number of discontinuities using a range of techniques. They further differed significantly on the
35 magnitude of corrections and even, at times, the sign of the needed correction, resulting in changes to observed
36 trends by between 35% and 80%. Thus, even identified and corrected discontinuities introduce some uncertainty in
37 the long-term stability of the record. Weatherhead et al. (1998) showed that these corrections can be quantified with
38 advanced statistical techniques. The resulting uncertainty in long-term stability increases the number of years
39 necessary to detect trends, independent of how large the correction is. Hurrell and Trenberth (1997) point to the
40 importance of two small, discrete, downward jumps in merged satellite records that dominated the trend results for
41 tropospheric temperature records.

42 Jumps can occur for a variety of reasons related to instrument changes (e.g. Brown, 2013). Depending on the
43 physical source of the jumps, the effects can last from less than a few hours to multiple years. In some cases, true
44 jumps occur in the parameter being observed. In many other situations, the observing system or assumptions used in
45 the algorithms are responsible for the jumps and there is a desire to identify and remove these spurious jumps. The
46 statistical removal of a spurious jump involves two steps: identifying a jump in the overlap differenced data and
47 estimating the magnitude of the needed correction. A variety of approaches are used to both identify jumps (Jaxk et
48 al., 2007; Vincent et al., 1998, 2002; Ducre' -Robitaille et al., 2003) and correct for these jumps in satellite and non-
49 satellite observing systems (Karl and Williams, 1987; Mitchell and Jones, 2005). When information is available to
50 identify the timing of a jump, there is considerably more confidence in the correction for the jump because physical
51 interpretation is easier and therefore corrections can be physically based rather than statistically based. If there is no

1 external information on timing, then one has to consider that there could be other jumps below the threshold for
 2 detection and estimates for how the instrument is behaving are more uncertain. For jumps in overlapping datasets,
 3 the correction brings into question the magnitude of any derived offset as well as the magnitude of any drift in the
 4 overlap period. For jumps in observing systems when there is no overlap, the challenge of appropriately identifying
 5 and correcting jumps is notably more difficult, again pointing to the value of redundant observing systems when
 6 possible. We focus on the ability to detect and understand the jumps that last more than a few months, as they may
 7 be the interruptions that can cause the most serious damage to long-term records, particularly those used in the
 8 context of climate research. We consider the two cases separately of how a jump affects the offset estimate and how
 9 a jump affects the drift estimate.

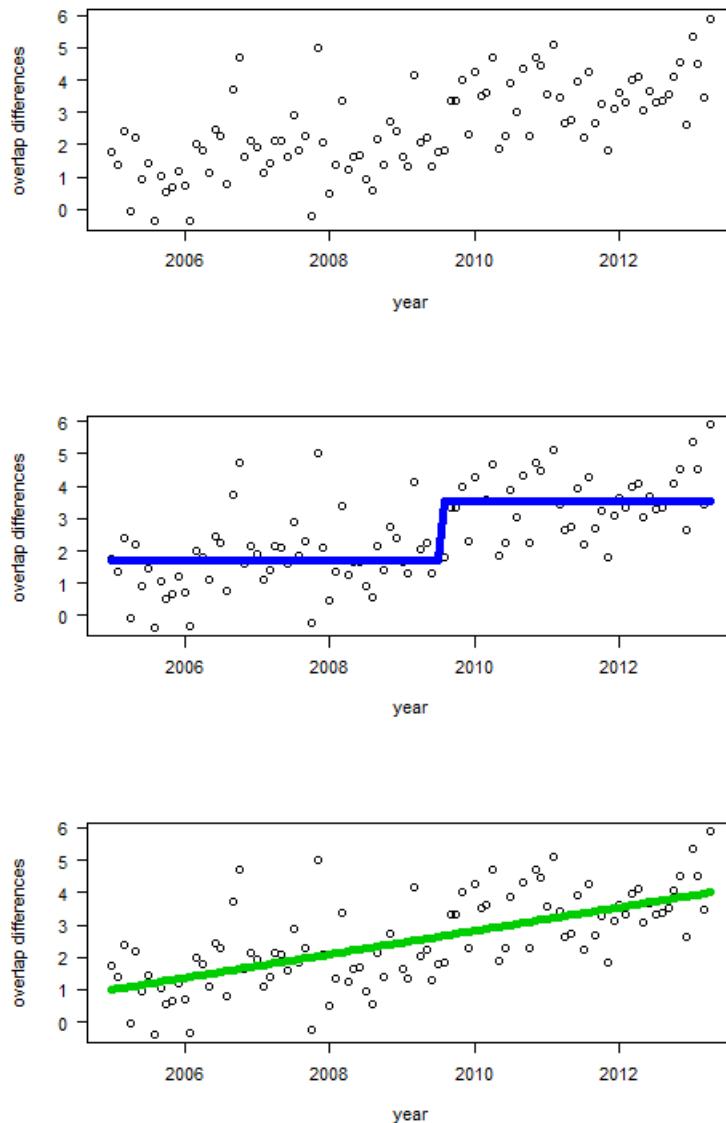
10 While ancillary data about the instrument or the observed parameter may be used to identify the existence and
 11 timing of a jump, deriving the magnitude of correction by examining the data is dependent on the amount of
 12 variability (both magnitude and autocorrelation) in the overlap difference data. For the case of an otherwise stable
 13 offset between the two sets of satellite data, the added uncertainty on the true offset is enhanced based on the
 14 uncertainty of the jump. The impact of the jump on offset estimates is a minimum when the offset occurs in the
 15 middle of the overlap period, because maximum information is available to identify the size of the jump. For the
 16 case of a drifting offset between the two sets of satellite data, the added uncertainty on the true drift is also enhanced
 17 based on the timing of the jump. Because the impact of the jump is co-linear with the derived drift, estimating the
 18 overall drift in instrument offset is more difficult in the presence of jumps. If we assume that we are going to fit the
 19 environmental data to a linear statistical model of the form:

20 $Environmental\ Data = Mean + Linear\ Drift + Offset_{\tau} + Noise(0,1)$ (4)

21 where the Mean, Linear Drift and Offset at time τ are derived from the data. The residuals ($Noise(0,1)$) are modeled
 22 as an AR(1) process with mean=0, and both σ and φ derived from the observations. With such a statistical model,
 23 Weatherhead et al. (1998) show that the impact of the additional jump term on deriving an accurate estimate of drift
 24 is dependent on the timing of the jump.

25 $Months\ to\ Estimate\ a\ Drift \cong 12 * \left[1.96 \frac{\sigma}{|drift|} * \sqrt{\frac{1+\varphi}{1-\varphi}} \right]^{2/3} \frac{1}{[1-3\tau(1-\tau)]^{1/3}}$ (5)

26 Where τ is the fraction of the data before the identified jump occurs. Thus, a longer time of overlap is required for
 27 accurately confining a drift in overlap data: As opposed to the impact of jumps on estimating offsets, the impact of
 28 jumps on derived drifts is largest when the jump occurs in the middle of the overlap period ($\tau=0.5$), where the
 29 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
 30 the similar temporal signature of both drifts and jumps on a long-term record as illustrated in Fig. 6. Equation 4
 31 assumes that any drift and offset will be fitted to the data simultaneously. If data will be fitted sequentially to an
 32 offset and then to a drift, the derived drift will be considerably smaller than if the data were fitted to a drift and then
 33 to an offset because the two functions (drift and offset) are not orthogonal and thus the derived results for magnitude
 34 of offset and magnitude of drift are not commutative. To be explicitly clear, correction of offsets in advance of
 35 deriving drifts can artificially minimize the amount of observed drift, while ignoring offsets (perhaps because they
 36 are not easily detectable) can either add to or diminish the derived drift.



1
 2 **Figure 6:** This dataset was created to visually show the potential impact of a spurious jump on the estimate of offsets and
 3 drifts. In all three plots, the same data are shown, with the second plot showing how the data could be modeled as an
 4 offset. The data were actually created by adding a linear drift to simulated autoregressive data, as shown in the third plot.
 5 The confounding nature of jumps and drifts cannot easily be separated; although ancillary data can be extremely helpful.
 6 Note these are synthetic data with arbitrary units.

7
 8 In the case of the SORCE instruments, these jumps are mostly prompted by changes in the performance of hardware
 9 subsystems of the spectrometers. For example, a significant jump occurred in the SIM on 14 May 2007 related to the
 10 instrument becoming very cold during a safe-hold event, and upon recovery, the instrument did not return to the
 11 same state as before the safe-hold. This event produced a significant change in the hardware that controls the
 12 wavelength drive (see Harder et al., 2005a for a description of the wavelength drive mechanism). Even after the
 13 wavelength correction a residual jump in irradiance level was still observed; the most likely cause of this jump is
 14 due to a change in light transmission caused by a change in optical path through the instrument that is different than
 15 before the safe-hold event.

1 For the particular case of Fig. 3, the jump appears to be adequately corrected, but other wavelengths show a
2 discernable discontinuity. A similar observation can be made about the SOLSTICE slit anomaly in 2006 (see Fig. 3).
3 Most disconcerting about these events is the possibility that the jumps can be disguised as a change in stability and
4 produce results similar to what appears in the discussion of Fig. 6.

5

6 **6. Impacts of Uncertainty in Drifts**

7 The primary purpose of overlap in instruments is to understand how two different instruments are responding to the
8 parameter each is intended to measure. Uncertainty in the long-term stability of data products directly affects the
9 level of confidence one can have in the long-term, merged datasets. Logan et al. (2012) noted that differences as
10 small as a few ppb in ozone records with 2–3 years overlap lead to different trends for 1995–2008. Frith et al. (2014)
11 use Monte Carlo techniques to estimate the uncertainty of the SBUV MOD merged data set based on both overlap
12 information and validation against independent data sets.

13 The impact of undetected drifts in overlapping data sets on the merged record can be illustrated with a simple Monte
14 Carlo representation, as shown in Fig. 7. Consider two satellites launched with seven year lifetimes. The second
15 satellite (blue trace) overlaps the first satellite (green trace) by either one or two years (top plot and bottom plot,
16 respectively). Each satellite launches with an uncertainty estimate on the pre-flight absolute calibration of its
17 measurements and some level of unidentified drift in the instrumental record, acknowledging that all identified
18 drifting factors have already been corrected to the best abilities of the instrument team. For this illustration, we
19 assume our simulated instruments have the same variability and autocorrelation during their overlap as the
20 SOLSTICE-SIM data used in the previous sections. Accordingly, an overlap of one year implies a drift in the
21 differences smaller than $3.28e^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$ (1 standard deviation; roughly 48 % yr^{-1}) would be
22 undetectable, and a relative drift of $1.16e^{-4}$ watts m^{-2} nm^{-1} $year^{-1}$ (1 standard deviation; roughly 17 % yr^{-1}) would be
23 undetectable with two years overlap (see Figure 5). For illustrative purposes we show 100 simulations generated
24 based on these potential relative drift values. In addition we depict a scenario in which the pre-launch absolute
25 calibration of the second instrument is better constrained, which can occur when advanced technologies are
26 introduced. Such a priori information about the individual data sets is important; in this case a user might choose to
27 adjust the first satellite record to the second to take advantage of the smaller absolute uncertainty in the second
28 instrument.

29

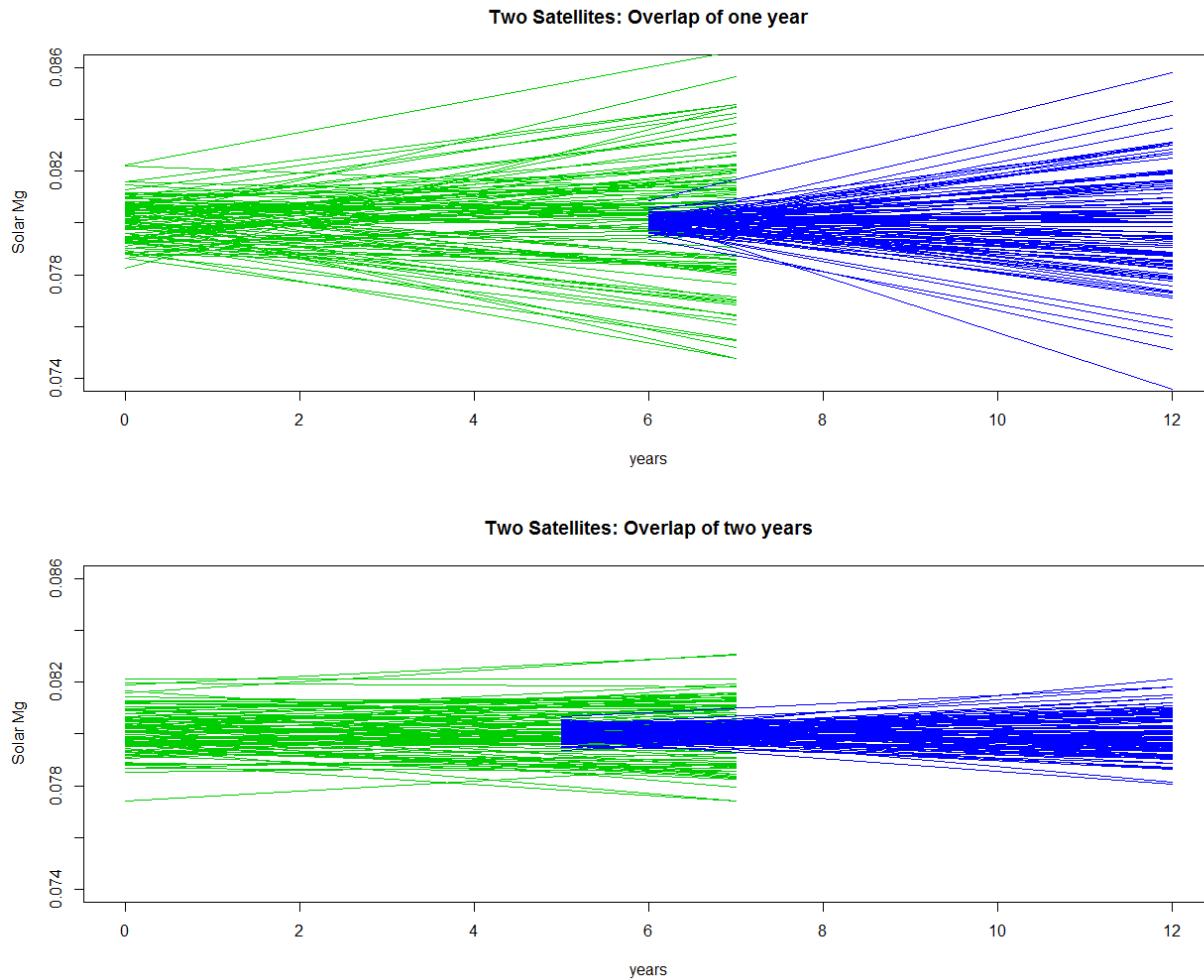


Figure 7: Two overlapping satellite observations (shown in green and blue) simulating 280 nm Mg II irradiance ($\text{watts m}^{-2} \text{ nm}^{-1}$) with an expectation for some absolute pre-flight calibration uncertainty and some unidentified on-orbit drift. The amount of unidentified drift can be 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.

We note that the drift uncertainty depicted here only represents the uncertainty associated with adjusting one data set to the other; it does not include the absolute drift. If, as an example, we assume our two simulated instruments had the same relative drift as observed in the SOLTICE-SIM example ($\sim 1.67 \text{e}^{-4} \text{ watts m}^{-2} \text{ nm}^{-1} \text{ year}^{-1}$, see Figure 4), this drift is not detectable with one year of overlap, but is detectable with two years overlap. However no amount of overlap can reduce the true absolute drift of either satellite dataset. As with any instrument intercomparison, satellite overlap by itself cannot replace calibration, but it does offer valuable information that significantly complements all other available information. However, this illustration can help guide how large an impact any confinement of potential drift may be and can offer insights into whether uncertainty on the level of drift is too large for the intended use of the data. Based on this type of analysis, some have argued for the value of redundancy in observing systems (Weber et al. 2016b) and the value of continued replacement of satellites to minimize the extent of unchecked drift in a single record (Stolarski and Frith, 2006, Frith et al., 2014 and Weber et al., 2016a) to potentially reduce the absolute uncertainty in merged records.

1 **7. Application to In Situ and Ground Based Observations**

2 Both in situ and ground-based Earth Observations often require targeted efforts and planning to assure a continuous
3 record based on discontinuous observing systems. In some cases, overlap is impossible when changes take place; for
4 instance when an expensive instrument moves location within a region or local changes, such as urban development,
5 directly affect the local observations. However, in some cases, the approaches presented in this paper can have direct
6 relevance. WMO's Global Climate Observing System's Reference Upper Air Network is currently evaluating the
7 transition from one sonde instrument package to another. Such a change in instrumentation is likely to affect the data
8 stream particularly when extreme values are concerned. The planned overlap period and detailed planning of the
9 intercomparison of results is currently underway within the GRUAN leadership team. In a similar situations, UK
10 Met Office evaluated the impact on a the switch from RS92 sondes to RS41 sondes, following a careful study
11 carried out jointly with Vaisala and the UK Met Office making use of simultaneous launches of 30 sets of two RS92
12 and two RS41 launches. The ability to compare the data simultaneously, using the same balloon system and doing
13 the intercomparisons with synchronized GPS times allowed for much smaller differences in the statistical limit to be
14 placed on the impact of transition from the RS92 to RS41 observations. This controlled study is the ideal situation
15 and contrasts with no overlap as noted by Weatherhead et al. (1997) which showed the large impact of sudden shifts
16 of radiation measurements, some of which were due to shadowing of the instruments by newly built structures. The
17 sudden shifts without overlap rendered the data inappropriate for trend analysis because the uncertainty due to the
18 level shifts were too large to allow detection of trends expected from ozone loss.

19 **8. Optimization and economic benefit**

20 Decisions to improve a single observing system can rarely be made without consideration of the potential impact on
21 the support of other approaches to improve monitoring. Figure 5 shows that drift detection accuracy improves as the
22 number of overlap years increases. Improvements in drift detection capability decrease as the number of overlap
23 years increases, but the optimal overlap duration is difficult to identify unless restrictions, such as cost, are
24 considered. If a specific stability criterion is the objective, the minimum required overlap can be directly determined
25 to meet that criterion. Assuming that the total costs increase as overlap time increases (this value can often be in
26 excess of \$1-2M per year), from an economic perspective the minimum required overlap time should also be chosen
27 as the most cost-effective. Although technical criteria are critical to understanding optimal data series overlap,
28 decisions in the policy arena are often based on economic analysis and thus it is important to better understand how
29 economic analysis frames the allocations of resources. "Optimal" in such discussions is structured on criteria
30 measured in societal benefits and costs.

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

41 Without specific stability criteria or budget constraints, economic criteria suggest choosing the overlap period or
42 combination of approaches to achieving data stability that provides the maximum net societal benefit (i.e., total
43 benefits minus total costs).² Identifying the societally optimal choice implies choosing the overlap (or possible mix
44 of calibration methods) where marginal cost equals marginal benefits. Marginal costs are the change in program
45 costs to achieve one unit of improvement (e.g. $\text{watts m}^{-2} \text{ nm}^{-1} \text{ year}^{-1}$ in detectable drift as per Fig. 5). From an

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

1 economic perspective, the marginal benefits of one unit of improvement in the data quality should be measured in
2 terms of potential changes in societal outcomes from the use of improved information (a value of information or
3 VOI approach (Laxminarayan R. and M.K. Macauley, eds. 2012)). While there has been some work on the value of
4 information from satellites (Donaldson and Storeygard, 2016; Cooke et al., 2014; Macauley, 2006), there have not
5 been many applications of economic analysis to determine optimal observational systems. Morss et al. (2005)
6 provide an overview of relevant economic concepts and theory for optimal design of observational systems based on
7 benefit cost tradeoffs.

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

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

24 While economists have extensive experience and applications in monetizing the value of potential changes in
25 societal outcomes (e.g. lives saved, reduced damages, improved crop yields, etc.), it is generally much more difficult
26 to and thus important to build more applications (1) identifying all of the potential stakeholders and potential
27 outcomes and (2) validly and reliably characterizing and quantifying the information value chain and how it
28 changes.

29 **9. Conclusion**

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

40 The uncertainty due to the merging of satellite records is unavoidable, but quantification of this uncertainty is
41 possible. In the case of identifying or verifying the offset in two satellite records, the uncertainty is inversely
42 proportional to the square root of the number of months of overlap. In the case of identifying or verifying the long-
43 term relative stability or potential drift, the number of months of overlap is inversely proportional to the drift to the
44 2/3 power. Both time estimates require some understanding of the variability in the overlap differences. If no
45 estimate of overlap variability is available, the behavior of the first few months of overlap can be evaluated to
46 estimate the length of overlap needed to achieve the prescribed tolerance. The impact of abrupt disruptions in the
47 overlap period on offsets can require up to 50% more overlap to be able to identify the offset and drift with the same
48 level of tolerance.

49 These algorithms are appropriate for a direct evaluation using only the satellite data. In some cases, particularly with
50 Earth observations, added benefits and challenges may exist. For instance, with Earth observations, additional in situ

1 and ground-based observations may be available to reduce uncertainty in satellite overlap. However, a challenge to
2 Earth observations is the need for direct temporal and spatial overlap, which can be difficult or impossible as
3 satellite observation approaches are considered within challenging budgetary constraints. All of the techniques
4 outlined here can be applied to identifying the level of overlap needed with existing satellite observations and next
5 generation observations or reference calibration satellites, such as proposed by CLARREO. In all cases, the level of
6 uncertainty in offsets and drifts will be determined not by the length of overlap, but in the quality of the match-up
7 between the reference and operational satellites. Under various constraints, choices of overlap can be optimized to
8 help assure climate records that are appropriate for advancing our understanding of the Earth system. The goal of
9 achieving the most stable observational data from existing and future observations is fundamental to understanding
10 the Earth and potential long-term changes. The value of this paper is the ability to estimate, either prior to satellite
11 launch or soon after satellite launch, the amount of time needed to achieve or verify tolerance for a stable merged
12 satellite record using objective criteria.

13

14 **Data and code availability:** Example data and code used will be available from the first author on request
15 (Betsy.Weatherhead@Colorado.edu).
16

17

18

1 **Appendix A: Comments on the usefulness of monthly data.**

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

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

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

	Mean (watts m ⁻² nm ⁻¹)	Standard Deviation (watts m ⁻² nm ⁻¹)	Auto- correlati on	Standard Deviation of de-trended Data (watts m ⁻² nm ⁻¹)	Auto- correlation of de- trended Data	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 0.00008 watts m ⁻² nm ⁻¹ year ⁻¹
SIM							
Monthly	0.07839	4.78×10^{-4}	0.939	1.528×10^{-4}	0.429	43.6	3.27
Solar	0.07837	4.70×10^{-4}	0.945	1.499×10^{-4}	0.460	47.0	3.31
SOLSTICE							
Monthly	0.07907	3.55×10^{-4}	0.890	1.733×10^{-4}	0.544	13.0	3.94
Solar	0.07905	3.41×10^{-4}	0.898	1.639×10^{-4}	0.533	12.9	3.75
SOLSTICE-SIM							
Monthly	6.802×10^{-4}	1.67×10^{-4}	0.890	8.586×10^{-5}	0.570	4.94	2.52
Solar	6.831×10^{-4}	1.69×10^{-4}	0.901	8.363×10^{-5}	0.609	5.70	2.58

20
 21 **Table A1. Fundamental descriptive values and calculations of overlap periods were calculated using monthly averaged**
 22 **data and data that were averaged on a 27-day time period, which is a more natural timeframe for these calculations. We**
 23 **note that little impact is observed from this small change in averaging period.**

24
 25 For calculations of detecting an offset, the magnitude of variability and autocorrelation of the data were used; for the
 26 calculations of detecting a drift, the magnitude of variability and autocorrelation of the detrended were used. The
 27 calculations for small number of months in the final column are adjusted from a 1.96 factor to a 2.8 to account for
 28 the uncertainty in small sample size. We note that the differences observed are remarkably small. Differences likely
 29 would have been larger if we had used data of a shorter duration (e.g. two years of data instead of just over three
 30 years). One note is that when the solar rotation period is used for averaging, 42 data points are derived, as opposed
 31 to the 39 data points derived from monthly averages. This “larger number” of data points is accompanied by slightly
 32 lower standard deviation and nearly constant auto-correlation and directly feeds into the standard error calculation.
 33

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

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

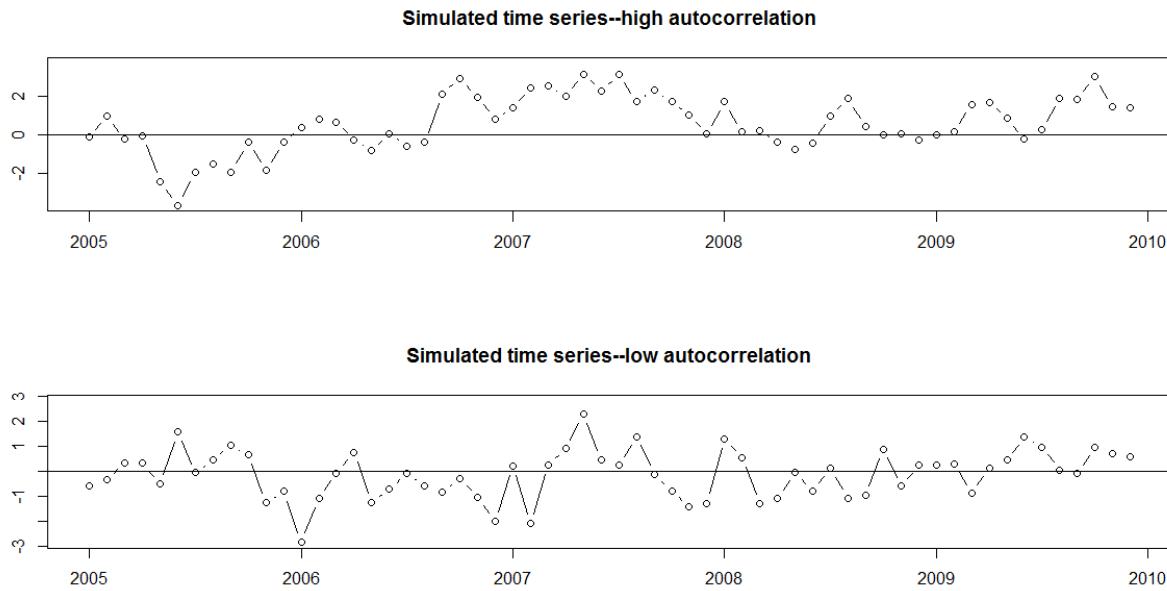
7 $Months\ to\ Estimate\ an\ Offset \cong 1.96^2 \sigma^2 / Offset\ Limit^2 \frac{1+\varphi}{1-\varphi}$ (2)

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

9 $Months\ to\ Estimate\ a\ Drift \cong 12 * \left[1.96 \frac{\sigma}{|trend|} * \sqrt{\frac{1+\varphi}{1-\varphi}} \right]^{2/3}$ (3)

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

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



17
 18 **Figure B1.** Two simulated time series are shown with high autocorrelation (top plot) and low autocorrelation (bottom
 19 plot).

20 In situations of high autocorrelation (0.7 in the top plot of Figure B1), a time series can deviate from the long-term
 21 mean (0 in both of these simulated time series) for many months. If a short time period is used to estimate phi, likely
 22 phi will be under-estimated and the error on the sample mean may be farther from the true population mean
 23 compared to a situation with low autocorrelation (0.2 in the bottom plot of Figure B1). For the data used in this
 24 paper, the autocorrelation is estimated at 0.1 once drifts are accounted for and therefore the overlap period of six
 25 years is more than adequate to derive a good estimate for the long-term value of phi. To test for AR(1) behavior in
 26 the SOLSTICE-SIM monthly overlap data, we calculate a partial autocorrelation function out to fifteen terms.

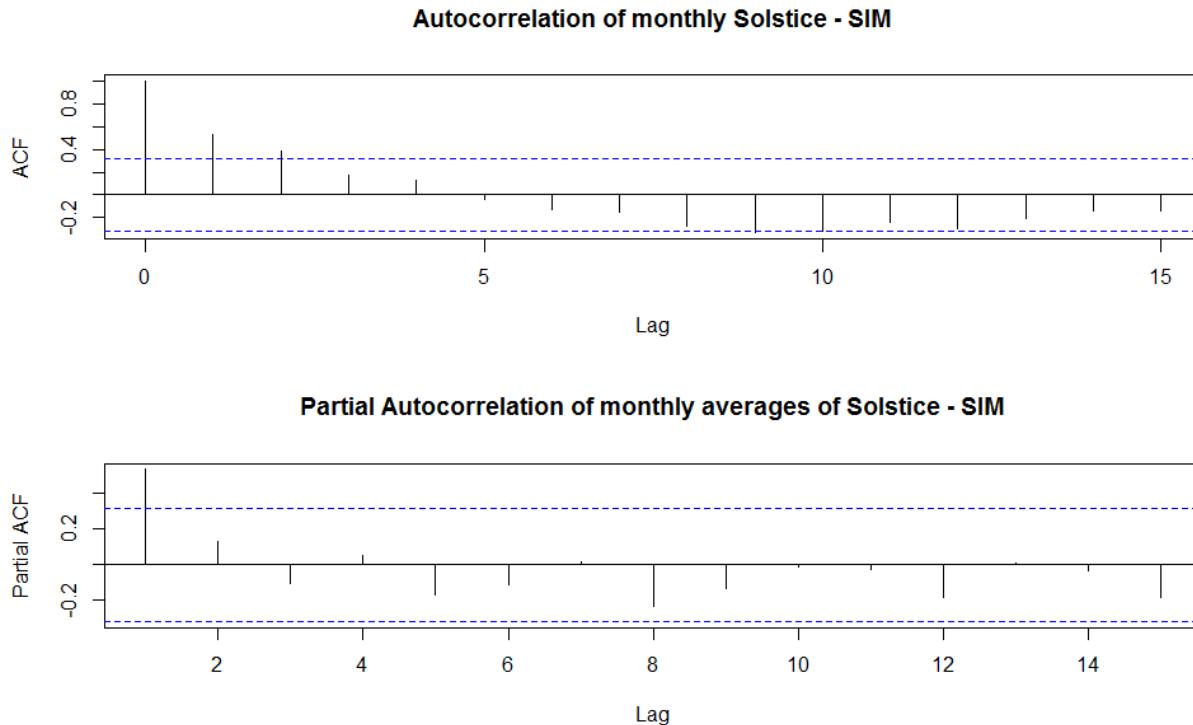
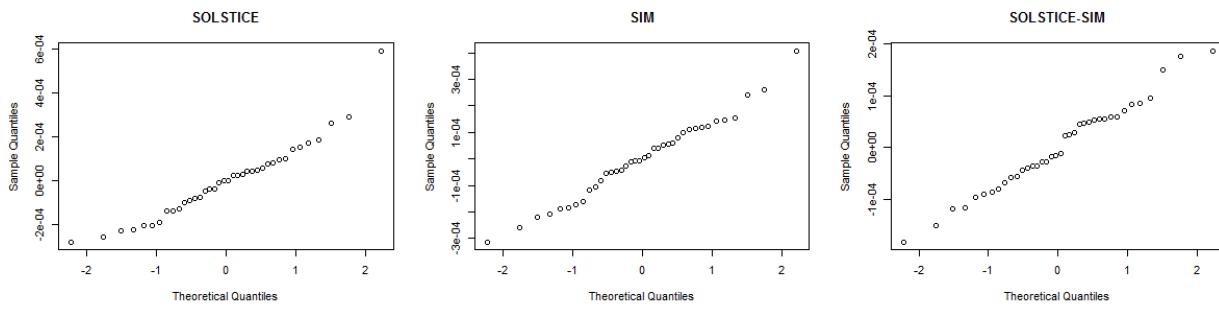
1
23
4
5
6
7

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 dashed blue lines indicate thresholds of statistical significance for results. Each lag, for this dataset represents calculations based on differences of one month.

8
9
10
11
12
13
14
15
16
17
18

Figure B2 shows both an AutoCorrelation Function (ACF) for the SOLSTICE-SIM data (top plot) and a Partial Autocorrelation Function (PACF) for the SOLSTICE-SIM data (bottom plot). The plots allow confirmation that the residuals of the data behave as an AR(1) process: the ACF shows significant autocorrelation for two months lag; the PACF shows that there is no significant correlation once a lag 1 correlation is accounted for. 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 Q-Q plot (e.g. Hamilton, 1994; Box et al., 2015):

19
20

1 **Figure B3. These three Q-Q plots compare the monthly averaged SOLSTICE, SIM and SOLSTICE-SIM differences to**
2 **theoretical Gaussian distributions.**

3 Q-Q plots, such as those shown in Figure B3, compare the distribution from a pure Gaussian distribution to the
4 distribution of deseasonalized data used in this study. The roughly linear relationship demonstrated in the Q-Q plots
5 of B3 shows that the three datasets do behave in a close to Gaussian nature and thus the use of Eq. (2) and Eq. (3)
6 are supported for the analyses presented in this paper.

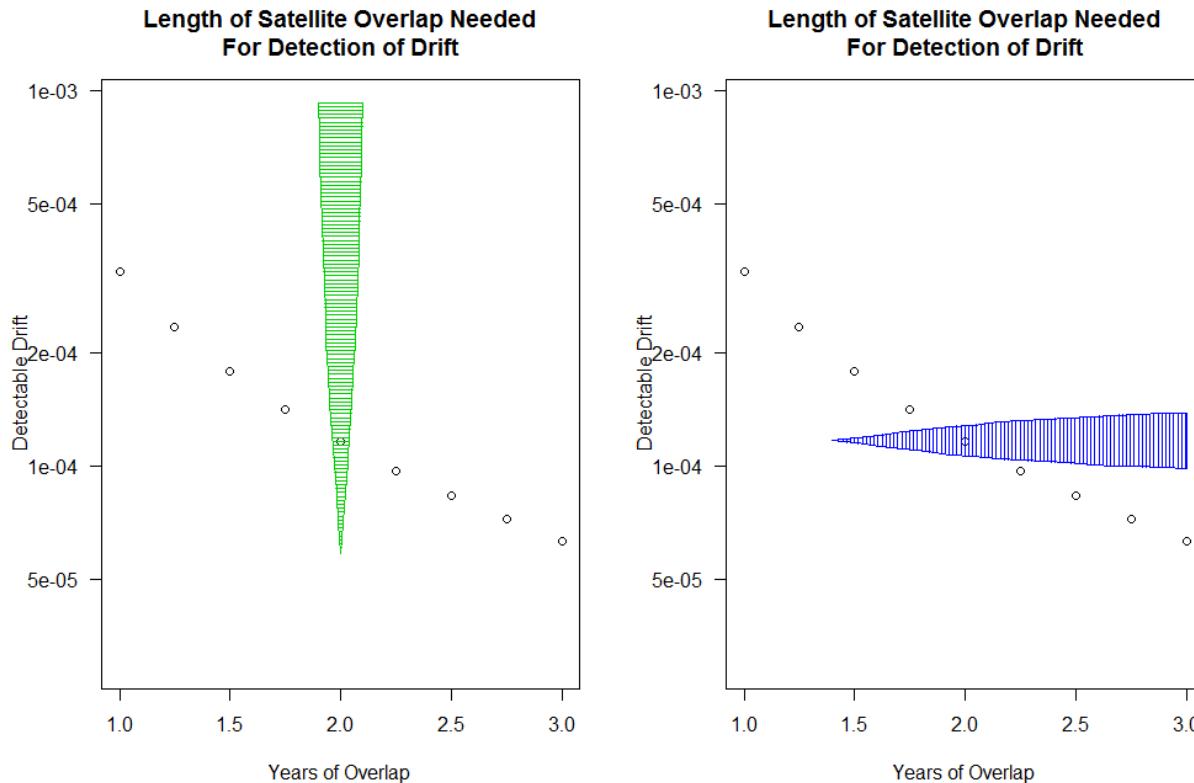
7 **Appendix C: Comments on the interpretation of time estimates.**

8 Any estimate of how long it will take to correctly identify a drift must be taken with some level of understanding of
9 how this estimate is made and what can be expected from using these estimates. Figure 5 offers estimates for a range
10 of times needed to estimate specific drifts, assuming no jumps occur in the record. As a reminder, this plot was
11 created assuming the type of overlap seen in the SOLSTICE-SIM overlap period; specifically, the calculations
12 assume the amount of variability and autocorrelation observed in the differences (shown in the second plot of Fig.
13 4). However, different observing systems are likely to have different levels of agreement.

14 When deriving drifts on existing data, only a single level of certainty is required: for example, what does it mean to
15 detect a drift? Often the community has focused on detection at a 95% confidence level or a 99% confidence level.
16 However, when estimating how long it will take to detect a drift, two statistical levels are required: one that
17 identifies what is meant by “detecting a drift” and one that identifies the likelihood that a drift will be detected in the
18 specified period of time, if that level of drift is the true, long-term drift in the overlap. For the first, we consider
19 detecting a drift to mean identifying a drift that, with 95% likelihood, is not zero, although other levels may be
20 considered. For the second, we consider the likelihood of detecting the drift (at the 95% confidence level) to be
21 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
22 time.

23 There are no error bars in Fig. 5. We’d like to begin the discussion of appropriate error bars in this section. As stated
24 in the previous paragraph, the data in Fig. 5 represent estimates of how long it will take to detect a specific level of
25 drift. If we focus on a single point, for instance the two year point that indicates a drift of 1.2×10^{-4} watts $m^{-2} nm^{-1}$
26 $year^{-1}$ could be detected, it is possible that a slightly smaller drift could be detected in that two year of overlap, if the
27 variability happens to result in a signal-to-noise for the overlap period that is slightly more favorable. Similarly, if
28 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
29 would be detectable within the two years. So the “error bars” on this one point would be slightly below the current
30 point and would extend infinitely upward, indicating that much larger drifts could be detected in the two year period.

31 Extending our discussion of error bars in Fig. 5, we can similarly think in terms of horizontal error bars. Again,
32 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
33 years, this drift, if it is the true underlying drift, may be detectable a few months shy of two years or may take a few
34 months more than two years. As stated above, the two years is a 50% likelihood of detection. It is highly likely that
35 such a drift could not be detected in a few months of monitoring, but it would very likely be detected in ten years of
36 monitoring. So, again, we have error bars that are non-standard in that they extend to the left in the plot and continue
37 indefinitely to the right.



1
2 **Figure C1.** Estimates of how long it will take to detect a drift can be interpreted as the likely time needed. Depending on
3 variability present, even small drifts can be detected (although with less than 50% likelihood of detection), probability
4 indicated by the width of the green area. For a given drift level, there is a chance that the drift can be detected in less than
5 the number of years indicated, although that likelihood is less than 50% for time less than the times indicated, and the
6 probability indicated by the blue area.

7 If we want to express this uncertainty of likelihood of detection in a visual manner, we could employ two
8 dimensional error bars, similar to violin plots which are often employed to express variable information. Figure C1
9 shows the likelihood of detecting a particular drift with two years of overlap. For drifts considerably smaller than
10 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
11 green area. Larger drifts can be detected with higher likelihood. Figure C1 shows the likelihood of a true drift of
12 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
13 likelihood, with the linear scale being defined such that the likelihood of detection being 50% at two years;
14 considerably higher likelihood of detection is indicated with more years of overlap. There is also a small, but less
15 than 50% likelihood that the true drift might be detected in less than two years; again, the height of the blue bars
16 indicates the likelihood of detection.

17
18

1 **Appendix D: Mathematical structure for optimizing overlap decision choices.**

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

6 $\mathbf{NB} = \mathbf{TB}\{IVC[q(o)]\} - \mathbf{TC}[o]$ (D1)

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

14 If the objective of “optimizing” satellite observation overlaps is to achieve a specific quality standard, economists
 15 would approach this as a cost-effectiveness issue. In this case, the objective is to minimize the costs to achieve the
 16 exogenously determined standard. There is no consideration of benefits in this case, o is set to achieve \bar{q} so
 17 $\mathbf{TB}\{IVC[\bar{q}(o)]\}$. The outcome is not necessarily societally optimal – the standard could be too strict in which
 18 case net benefits could be negative or the standard could be too lax in which case it would be possible to improve
 19 societal outcomes by increasing the observational overlap period and increase net benefits.³

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

23 $\partial \mathbf{NB} / \partial o = \partial \mathbf{TB}\{IVC[q(o)]\} / \partial o - \partial \mathbf{TC} / \partial o = \frac{\partial \mathbf{TB}}{\partial IVC} \frac{\partial IVC}{\partial q} \frac{\partial q}{\partial o} - \partial \mathbf{TC} / \partial o = \mathbf{MB}_o - \mathbf{MC}_o$ (D2)

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

1 To dependably quantify societal benefits from satellite observations requires understanding the complex relationship
2 between individual instruments, data streams, modeling, communication, decision making, and potential and actual
3 societal outcomes. This requires understanding stakeholders and processes of information creation, transformation,
4 transmission, and use in decision making along the entire information value chain through multiple stakeholders
5 with a variety of objectives, resources, and constraints. Cooke et al. (2014) develop an illustrative example of such
6 an analysis using the social costs of carbon (SCC) as a measure of societal benefits and a hypothetical decision
7 framework (i.e., a tipping point that would lead to global climate impact mitigation efforts).
8

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

3
4 **Author contribution:** Jerald Harder conceived of the idea for this paper and the general techniques to be used; he
5 also provided the data for analysis. Elizabeth Weatherhead carried out the calculations, including adapting statistical
6 approaches to the problem and oversaw the writing. Eduardo Araujo worked on the calculations and identified
7 stability requirements for solar and environmental data. Larry Flynn supplied input on GSICS and complementary
8 methods of verifying stability in satellite records as well as cross verification of results, including those presented in
9 Table 1. Stacey Frith and Mark Weber provided valuable input on merging of satellite records and applicability of
10 results to developing long-term records. Greg Bodeker offered data and insights into merging of satellite
11 temperature datasets. Jeff Lazo, in consultation with the other co-authors, provided the optimization methodology
12 and economic analysis. Thomas Woods and Peter Pilewskie, as respective Principal Investigators for the NASA
13 SORCE and the upcoming TSIS missions advised on the requirements and needs for mission overlap for the solar
14 spectral irradiance climate record. Jason English supplied input on applicability to climate studies.

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

17 **Disclaimer:** None

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

22 References

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

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

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

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

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

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

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

40 Bodeker, G. E., Scott, J. C., Kreher, K., and McKenzie, R. L.: Global ozone trends in potential vorticity coordinates
41 using TOMS and GOME intercompared against the Dobson network: 1978–1998, *J. Geophys. Res. Atmos.*,
42 106(D19), 23029-23042, 2001.

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

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

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

5 Chander, G., Helder, D. L., Aaron, D., Mishra, N., and Shrestha, A. K.: Assessment of spectral, misregistration and
6 spatial uncertainties inherent in the cross-calibration study, *IEEE T. Geosci. and Remote Sensing*, 51, 1282-
7 1296, 2013a.

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

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

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

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

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

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

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

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

24 Dudok de Wit, T.: A method for filling gaps in solar irradiance and solar proxy data, *Astron. Astrophys.*, 533, A29,
25 2011.

26 Eckert, E., T. von Clarmann, M. Kiefer, G. P. Stiller, S. Lossow, N. Glatthor, D. A. Degenstein, L. Froidevaux, S.
27 Godin-Beckmann, T. Leblanc, S. McDermid, M. Pastel, W. Steinbrecht, D. P. J. Swart, K. A. Walker, and
28 P. F. Bernath. Drift-corrected trends and periodic variations in MIPAS IMK/IAA ozone measurements,
29 *Atmos. Chem. Phys.*, 14, 2571-2589, 2014.

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

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

36 Fioletov, V. E., G. E. Bodeker, A. J. Miller, R. D. McPeters, and R. Stolarski, 2002: Global and zonal total ozone
37 variations estimated from ground-based and satellite measurements: 1964–2000. *J. Geophys. Res.*, 107,
38 4647, doi:10.1029/2001JD001350.

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

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

1 Frith, S. M., Kramarova, N. A., Stolarski, R. S., McPeters, R. D., Bhartia, P. K., and Labow, G. J.: Recent changes
2 in total column ozone based on the SBUV Version 8.6 merged ozone data set, *J. Geophys. Res. Atmos.*,
3 119(16), 9735-9751, 2014.

4 Fruit, M., Gusarov, A., and Doyle, D.: Testing and qualification of optical glasses for use in a space radiation
5 environment; the advantages and pitfalls of using a parametric approach, *Proc. SPIE Int. Soc. Opt. Eng.*,
6 4823, 132, 2002.

7 Fröhlich C.: Evidence of a long-term trend in total solar irradiance, *Astron. Astrophys.*, 501, L27-L30, doi:
8 10.1051/0004-6361/200912318, 2009.

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

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

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

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

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

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

21 Hubert, D., and 39 co-authors, Ground-based assessment of the bias and long-term stability of 14 limb and
22 occultation ozone profile data records, *AMT*, 2016,

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

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

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

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

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

35 Laxminarayan, R., and Macauley, M. K. (Eds.): The value of information: methodological frontiers and new
36 applications in environment and health. Springer Netherlands. 304 pages, doi: 10.1007/978-94-007-4839-2,
37 2012.

38 Loeb, N. G., Nanalo-Smith, N., Su, W., Shankar, M., and Thomas, S.: CERES top-of-Atmosphere Earth radiation
39 budget climate data record: Accounting for in-orbit changes in instrument calibration, *Remote Sens.*, 8,
40 182, doi: 10.3390/rs8030182, 2016.

41 Logan, J. A., Staehelin, J., Megretskaya, I. A., Cammas, J. P., Thouret, V., Claude, H., & Fröhlich, M. Changes in
42 ozone over Europe: Analysis of ozone measurements from sondes, regular aircraft (MOZAIC) and alpine
43 surface sites. *Journal of Geophysical Research: Atmospheres*, 117(D9), 2012

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

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

3 MacDonald, A. E.: A global profiling system for improved weather and climate prediction, *Bull. Am. Met. Soc.*, 86,
4 1747, doi: 10.1175/BAMS-86-12-1747, 2005.

5 Marchenko, S. V., and DeLand, M. T.: Solar spectral irradiance changes during cycle 24, *The Astrophys. J.*, 789, 2,
6 117, doi: 10.1088/0004-637X/789/2/117, 2014.

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

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

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

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

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

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

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

22 Ohring, G., Wielicki, B., Spencer, R., Emery, B., and Datla, R.: Satellite instrument calibration for measuring global
23 climate change: report of a workshop, *Bull. Am. Meteorol. Soc.*, 86, 1303-13, doi: 10.1175/BAMS-86-9-
24 1303, 2005.

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

27 Penckwitt, A. A., G. E. Bodeker, P. Stoll, J. Lewis, T. von Clarmann, and A. Jones. Validation of merged MSU4 and
28 AMSU9 temperature climate records with a new 2002–2012 vertically resolved temperature record,
29 submitted to AMTD 8, 235–267, 2015.

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

34 Rahpoe, N., Weber, M., Rozanov, A. V., Weigel, K., Bovensmann, H., Burrows, J. P., and Laeng, A.: Relative drifts
35 and biases between six ozone limb satellite measurements from the last decade, *Atmos. Meas. Tech.*, 8, 10,
36 4369-4381, 2015.

37 Randel, W. J., and Thompson, A. M.: Interannual variability and trends in tropical ozone derived from SAGE II
38 satellite data and SHADOZ ozonesondes, *J. Geophys. Res.*, 116, D07303, doi: 10.1029/2010JD015195,
39 2011.

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

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

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

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

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

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

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

13 Stolarski, R., and Frith, S.: Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data
14 record: the importance of instrument drift uncertainty, *Atmos. Chem. Phys.*, 6, 4057-4065. doi:
15 10.5194/acp-6-4057, 2006.

16 Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., ... & Jones, A. SPARC Data
17 Initiative: A comparison of ozone climatologies from international satellite limb sounders. *Journal of
18 Geophysical Research: Atmospheres*, 118(21), 2013.

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

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

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

26 Toohey, M., Hegglin, M. I., Tegtmeier, S., Anderson, J., Añel, J. A., Bourassa, A., and Brohede, S.: Characterizing
27 sampling biases in the trace gas climatologies of the SPARC Data Initiative, *J. Geophys. Res. Atmos.*, 118,
28 20, 2013.

29 Toohey, M. and von Clarmann, T.: Climatologies from satellite measurements: the impact of orbital sampling on the
30 standard error of the mean. *Atmo. Meas. Tech.*, 6(4), pp.937-948, 2013.

31 Viereck, R. A., Floyd, L. E., Crane, P. C., Woods, T. N., Knapp, B. G., Rottman, G., Weber, M., Puga, L. C., and
32 DeLand, M. T.: A composite Mg II index spanning from 1978 to 2003, *Space Weather*, 2, S10005, doi:
33 10.1029/2004SW000084, 2004.

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

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

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

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

1 Weatherhead, E. C., Tiao, G. C., Reinsel, G. C., Frederick, J. E., DeLuisi, J. J., Choi, D., & Tam, W. K. Analysis of
2 long-term behavior of ultraviolet radiation measured by Robertson-Berger meters at 14 sites in the United
3 States. *Journal of Geophysical Research: Atmospheres*, 102(D7), 8737-8754, 1997.

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

6 Weber, M., W. Steinbrecht, C. Roth, M. Coldewey-Egbers, D. Degenstein, V.E. Fioletov, S. M. Frith, L. Froidevaux,
7 J. de Laat, C. S. Long, D. Loyola, and J. D. Wild: [Global Climate] Stratospheric ozone [in "State of the
8 Climate in 2015"], *Bull. Amer. Meteor. Soc.*, 97, S49–S51, 2016a.

9 Weber, M., S. Dikty, J. P. Burrows, H. Garny, M. Dameris, A. Kubin, J. Abalichin, and U. Langematz, The Brewer-
10 Dobson circulation and total ozone from seasonal to decadal time scales, *Atmos. Chem. Phys.*, 11, 11221-
11 11235, doi:10.5194/acp-11-11221-2011, 2011.

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

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

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

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

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

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

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

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

27 WMO, UNEP. Scientific Assessment of Ozone Depletion: 2014.

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

30 Wulfmeyer, V., Hardesty, M., Turner, D., Behrendt, A., Cadeddu, M., Di Girolamo, P., Schlüssel, P., van Baelen, J.,
31 and Zus, F.: A review of the remote sensing of lower-tropospheric thermodynamic profiles and its
32 indispensable role for the understanding and the simulation of water and energy cycles, *Rev. Geophys.*, doi:
33 10.1002/2014RG000476, 2015.

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

36