

# Response to Reviewers

We would like to sincerely thank both reviewers for helping us make this a significantly better paper. Specifically, we feel the paper has benefited by inclusion of more geophysical datasets and a re-organization that has resulted in a reduction in redundancy. We are grateful for the thought and care that both reviewers showed in their reading and comments. Specific responses are below. Our responses are in black font.

## Anonymous Referee #1

Received and published: 27 February 2017

### General comments

This paper examines the issue of quantifying differences in satellite measurements sets based on overlapping measurement periods, addressing the question of how long overlap periods need to be to accurately estimate offsets and drifts between instruments. A few general formulas are presented to calculate required overlap periods for given desired precision requirements of offset and drift estimation, and examples are presented.

While the issues discussed are definitely relevant to the construction of long-term atmospheric data records, this paper does not actually directly deal with any atmospheric measurements, instead focusing on examples based on satellite measurements of solar irradiance. As a result, the fit between this paper and ACP (or even AMT for that matter) is somewhat questionable. Readers of ACP would likely benefit greatly from examples using actual atmospheric data, and given the ready availability of multi-instrument data sets like stratospheric ozone (e.g., Tegtmeier et al., 2013), it wouldn't be hard to include such examples. Of course, atmospheric data, with the temporal and spatial variability that comes with it, may present some additional complications to the analysis (which is mentioned in passing in the manuscript), but a discussion of these complications seems warranted in such a paper if it truly wants to address the analysis (and merging) of atmospheric data.

We appreciate this comment and in response have included two additional atmospheric datasets (ozone and temperature) to help illustrate the problems in Earth observations from satellite and the importance of these techniques for estimating appropriate satellite overlap. We show how atmospheric data have more complications than solar irradiance data for a variety of reasons including spatial and temporal matchup as well as differences with latitude and season. We have also added more clarity on the importance of solar data to understanding atmospheric behavior, including variability, climatology and change.

The utility of this paper to atmospheric community could also be improved by a fuller description of the general implications of the analysis before descending to the focused solar irradiance example. For example, Fig 3 displays the detectable drifts in the solar irradiance data sets as a function of years of overlap, which suggests a general form of

the solution, but won't provide any quantitative information to anyone working with other data sets. Instead, a plot of ratio of drift to variability as a function of n (perhaps for different sample values of autocorrelation) would be directly relevant to users of other data sets.

We have now highlighted the general applicability of this analysis. We have also changed Plot 3 (now Plot 5), to include results for various levels of variability and autocorrelation. This is in direct response to the reviewer's suggestion and we feel that it improves the paper. While we assume anyone making use of this paper will make use of the formulae not the figures, we feel that Plot 5 is now more useful in helping users understand the implication of the formulae.

Not written in this paper, although I think the two reviewers seem to have understood this point, is that we would like decisions about pre-flight calibration, satellite overlap, etc. to be based on scientific criteria and not on bureaucratic guesses as to what level of overlap is adequate. We do not feel this paper has completely answered the question of overlap, but we do hope we have helped start a series of scientific discussions on the subject.

Stylistically, I found the paper repetitious in places, often returning to discussions of issues that aren't, in my opinion, of central importance. For example, the issue of requirements (or the desire) for self-calibrating, consistent systems for atmospheric measurements is often brought up, but this paper deals specifically with techniques to deal with situations where measurements are not self-calibrated. This point can be made succinctly in the introduction, and thereafter neglected, at least until the discussion and conclusions.

We have taken these comments to heart and agree with the reviewer. In response, we have brought all discussions of self-calibrating satellites into one place. Similarly, we brought discussions on the complications of Earth observations (spatial match-up, etc.) into one place. We have merged the discussion of what had previously done on offsets, drifts and jumps that occurred in the rather long introduction into the body of the paper where we deal with offsets, drifts and jumps. In doing this merging we were able to eliminate many redundancies. Again, we sincerely thank the reviewers for pointing this out. The paper is greatly improved by these comments.

#### Specific comments:

Pg 1, 127: "offset or a drift in the offsets": If "offset" is singular in the first case, then "drift in the offset" seems more appropriate. But the sentence was confusing to me at first, and I wonder if just "drift" is easier to understand.

We have made the change and tried to make the point more clear.

Pg 1, 137-38: "may also benefit. . ." this issue is not dealt with in any substantial way in the paper, so this statement's inclusion in the abstract seems superfluous.

Thank you. This is dealt with in the re-write of our paper by including examples of ozone and temperature in Section 2 of the revised manuscript.

Pg 2, 116: “tying data to absolute reference standards with the intent of developing traceability to reference standards” sounds a little tautological.

We agree. This wording has been completely redone—see the new wording in the response to the next point.

Pg 2, 117: It’s not clear to me why reference standards are brought into the argument here, is the point that if one of the two overlapping measurement sets is a standard, then you can extend a standard through identification of an offset and drift in the second instrument?

Thank you. We have hoped to make the point more clear with the following:

Page 2, Line 35: *“Another approach to addressing satellite uncertainty, based on maintaining traceability through on-board calibration capabilities using absolute references, has been advocated through the CLARREO and TRUTHS programs (Wielicki et al., 2013; Fox et al., 2013). For both programs, verification of merging of these new approaches will be important for validation of expected agreement.”*

Pg 2, 136: Do “wavelength scale corrections” etc. really help instrument scientists understand the fundamental observations? Or does an understanding of the fundamental observations allow for valid corrections?

Thank you. We think the wavelength scale corrections truly are helpful—and nearly universally examined by satellite instrument scientists. We need to understand these details of the instrument (including those related to wavelength scale corrections) before we can understand the observation. All of these effects have a different influence on the data. All of these components are terms in the measurement equation, and if there is something ‘missing’ in the measurement equation then it is possible for the observation to be misinterpreted. We have not modified this sentence in the text, but hope the new context makes the point more clear. The new wording starts on Page 2, around line 32.

Pg 3, 115: Does removing a bias affect the precision of the merged data set? And, does one really need to remove a bias to identify a drift? If you look at changes with time (time derivatives) the absolute value doesn’t matter.

This is part of a re-written paragraph. We believe that the answer to the question: ‘Does removing a bias affect the precision of the merged data set?’ is yes it does. If left in place in a merged data set, a bias in one component of a merged dataset can be misinterpreted as drift. This situation has much in common with our discussion on jumps – see the discussion regarding Figure 6 in our paper. The ability to use time derivatives is strongly dependent on the signal-to-noise ratio of the measurement and the autocorrelation of the data. Our revised Figure 5 helps illustrate this matter. We hope the new text is clearer.

Pg 3, 128: It would seem that the paper is of interest to a wider group than just the users of merged data sets, specifically to the creators of merged data sets.

Thank you. We have adjusted the paper so that users of merged data sets may be able to more fully appreciate the uncertainties add to the data—uncertainties that aren't always fully explained when one is interested in downloading a dataset. The addition of creators as an important group is now added on Page 3, Line 19-20.

Pg 3, ll40: The first two paragraphs of Sec 1.1 have no apparent specific connection to “Offsets”, and seem to set the scene for an analysis of ozone data which never arrives. Actually, there doesn't seem to be much of any specific introduction of the issue of offsets in this subsection.

Thank you. I believe this is addressed in the larger re-write of the manuscript.

Pg 3, l32: “but this will not. . .” If two measurement sets were both traceable to a reference, why wouldn't this fully address the challenge of merging the data sets? Is the calibration referred to here only at a single time, or could it be continuous?

Removed in revised manuscript. The point we still want to make is for the case when only one measurement is traceable. I think the re-write addresses this on See Page 2, line 21. I believe it is now clearer. However, to the reviewer's point of two traceable datasets should make merging of datasets quite simple, I would still say that challenges will exist. For instance, if each of two temperature satellites are traceable with an uncertainty on the temperature of  $\pm 1$  degree, we could easily have an offset between the two satellites of 2 degrees. If the uncertainty on the traceable products is quite, quite small, then I agree with the reviewer, although we still probably want to verify.

Pg 3, l36: Temporal changes in sampling can also contribute to drifts. C3

Thank you. We agree. We now include this on page 3 starting at line 12.

Pg 3, l39: If drifts in ozone were up to 5%, but were statistically insignificant, then the case of ozone seems to be very different than that of solar irradiance.

~~The objection here is unclear.~~ Indeed, the ozone and irradiance cases are very different, but the kind of analysis we perform here is applicable regardless of the case under study. The climate data record mandates different levels of knowledge for these two cases. We address this on Page 6 starting at line 13.

Pg 5, l9: A clear definition of “jump” is needed: I assumed it to be the instantaneous addition of a constant offset, but if a jump “can last from less than a few hours to multiple years” it sounds more like a two-step process.

Thank you. We now write: *“Jumps are permanent or semi-permanent level shifts in the data that occur at specific points in time and are not attributable to the parameter being observed; jumps could represent a change in sensitivity of an instrument or a change in location or orientation of the satellite.”*

Pg 5, l15: Why is the requirement for a long-term stable record difficult to justify?

Context changed in the formulation of section 1.2 ‘Planning for need homogeneity’. Some of the difficulty comes from the concept that we don’t want to require stability that is unattainable. There can be a bit of a dance between what we want and what is possible. Another reason for the difficulty is that there are many, many uses of any environmental dataset, so the requirement for one application may be different from the requirement (on the same dataset) for another application. It may be tempting to take the smallest requirement from any user community, but strict stability can be hard to achieve and very expensive.

Pg 10, 17: “better behaved” is not very helpful: this sentence doesn’t explain what monthly data is better than.

The sentence has been rewritten to avoid the unhelpful and unclear wording. New text is on Page 12, line 4: *“These constraints for the formula are some of the reasons that monthly averages are used as often as they are: monthly averaged data remove higher frequency noise and sampling match-up problems from different instruments are minimized.”*

Pg 10, 112: Given that the example below gives a case in which “1.96” is not the valid multiplier, it would seem appropriate to replace “1.96” in equation 2 with a placeholder variable for the student-t distribution (as a function of n). Otherwise, a quick reader may overlook the fact that 1.96 holds only for large n.

We have rewritten this section and believe the issues are now addressed appropriately on page 12, line 35. We worry that removing the 1.96 from the formula would make it less useful and intuitive. In reality, most applications will be in the large n limit where 1.96 holds or is at least close.

Pg 11, 18: “we can increase our measurements per month” in this example case, but not in all circumstances. I think the point is that with enough measurements, the random measurement errors in the mean are small enough to ignore, the only source of variance is the natural variability.

Small changes in the text have been made to clarify this point. I’m not sure that it is safe to write that “the only source of variance is the natural variability.” If this were true, good observations would match up perfectly in terms of variability, but they don’t for reasons now more clearly itemized in the text.

Pg 13: 11: A short derivation of Eq. 4 would be useful here if possible, otherwise the term introduced to account for the jump is not intuitive.

We have added a new equation that shows how Eq. 4 is derived. The added equation and associated sentences address this concern—those interested can derive what is now equation (5) from given equation (4), which is the new addition. This work has already been published in Weatherhead et al., 1998.

Pg 13: 16-8: These sentences talk about fitting of the offset and drift, simultaneously and

sequentially. However, to this point there has been no discussion of “fitting”, only using the equations to estimate the length of time needed to estimate an offset or drift. How does the concept of fitting, simultaneously and sequentially, affect the use of Equations 2-4?

Thank you. We’ve adjusted the text so that we talk about the future fitting of the data. We hope this is more clear. Fundamentally, how long we need to monitor depends on how we will fit to an offset and a drift term in the data. We now write in terms of, “...*If we assume that we are going to fit the environmental data to a linear statistical model of the form:...*” We hope this is now clear.

Pg 16, 14: Is a reference really needed to support the statement that “Earth observations often invoke spatial and temporal variations”?

Statement lost in rewrite.

Pg16, 117: An error in the drift which is half the trend one is seeking to detect seems large: does it mean that in a worse-case, the detection of the trend might take twice as long (as the case with no drift in the measurement?). How was this threshold decided?

This is being introduced in this paper after multiple discussions among co-authors who correctly pointed out that for many parameters there is not requirement for an error in the drift. WMO, under the guidance of Global Climate Observing System, is working on establishing these requirements, but the choices of drift limit are, in most cases quite arbitrary. We agree that this drift is large, but it is also quite likely that this large of a drift (half the size of the trend) is not attainable for many important observations. Evaluation of stability and limitations of drift for various parameters must be done very, very carefully and is therefore beyond the scope of this paper.

Pg 16, 142: Given only Fig 3, the optimal is obviously as many years as possible. The optimality issue only is apparent when you consider the costs (which are discussed below).

This section is reworded in a very similar way. Please see page 20, line 31.

Pg 17, 116: First, a subjective evaluation (“nice”) of the work of Morss et al. (2005) is probably not appropriate here, and secondly, there’s not much in the sentence to really inform the reader of the relevance of this work to the present study: “a case study based on primarily hypothetical valuation estimates” doesn’t help much.

On page 21, line 5 the sentence now reads:

Morss et al. (2005) provide an overview of relevant economic concepts and theory for optimal design of observational systems based on benefit cost tradeoffs.

Editorial comments Pg 1, 133: delete “may” Pg 1, 136: either “Extensions . . . are” or “Extension . . . is”

done

Pg 2, 12: “assess the stability” relates to identifying and quantifying drift, but much of the paper deals also with identifying and quantifying offsets, so I wonder if this first sentence of the paper should be more general.

Thank you. It now reads, “...assess the characteristics..”

Pg 2, 18: The “sensitivity degradation mechanisms” described in the prior sentence will firstly impact the individual satellite record, not just the merged record referred to here.

We agree with the reviewer on this point. For an individual satellite record, analysis of the degradation measurement equation must be evaluated. Data are adjusted for all known factors. But only through the process of comparing simultaneous data records (i.e. overlapping) can any potential systematic errors be evaluated. We have not edited this comment since the purpose of the paper is to discuss the merging of records, but we think the new context makes it more clear.

Pg 2: 128: “Should an offset. . .”, This sentence joins two statements with a semicolon, but it’s not clear how or why the two statements are linked.

Sentence rewritten

Pg 3, 143: why “potential”?

‘potential’ removed

Pg 5, 1135-40: Some pretty general material here which seems repetitive to the introduction.

We believe we have cut down significantly on the redundancy. The new sentence reads:

“The techniques discussed herein are applicable to instrument scientists pursuing improvements in on-board instrument corrections, but also for mission planning by program managers to ensure the best overlap characteristics of adjoining missions.”

Pg 6, 119-10: Intra-sentence repetition.

On page 7, line 40, the sentence now reads:

*The instruments for the SORCE mission are described in a series of papers published in Solar Physics related to the design, operation, calibration, and performance of the SORCE instruments. Harder et al. (2005a) describes the scientific requirements, design, and operation modes for the instrument.*

Pg 6, 123: deploy->deployment

Pg 6, 126: Here w

done

Pg 10, 110: Here and elsewhere, “years” are discussed in the text, while the equation is

written in terms of months.

Cleared up in text modifications. We now present the equations as years and the words are more often in “length of time” or “amount of time.” We have only left “months” when we are referring to fractions of a year, or think it is more clear

Pg 10, 136: why brackets around “%”? Pg 10, 147: The sentence which includes the equation seems to not quite make sense. Pg 11, 12: do you not specify the drift, rather than “estimate” it?

Done.

Again, thank you to the reviewer.