Response to RC1

Specific Comments: In the abstract, I suggest including a sentence with the improved calibration numbers between the three methods, moving average (MA), the current operational version (OPER), and the Gaussian Process (GP) with improved dynamic input uncertainty for at least one site (HI01, IL02, or OK02).

We have revised and added the following sentences with the improved calibration numbers in the abstract (lines 23 to 29).

The validation results at the three test sites (i.e. HI02 at Mauna Loa, Hawaii, IL02 at Bondville, Illinois, and OK02 at Billings, Oklahoma) demonstrated that the agreement between aerosol optical depths (AODs) at the 368 nm channel calculated using Vo determined by the GP mean function and the equivalent AERONET AODs were consistently better than those calculated using Vo from standard techniques (e.g. moving average). For example, the average AOD biases by the GP method (0.0036 and 0.0032) are much lower than those by the moving average method (0.0119 and 0.0119) at IL02 and OK02, respectively. The GP method's absolute differences between UV-MFRSR and AERONET AOD values are approximately 4.5%, 21.6%, and 16.0% lower than those of the moving average method at HI02, IL02, and OK02, respectively.

Pg 8, line 193: Just a note to correct the wording of this sentence (though the sentence refers to irradiance at 369-nm). The Physikalish-Meteorological Observatorium Davos, World Calibration Center has a Precision Filter Radiometer (PFR) that measures AOD at 368-nm. Using this type of instrument would avoid additional uncertainties in AOD caused by the interpolation between wavelengths when comparing the MFRSR with the AERONET CIMEL. At the sites used for the comparison in this paper, the site HI02 has a PFR but I do not know about the other two sites. This isn't essential for the analysis, nor conclusions of the paper, only suggest the sentence be modified.

Thanks for pointing out the existence of the WMO reference instruments that measure at 368 nm. We have changed the beginning sentence of section 2.4 to the following (lines 199 to 205).

Ideally, to avoid additional uncertainties caused by the interpolation between wavelengths, the calibration factors should be validated via a direct comparison of direct sun signals from the to-be-calibrated UV-MFRSR and a reference instrument measuring at the 368 nm channel (e.g. the standard precision Filter radiometer (PFR) operated by the Physikalisches-Meteorologisches Observatorium Davos, World Optical Depth Research Calibration Center (WORCC)). However, such reference measurements are not available at most UVMRP stations. Therefore, the estimated mean normalized Vo (Vo_norm) values from the Gaussian Process regression and the other two comparison methods (i.e. MA and OPER) are validated indirectly in terms of aerosol optical depth (AOD) against those obtained from the collocated AERONET sites.

For validation of the technique, the authors compare AOD at 368-nm from the UVMFRSR indirectly to the AERONET CIMEL using information of AOD at two wavelengths (340 and 380 nm). Different types of measurement techniques have their own source of uncertainties as with the CIMEL and the addition of the few paragraphs on previous literature that highlights these differences is crucial to the understanding the improvements using the GP technique.

We agree that the discrepancy/uncertainties in deriving AOD values from the two instruments' measurements should be highlighted explicitly. The following sentences were added at the end of section 2.4 (lines 273 to 288).

Since AERONET and UV-MFRSR AOD values at 368 nm are derived from measurements involving different instruments and wavelengths, the uncertainties when comparing these AOD values should be noted. Some important sources of uncertainties include:

1) AERONET calibration error – At the time of calibration at MLO, AERONET reference instruments have an uncertainty of ~0.2 to 0.5%, which is equivalent to a 0.002 to 0.005 uncertainty in AERONET AOD (Holben et al. 2001). These calibration factors are likely to shift within the year following calibration, which may result in a total AOD uncertainty of ~0.01 to 0.02 (wavelength dependent, higher in the UV) (Holben et al. 2001).
2) Instrument Field of View (FOV) - AERONET CIMELs have a field-of-view (FOV) of 1.2° while the UV-MFRSR has a larger FOV (e.g. ~6.5°, reported by Kazadzis et al. 2018). AODs obtained from instruments with larger FOVs are associated with greater AOD uncertainty due to larger contributions of scattered light to the direct irradiance measurement (Kim et al. 2005).
3) Instrument maintenance – Periodic soiling and cleaning of the UV-MFRSR diffuser can result in spurious increases and decreases in AOD, respectively. The frequency of on-site maintenance (e.g. cleaning of the UV-MFRSR dome) as well as rainfall events may therefore account for some of the AOD difference (Kim et al. 2005; Kim et al. 2008).
4) Trace gases - As mentioned above, AERONET AOD accounts for NO_2 optical depth (e.g. ~0.002 at OK02) while UV-MFRSR AOD does not.

Pg. 7, section 2.2 on Moving Average. This doesn't describe the moving window size used in the analysis.

We have added the following sentence describing the moving window size in section 2.2 (lines 183 to 184).

The parameter *win_size* of MA is set at 20 for all applicable cases in this study.

Technical corrections:

Pg 18, line 395. AEROENT needs to be AERONET.

We have corrected the error accordingly (5 instances).

Pg 420, line 422-423. Incomplete sentence.

The sentence has been revised as below.

As a result, higher accuracy of Rayleigh and other optical depth components is required to discern small improvement on AOD for HI02.

Response to RC2

Concerning the introduction
I am missing previous results of (UV-) MFRSR comparisons with other standard AOD measuring instruments. For example in the 2015 the Filter Radiometer comparison in Davos, Switzerland various of this instrument types have participated and the results have been discussed. There are also earlier studies of such comparisons.
In Kazadzis et al., 2018: There were 4 MFRSR instruments in this campaign. The results were summarized in the following paragraph:
"The four MFR instruments showed good agreement for the medians compared to the PFR triad, however, they exhibit larger scatter than the sun-pointing instruments resulting in a lower precision. McArthur et al. (2003) had previously reported that the MFR-derived AOD does not quite meet the accuracy of the sun-pointing instruments under clean atmospheric conditions. MFR_DE showed an AOD overestimation in various instances that gave results that are outside the WMO defined AOD limits (Fig. 2d). This small overestimation of the MFR_DE instrument compared to the PFR Triad could be due to uncertainties introduced while correcting for their angular response, the calibration procedure, or incomplete blocking of the diffuser by the shadow-band. The MFRSRs that are part of the SURFRAD network (MFR_US2 and MFR_US3) give a median AOD at 500-nm that is in very good agreement with the PFR triad and as good or better than some of the other sun-pointing instruments, e.g., CIMEL and POM; these two slightly underestimate the AOD at 865 nm, but are within the WMO defined limits. Again, these two MFRs' medians are comparable to the better sun-pointing instruments, but give larger scatter. These two MFRs are representative of the SURFRAD network that follow network protocols for calibration and alignment and frequent characterizations of the spectral and angular responses (Augustine et al., 2003, Michalsky et al., 2001)."
Would be helpful some of the above aspects to be included in the introduction section or in section 2.4

We have added the following sentences in section 2.4 (lines 214 to 222):

During the recent Fourth Filter Radiometer Comparison held in Davos, Switzerland (between 28 September and 16 October 2015), most AOD values derived from the three AERONET CIMEL sunphotometers are within the ±0.01 range compared with the PFR triad standard (Kazadzis et al. 2018). This includes those determined at 368nm from the extrapolation of AERONET AODs at 340nm and 380nm. The 2015 Davos campaign also included four MFRSR instruments. Overall, the results showed good agreement between the four MFRSRs and the PFR triad standard, though one instrument exhibited a positive bias and low precision compared to the sun-pointing instruments (Kazadzis et al. 2018). However, such errors were likely explained by instrument-specific uncertainties (e.g. angular response correction, responsivity calibration, and shadowband position issues) and do not suggest inherent error in MFRSR AODs (Kazadzis et al. 2018).

Authors are mentioning: “There are no other instruments measuring at 368nm”
The World Meteorological Organization (WMO) instigated the Global Atmosphere Watch (GAW) program in 1989. Based on a recommendation by GAW experts, the World Optical Depth Research Calibration Center (WORCC) was established in 1996 at the PMOD/WRC in Switzerland. WORCC has since been advised by the GAW Scientific Advisory Group for Aerosols. The standard instrument consist of a precision Filter radiometer (PFR measuring at 368, 412, 500 and 862 nm. So actually the WMO reference instruments (triad) is measuring at 368nm.
So the argument of the non existence of instruments measuring at 368nm (thus the choice of the AOD based comparison) is not correct. However, as it is not possible to repeat this study with one of the PFR instruments, probably the only solution could
be the AOD comparison with the collocated cimels as the authors have initiated. Nevertheless, a short comment on the above text could be included in the paper.

Thanks for pointing out the existence of the WMO reference instruments that measure at 368 nm. We have changed the beginning sentence of section 2.4 to the following (lines 199 to 205).

Ideally, to avoid additional uncertainties caused by the interpolation between wavelengths, the calibration factors should be validated via a direct comparison of direct sun signals from the to-be-calibrated UV-MFRSR and a reference instrument measuring at the 368 nm channel (e.g. the standard precision Filter radiometer (PFR) operated by the Physikalisches-Meteorologisches Observatorium Davos, World Optical Depth Research Calibration Center (WORCC)). However, such reference measurements are not available at most UVMRP stations. Therefore, the estimated mean normalized Vo (Vo_norm) values from the Gaussian Process regression and the other two comparison methods (i.e. MA and OPER) are validated indirectly in terms of aerosol optical depth (AOD) against those obtained from the collocated AERONET sites.

Line 224 : probably you should comment also that ozone is also ignored .

We changed the word “ozone” to “O_3” to make it clear ozone is also ignored.

Line 235: I guess that the cloud flagging method is not evaluated here, as comparison with CIMEL data includes only data that CIMEL algorithm considers as cloud free.

Yes, we did not evaluate the cloud flagging method in this study. We mainly relied on AERONET’s cloud flagging for selecting cloud-free UV-MFRSR measurements. We did perform a simple variation check using similar methodologies to those found in many cloud screening algorithms (e.g. Alexandrov et al. 2004) to reduce the potential contamination of aerosol optical depth by broken clouds.

Line 250 : It would be informative to explain why S(lamda) appears in equation 9.

Since F(lamda) is ~4nm the integrated range is ~366-370 nm. There you mention that AOD is the “interpolated ADO spectrum” which I guess you mean the linear (?) interpolated using 340 and 380 AERONET AODs ? Then S(lamda) is used for normalization in this small 4 nm range? Is this so different that the actual interpolated value of AOD at 368 nm? And if S(lamda) is used, why not S(lamda) – Rayleigh optical depth ? Have in mind that spectral function FWHM of the CIMEL is larger than 2nm.

We agree that the S(lambda) in equation 9 has minimal effects on deriving the UVRSR 368 band passed AERONET AOD. For example, for Mauna Loa, Hawaii site, the mean (2.6×10^{-6}) and standard deviation (2.5×10^{-6}) of the difference between the AOD with and without S(lambda) are several magnitudes smaller than the instrument resolution. Therefore, we removed the S(lambda) term in equation 9 for simplicity (lines 266 to 271). The change is so small that it doesn’t impact further analysis in the manuscript.

Comparing the AOD directly interpolated to 368 nm with the bandpass AOD (calculated with updated equation 9) for Mauna Loa, Hawaii site, we found that there was a ~0.1% discrepancy. The small discrepancy could be explained by the small difference between the effective wavelength of the instrument (367.91 nm for the MLO UV-MFRSR) and 368 nm. With larger wavelength differences for some instruments, we expect a larger discrepancy. Therefore, we decided to keep using the more accurate bandpass AOD in the manuscript.

The authors chose to evaluate their method by comparing the retrieved (from their Vos data) AODs. Here are my comments on this section:

- The comparison of UVMFR with AERONET would be essential to follow criteria that are defined by WMO –CIMO in order to assess the results in detail.

https://library.wmo.int/pmb_ged/wmo-td_1287.pdf

There (page 8) such conditions and formulas are defined.

For example the U95 criterion where a number (here the lower limit is 95%) of measurements have to be in the range of ±0.005 + 0.010/m

Where 0.005 accounts for instrument related uncertainties and 0.01/m for calibration related uncertainties (calibration uncertainty better than 1%).

Changing the analysis figures with the use of this criterion authors can:

a. better show the agreement and the improvements with their methods by showing the percentage of data within these limits for each case.

b. having in mind that calibration related uncertainties will be inherited in AOD retrievals as a function of air mass, the figures (a) including aod differences vs air mass can point out on Vo related issues.

Still slopes and cor. Coefficients can be reported in the form of a table.

Thank you for pointing out the U95 criterion. We understand that complying with such criterion is critical for achieving traceability of AOD products generated from radiometer measurements.

In this study, the comparison of AOD values derived from UV-MFRSR measurements and AERONET AOD values only serves as an indirect evidence that the calibration of UV-MFRSR is reasonably accurate. However, we do not attempt to argue that the calibration of UVMRP UV-MFRSR is accurate enough to produce AOD values that meet the U95 criterion. Nor do we attempt to argue that UVMRP AOD products are traceable to the AERONET or WMO AOD standard. As mentioned in the manuscript and the citations therein, the stability assumption of the Langley method may not be strictly fulfilled at many UVMRP sites, rendering larger uncertainty of Langley calibration factors at these sites compared with those derived under ideal conditions (such as at Mauna Loa, Hawaii).

This conclusion is supported by the following analysis. Based on the GP uncertainty results (using 1.96 instead of 4.42 standard deviations to mimic 95% CI, data not shown in the manuscript) among the three test sites, only at HI02 (Mauna Loa, Hawaii), the AOD uncertainty (at 95% level) caused by calibration (~0.0078) is lower than the U95 criterion (0.01). The AOD uncertainties (at 95% level) caused by calibration at the other two sites are much larger than 0.01 (i.e. ~0.043 for IL02, and ~0.028 for OK02).

We recognize that the assumption that AERONET AOD represents “ground-truth” is not ideal, however, as indicated by previous work, it is a reasonable assumption for this study. AERONET sunphotometers are routinely calibrated with the uncertainty of AOD around 0.002 to 0.005 in the visible and up to 0.01 in the UV region (Eck et al., 1999; Holben et al., 2001). Additionally, during the recent Fourth Filter

Radiometer Comparison held in Davos, Switzerland (between 28 September and 16 October 2015), most AOD values derived from the three AERONET CIMEL sunphotometers are within the ±0.01 range compared with the PFR triad (Kazadzis et al. 2018). This includes those determined at 368nm from the extrapolation of AERONET AODs at 340nm and 380nm. Therefore, we are confident that the AERONET AOD products used in this study are accurate enough to be a reliable source of AOD values for validation purposes. This assumption is consistent with numerous other field-based evaluations of radiometric AOD accuracy as detailed between lines 213 and 214 in the manuscript. To emphasize the validity of using AERONET as an effective standard, we have revised the following text in the manuscript summarizing the results from the Fourth Filter Radiometer Comparison and Kazadzis et al. (2018) (lines 199 to 222).

Ideally, to avoid additional uncertainties caused by the interpolation between wavelengths, the calibration factors should be validated via a direct comparison of direct sun signals from the to-be-calibrated UV-MFRSR and a reference instrument measuring at the 368 nm channel (e.g. the standard precision Filter radiometer (PFR) operated by the Physikalisches-Meteorologisches Observatorium Davos, World Optical Depth Research Calibration Center (WORCC)). However, such reference measurements are not available at most UVMRP stations. Therefore, the estimated mean normalized Vo (Vo_norm) values from the Gaussian Process regression and the other two comparison methods (i.e. MA and OPER) are validated indirectly in terms of aerosol optical depth (AOD) against those obtained at the collocated AERONET sites. We admit that the uncertainty of UV-MFRSR AODs could exceed the World Meteorological Organization (WMO) U95 criterion (e.g. 95% of the measured data have uncertainty in the range of 0.005 ± 0.01 / airmass, Kazadzis et al. 2018) at many UVMRP sites because the stability assumption of the Langley method may not be strictly fulfilled. Therefore, the AOD comparison in this study can only serve as an indirect evidence to verify whether the calibration of UV-MFRSR is reasonably accurate.

AERONET sunphotometers are routinely calibrated with the uncertainty of AOD around 0.002 to 0.005 in the visible and up to 0.01 in the UV region (Eck et al., 1999; Holben et al., 2001) and are therefore considered a reliable source for AOD intercomparison and radiometer validation [e.g. (Alexandrov et al., 2002, 2008;Augustine et al., 2003;Krotkov et al., 2005a;Krotkov et al., 2005b;Kassianov et al., 2007;Tang et al., 2013;Yin et al., 2015;Zhang et al., 2016)]. During the recent Fourth Filter Radiometer Comparison held in Davos, Switzerland (between 28 September and 16 October 2015), most AOD values derived from the three AERONET CIMEL sunphotometers are within the ±0.01 range compared with the PFR triad standard (Kazadzis et al. 2018). This includes those determined at 368nm from the extrapolation of AERONET AODs at 340nm and 380nm. The 2015 Davos campaign also included four MFRSR instruments. Overall, the results showed good agreement between the four MFRSRs and the PFR triad standard, though one instrument exhibited a positive bias and low precision compared to the sun-pointing instruments (Kazadzis et al. 2018). However, such errors were likely explained by instrument-specific uncertainties (e.g. angular response correction, responsivity calibration, and shadowband position issues) and do not suggest inherent error in MFRSR AODs (Kazadzis et al. 2018).

The AOD retrieval and the differences among two instruments are a consequence not only an uncertainty on the instrument calibrations but also other factors.
Here is a list:
- The calculation of Rayleigh optical depth from both instruments including the pressure measurement. Are the two instruments (UVMFR and CIMEL) use the same formulas ?
- The calculation of Rayleigh and aerosol air mass factors
- The potential differences in the field of view of the instrument
- CIMEL includes NO2 and Ozone optical depths
- The wavelength interpolation from 340 and 380 nm to 368 nm is not by definition linear but aerosol type related.

So in order to assess their results the authors at least have to mention the related uncertainties and the above issues raised by using retrieved AODs from two different instruments with different instrument characteristics and post processing AOD algorithms and procedures, in order to validate the Vos.

In theory a direct comparison of direct sun signals for the UVMFR instrument and a reference instrument measuring at 368nm could be used in order to assess the differences in the Vo calculation, without having the AOD calculation related uncertainties.

We agree that the discrepancy/uncertainties in deriving AOD values from the two instruments' measurements should be highlighted explicitly. The following sentences were added at the end of section 2.4 (lines 273 to 288).

Since AERONET and UV-MFRSR AOD values at 368 nm are derived from measurements involving different instruments and wavelengths, the uncertainties when comparing these AOD values should be noted. Some important sources of uncertainties include:

1) AERONET calibration error – At the time of calibration at MLO, AERONET reference instruments have an uncertainty of ~0.2 to 0.5%, which is equivalent to a 0.002 to 0.005 uncertainty in AERONET AOD (Holben et al. 2001). These calibration factors are likely to shift within the year following calibration, which may result in a total AOD uncertainty of ~0.01 to 0.02 (wavelength dependent, higher in the UV) (Holben et al. 2001).
2) Instrument Field of View (FOV) - AERONET CIMELs have a field-of-view (FOV) of 1.2° while the UV-MFRSR has a larger FOV (e.g. ~6.5°, reported by Kazadzis et al. 2018). AODs obtained from instruments with larger FOVs are associated with greater AOD uncertainty due to larger contributions of scattered light to the direct irradiance measurement (Kim et al. 2005).
3) Instrument maintenance – Periodic soiling and cleaning of the UV-MFRSR diffuser can result in spurious increases and decreases in AOD, respectively. The frequency of on-site maintenance (e.g. cleaning of the UV-MFRSR dome) as well as rainfall events may therefore account for some of the AOD difference (Kim et al. 2005; Kim et al. 2008).
4) Trace gases - As mentioned above, AERONET AOD accounts for NO_2 optical depth (e.g. ~0.002 at OK02) while UV-MFRSR AOD does not.

We do not include the Rayleigh optical depth formula, airmass formula, and interpolation methods in the list.

For the Rayleigh optical depth (RLOD) calculation, both UV-MFRSR and AERONET (version 2, https://aeronet.gsfc.nasa.gov/version2_table.pdf) use the same formula described in Bodhaine et al. 1999. The instantaneous pressure values for UV-MFRSR are obtained from the collocated AERONET measurements. Therefore, RLOD should not introduce additional uncertainty in this study.

For the airmass calculation, both UV-MFRSR and AERONET (version 2, https://aeronet.gsfc.nasa.gov/version2_table.pdf) use the same formula described in Kasten and Young 1989.

For the interpolation between 340 and 380 nm AOD values, we agree that the spectra of aerosol optical depth is aerosol type related and may not be strictly linear (e.g. slightly quadratic). However, we believe that the difference in the interpolated AOD spectrum among different interpolation methods/equations should be negligible because the two wavelengths are so close (e.g. Figure 6 in Krotkov et al. 2005a).

A list of all relevant changes made in the manuscript

1. In the abstract, we added a few sentences with the improved calibration numbers.
2. In section 2.2 Moving Average (MA), we added one sentence describing the win_size of MA used in this study.
3. In section 2.4 Validation method for 368-nm in-situ calibration factors,
    a. We correct our previous statement regarding the existence of other instruments measuring at 368 nm (i.e. the PFR operated by WORCC).
    b. We discussed the WMO U95 criterion and explained why UVMRP UV-MFRSR AODs could exceed the criterion.
    c. We added evidence from the recent Fourth Filter Radiometer Comparison (Davos, Switzerland) in 2005 to show the good agreement among AERONET CIMEL, MFRSR, and the PRF triad standard AOD measurements.
    d. We simplified the equation 9 for the equivalent AERONET AOD at the 368 nm channel.
    e. We added a paragraph discussing the potential uncertainties when comparing AERONET and UV-MFRSR AOD values at 368 nm.

Improving the Mean and Uncertainty of Ultraviolet Multi-Filter Rotating Shadowband Radiometer In-Situ Calibration Factors: Utilizing Gaussian Process Regression with a New Method to Estimate Dynamic Input Uncertainty

Maosi Chen [1], Zhibin Sun [1], John M. Davis [1], Yan-An Liu [2,3,4], Chelsea A. Corr [1], and Wei Gao [1,5]

[1]United States Department of Agriculture UV-B Monitoring and Research Program, Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, USA
[2]Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China
[3]School of Geographic Sciences, East China Normal University, Shanghai 200241, China
[4]ECNU-CSU Joint Research Institute for New Energy and the Environment, Shanghai 200062, China
[5]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO 80523, USA

Correspondence to: Maosi Chen (maosi.chen@colostate.edu), Zhibin Sun (zhibin.sun@colostate.edu), Yan-An Liu (yaliu@geo.ecnu.edu.cn)

Abstract. To recover the actual responsivity for Ultraviolet Multi-Filter Rotating Shadowband Radiometer (UV-MFRSR), the complex (e.g. unstable, noisy, and with gaps) time series of its in-situ calibration factors (Vo) need to be smoothed. Many smoothing techniques require accurate input uncertainty of the time series. A new method is proposed to estimate the dynamic input uncertainty by examining overall variation and subgroup means within a moving time window. Using this calculated dynamic input uncertainty within Gaussian Process regression (GP) provides the mean and uncertainty functions of the time series. This proposed GP solution was first applied on a synthetic signal and showed significant smaller RMSEs than a Gaussian Process regression performed with constant values of input uncertainty and the mean function. GP was then applied to three UV-MFRSR Vo time series at three ground sites; The method appropriately accounted for variation in slopes, noises, and gaps at all sites. The validation results at the three test sites (i.e. HI02 at Mauna Loa, Hawaii, IL02 at Bondville, Illinois, and OK02 at Billings, Oklahoma) demonstrated that the agreement between aerosol optical depths (AODs) at the 368 nm channel calculated using Vo determined by the GP mean function and the equivalent AERONET AODs were consistently better than those calculated using Vo from standard techniques (e.g. moving average). For example, the average AOD biases by the GP method (0.0036 and 0.0032) are much lower than those by the moving average method (0.0119 and 0.0119) at IL02 and OK02, respectively. The GP method's absolute differences between UV-MFRSR and AERONET AOD values are approximately 4.5%, 21.6%, and 16.0% lower than those of the moving average method at HI02, IL02, and OK02, respectively.~~The validation results demonstrated that the agreement between aerosol optical depths (AODs) calculated using Vo determined by the GP mean function and AERONET AODs were consistently better than those calculated using Vo from standard techniques (e.g. moving average).~~ The improved accuracy of in-situ UVMRP Vo values suggests the GP solution is a robust technique for accurate analysis of complex time series and may be applicable to other fields.

1 Introduction

While many instruments generate relatively stable data time series over short time windows, dynamic uncertainty levels, variable sampling densities, and/or different lengths of gaps with missing data can complicate the analysis of long-term datasets. For example, the five-year time series of a solar variability indicator (Mg II core to wing index) shows consistency on the order of days but increasing noise level and gaps are observed at the month-scale (Cebula et al., 1992). The time series of the geopotential scale factor, a function of the geoidal potential, is also relatively stable on shorter time scales but demonstrates a slowly increasing long-term pattern (Burša et al., 1997). Additionally, the time series of a ratio (F factor) for calibrating a satellite radiometer suite (i.e. VIIRS) shows band-specific gap distributions and variable trends (Cardema et al., 2012). As a result, these time series may not be described as a simple deterministic function of time due to possible noise and gaps.

Long term measurements of irradiance by Multi-Filter Rotating Shadowband Radiometers (MFRSRs) are also subject to errors imposed by the factors mentioned above. The MFRSR measures direct normal, diffuse horizontal, and total horizontal irradiances at seven visible channels with a roughly 10 nm full half-maximum width (FHMW) (Harrison and Michalsky, 1994). The Ultraviolet (UV) version of MFRSR measures the same three irradiance components at seven UV channels (i.e. 300, 305, 311, 317, 325, 332, and 368 nm) with a 2 nm FHMW (Gao et al., 2010). Currently, the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility (Mather and Voyles, 2013), the NOAA Surface Radiation (SURFRAD) (Augustine et al., 2005) and the U.S. Department of Agriculture (USDA) UV-B Monitoring and Research Program (UVMRP) (Gao et al., 2010) maintain their own MFRSR and/or UV-MFRSR at multiple sites across the U.S. To capture immediate instrument responsivity variation, the UVMRP performs in-situ calibrations using the Langley method (Slusser et al., 2000;Harrison and Michalsky, 1994) or derived approaches [e.g. (Chen et al., 2013;Chen et al., 2016;Chen et al., 2015)] on (UV-)MFRSR direct beam measurements on days with extended clear-sky periods (Gao et al., 2010).

Many factors contribute to the error or uncertainty of the Langley method including variations in aerosol and/or other atmospheric constituents over the course of the calibration period (Augustine et al., 2003;Chen et al., 2015;Zhang et al., 2016), the presence of thin cirrus (Shaw, 1976), as well as instrument errors (e.g. instrument tilt and misalignment, incorrect night-time offset and angular corrections) (Alexandrov et al., 2007). Thus, the sequence of original UVMRP (UV-)MFRSR in-situ calibration factors exhibits certain levels of noise. Among these uncertainties, variable AOD is considered the major contributor to the variability of the Langley calibration factors obtained in typical atmospheric conditions over the continental United States (Alexandrov et al., 2008), even with careful cloud screening [e.g. (Chen et al., 2014;Alexandrov et al., 2004)]. In addition, extended cloudy periods and low solar zenith angles during winter months further reduce the sequence quality, which appear as large time gaps in the datasets. Since the in-situ calibration factor represents the instrument's responsivity which is assumed to be relatively stable, it has been suggested that one applies some smoothing methods (e.g. averaging or fitting a smooth curve) on the daily calibration time series (Alexandrov et al., 2008) to reduce the issue. Currently, UVMRP implements an outlier detection and moving smoothing technique to overcome these issues. However, the process involves manual interaction, performs unreliably during sparse and gapped periods, and lacks the uncertainty estimation.

Analyses of complex long-term time series, such as those of (UV-)MFRSR Vo values, must consider: (i) the underlying continuous trend (i.e. the mean function) and the corresponding trend uncertainty and (ii) the (dynamic) input uncertainty. For problem (i), there is a variety of available approaches, such as local polynomial regression, smoothing splines, and Gaussian Process regression (Proietti, 2011). Local polynomial regression (LPR) constructs a polynomial within each local time window, and fits its coefficients by locally weighted least squares. LPR's computational complexity is low, and it can eliminate some of the randomness in the data (Hyndman, 2011). However, LPR may have difficulty on the cases with varying sampling densities or gaps. In addition, LPR does not allow estimating the trend near the ends of the time series and cannot be used for forecasting (Hyndman, 2011). A spline is a piecewise polynomial function with continuous derivatives (Proietti, 2011), and smoothing splines estimate the underlying spline by minimizing the distance between the spline and the observations while penalizing the roughness of the spline (Wahba, 2011). For example, a cubic spline fit was used to fill the large gaps in the Mg II index time series (Viereck et al., 2004). Both LPR and smoothing splines are unable to utilize the information about the input uncertainties or to estimate the uncertainty associated with the trend. Unlike the two methods above, Gaussian Process does not restrict the class of the underlying functions because it is not a parametric model (Rasmussen and Williams, 2006). Instead, it gives a priori probability to every possible function based on the desired function characteristics such as smoothness (Rasmussen and Williams, 2006). Gaussian Process regression assumes both the observations and the underlying function are from one joint (prior) Gaussian distribution, and derives the underlying function distribution by conditioning the joint (prior) distribution on the observations (Rasmussen and Williams, 2006). The method takes the observational error into consideration and naturally gives the uncertainty of the underlying function, making itself an appropriate tool for problem (i). Gaussian Process regression has been widely used in many fields [e.g. forecasting of mortality rates (Wu and Wang, 2018), prediction of spatial-temporal violent events (Kupilik and Witmer, 2018), and modelling received signal strength for wireless local area network location fingerprinting (Richter and Toledano-Ayala, 2015)].

For problem (ii), the input error statistics (e.g. input uncertainty) is often assumed to be known or roughly estimated in advance; In practice, a typical approach may use some predetermined constant (e.g. the nominal uncertainty of an instrument, or the standard deviation of its observation) to estimate input uncertainty for the entire dataset. However, this kind of approach omits the information of the possible time-varying observation error, leading to over- or under- estimation of the input uncertainty at a given (temporal) location (Chandorkar et al., 2017). A sophisticated approach may treat the dynamic input uncertainty as additional parameters and solve them together with other model parameters through optimization under the Bayesian framework (Kavetski et al., 2006b, a). However, this method requires the specification of valid error/uncertainty models, which are normally poorly understood in practice (Kavetski et al., 2006b, a).

In this study, we developed and validated a generic solution that combines Gaussian Process regression with a new dynamic input uncertainty estimation method, to determine the underlying continuous trend and the corresponding uncertainty for the given time series. In section 2, we briefly summarize the basics of the Gaussian Process regression and develop the dynamic input uncertainty estimation method. We also describe a complex (noisy, gapped, etc.) synthetic time series and real UV-MFRSR in-situ calibration factor time series used in the analysis. In section 3, we present and discuss the performance of the Gaussian Process method on the test data, in comparison with the UVMRP current operational method and a moving average technique. Validation of the calibration factors determined with the Gaussian Process method via the comparison of AODs calculated with these factors and those reported by the AErosol RObotic NETwork (AERONET)(Holben et al., 1998) is also discussed in section 3.

2 Materials and Methods

2.1 Gaussian Process regression (GP)

2.1.1 Main Procedure

A Gaussian Process is a technique used in the analysis of a finite number of random variables with a joint Gaussian distribution (Rasmussen and Williams, 2006). The following introduces briefly the theory of GP regression. An observed dataset, $\mathcal{D}_{obs} = (\mathbf{X}, \boldsymbol{y}) = \{(\boldsymbol{x}_i, y_i) \mid i = 1, \ldots, N\}$, has N pairs of inputs ($\mathbf{X} = \{\boldsymbol{x}_i\} \in \mathbf{R}^{N \times D}$) and corresponding observed values ($\boldsymbol{y} = \{y_i\} \in \mathbf{R}^N$), where D is the length of input vector $\boldsymbol{x}_i$. $\boldsymbol{y}$ is the combination of a function of $\mathbf{X}$ and noises $\boldsymbol{\varepsilon}$: $\boldsymbol{y}=\mathbf{f}(\mathbf{X})+\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ follows an independent distributed Gaussian distribution $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\boldsymbol{0}, diag(\boldsymbol{\sigma}_{\mathrm{y}}^2)\right)$ and $\boldsymbol{\sigma}_{\mathrm{y}} \in \mathbf{R}^N$ is the given or estimated uncertainty (standard deviation) on the N observations. In practice, $\boldsymbol{\sigma}_{\mathrm{y}}$ is not always known in advance. The section below, "Proposed Dynamic Input Uncertainty Estimation", provides an empirical approach to estimating $\boldsymbol{\sigma}_{\mathrm{y}}$. It is assumed that the test dataset $\mathcal{D}_* = (\mathbf{X}_*, \boldsymbol{f}_*) = \{(\boldsymbol{x}_{*i}, f_{*i}) \mid i = 1, \ldots, N_*\}$ and the observed dataset ($\mathcal{D}_{obs}$) have the joint Gaussian distribution but the test function values ($\boldsymbol{f}_*$) are unknown:

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_{\mathbf{XX}} + \boldsymbol{\sigma}_{\mathrm{y}}^2\mathbf{I} & \mathbf{K}_{\mathbf{XX}_*} \\ \mathbf{K}_{\mathbf{X}_*\mathbf{X}} & \mathbf{K}_{\mathbf{X}_*\mathbf{X}_*} \end{bmatrix}\right), \tag{1}$$

where, **I** is the identity matrix, $\mathbf{K}_{\mathbf{X}_*\mathbf{X}} \in \mathbf{R}^{N_* \times N}$ denotes the covariance matrix between observed ($\mathbf{X}_*$) and test inputs (**X**), and similarly for the other three terms $\mathbf{K}_{\mathbf{XX}} \in \mathbf{R}^{N \times N}$, $\mathbf{K}_{\mathbf{XX}_*} \in \mathbf{R}^{N \times N_*}$, and $\mathbf{K}_{\mathbf{X}_*\mathbf{X}} \in \mathbf{R}^{N_* \times N}$. Each element of these covariance matrices is determined by a kernel function $K(\boldsymbol{z}_1,\boldsymbol{z}_2)$, which maps any pair of inputs ($\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbf{R}^D$) into **R**. There are a wide variety of kernel functions such as the radial basis function (RBF) and the rational quadratic (RQ) kernel (Rasmussen and Williams, 2006). For example, The RQ kernel is defined by the following equation with length scale (l) and alpha (α) as its two parameters (Rasmussen and Williams, 2006):

$$k_{RQ}(r) = \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}, r = \left\|\boldsymbol{z}_1 - \boldsymbol{z}_2\right\|. \tag{2}$$

In practice, users need to use prior knowledge or techniques such as autocorrelation to choose the best kernel function to represent the correlation among input data. The hyperparameters ($\boldsymbol{\theta}$) of the chosen kernel function are then optimized by maximizing the log transformed marginal likelihood (Rasmussen and Williams, 2006):

$$\log p\left(\boldsymbol{y} \mid \mathbf{X}, \boldsymbol{\theta}\right) = -\frac{1}{2}\boldsymbol{y}^T\left[\mathbf{K}_{\mathbf{XX}}\left(\boldsymbol{\theta}\right) + \boldsymbol{\sigma}_{\mathrm{y}}^2\mathbf{I}\right]^{-1}\boldsymbol{y} - \frac{1}{2}\log\left|\mathbf{K}_{\mathbf{XX}}\left(\boldsymbol{\theta}\right) + \boldsymbol{\sigma}_{\mathrm{y}}^2\mathbf{I}\right| - \frac{N}{2}\log 2\pi \,. \tag{3}$$

To simplify the calculation, the mean of $\boldsymbol{y}$ has been subtracted from both the actual observed values and the test function values. Therefore, the joint distribution has a mean equal to zero.

Based on the (optimized) joint distribution [Eq. (1)], the theorem that derives the conditional distribution from the joint Gaussian distribution (Eaton, 1983), and the inversion equations of a partitioned matrix (Press, 1992), the Gaussian Process regression predicts $\boldsymbol{f}_*$ from given **X**, $\boldsymbol{y}$, and $\mathbf{X}_*$ (Rasmussen and Williams, 2006):

$$\left\{\boldsymbol{f}_* \mid \mathbf{X}, \boldsymbol{y}, \mathbf{X}_*\right\} \sim \mathcal{N}\left(\overline{\boldsymbol{f}}_*, \operatorname{cov}(\boldsymbol{f}_*)\right), \tag{4}$$

where,

$$\overline{\boldsymbol{f}}_* = \mathbf{K}_{\mathbf{X}_*\mathbf{X}}\left[\mathbf{K}_{\mathbf{XX}} + \boldsymbol{\sigma}_{\mathrm{y}}^2\mathbf{I}\right]^{-1}\boldsymbol{y}\,, \tag{5}$$

$$\operatorname{cov}(\boldsymbol{f}_*) = \mathbf{K}_{\mathbf{X}_*\mathbf{X}_*} - \mathbf{K}_{\mathbf{X}_*\mathbf{X}}\left[\mathbf{K}_{\mathbf{XX}} + \boldsymbol{\sigma}_{\mathrm{y}}^2\mathbf{I}\right]^{-1}\mathbf{K}_{\mathbf{XX}_*}\,. \tag{6}$$

The GP predicted sample standard deviations [i.e. the square root of the diagonal elements in $\operatorname{cov}(\boldsymbol{f}_*)$] can be converted to the predicted confidence intervals. For example, the predicted 0.99999 confidence intervals used in this study are obtained by multiplying a constant (i.e. 4.42) with predicted sample standard deviation. Points outside the predicted confidence intervals may be considered as outliers and can be excluded iteratively until all points are within the confidence intervals or the average ratio between GP predicted means and standard deviations are less than a threshold (e.g. the threshold is 0.01 in this study).

2.1.2 Proposed Dynamic Input Uncertainty Estimation

As mentioned before, the statistical properties of the noise $\boldsymbol{\varepsilon}$ of the observed time series $\boldsymbol{y}$ might be unknown. Even if assuming $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\boldsymbol{0}, diag(\boldsymbol{\sigma}_{\mathrm{y}}^2)\right)$ in practice, $\boldsymbol{\sigma}_{\mathrm{y}}$ is not always a constant and could vary in time. Therefore, we propose to estimate $\boldsymbol{\sigma}_{\mathrm{y}}$ with a moving window approach. Within each moving window (W), the input uncertainty (denoted as s_i) is assumed to be relatively stable and can be estimated using all points in the window (W). Note that s_i is not equivalent to the standard deviation of all points within the period (s_W), unless the mean function of the time series is invariant. We derive the relationship between s_i and s_W (see Appendix A for the detailed derivation) to estimate s_i:

$$s_W^2 = \frac{N-J}{N-1} s_i^2 + \frac{1}{N-1} \sum_{j=1}^{J} N_j \left(\mu_j - \mu_W\right)^2 , \tag{7}$$

where, all points within W are clustered into J subgroups based on their similarity in both time and value; N_j is the number of points in each subgroup j; $N = \sum_{j=1}^{J} N_j$ is the number of all points within W; μ_j is the mean of subgroup j, which can vary among subgroups; s_i is the estimated uncertainty of each point within W, acting as the sample standard deviation across all subgroups; μ_W and s_W are the mean and sample standard deviation of all points within W. The classic K-Means algorithm was used for the clustering process. To increase the reliability to estimate statistics (mean or sample standard deviation), small subgroups are merged with adjacent ones to ensure each subgroup has more than required minimum points. The numbers of initial subgroups and the required minimum points depend on the prior knowledge on the variability and availability of the data. Sensitivity studies (not shown) indicate that 5 initial subgroups per moving window and 3 required minimum points per subgroup worked well for our applications. The dynamic input uncertainty estimation process is applied on every data point in a sequence. The squares of the estimated input standard deviations [i.e. s_i^2 in Eq. (7)] are stored on the respective diagnostic positions in $\boldsymbol{\sigma}_{\mathrm{y}}$.

The flowchart of the proposed Dynamic Input Uncertainty Estimation method and the complete GP procedure of estimating the mean and confidence interval functions of a given time series is presented in Figure 1.

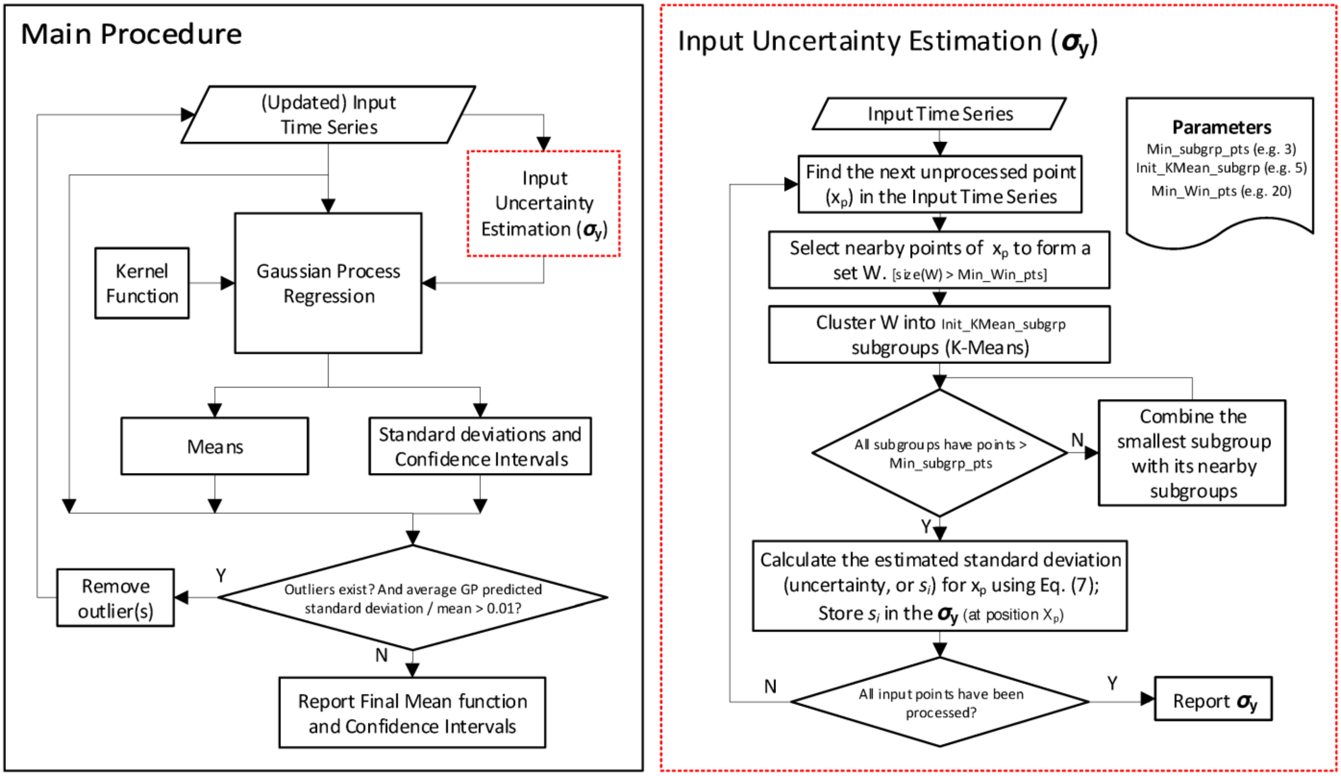

Figure 1. Main procedure for deriving the mean and confidence interval functions using Gaussian Process Regression (left black box) and detailed procedure of the proposed Input Uncertainty Estimation method (right red box).

2.2 Moving Average (MA)

Moving Average (MA) is a simple smoothing technique. To assess the performance of the GP regression with other methods, this study implements MA for one-dimensional case as follows. For a given x_{*_i}, we first choose its nearby observations $\left\{(x, y) \| x - x_{*_i} | \leq win\_size, (x, y) \in \mathcal{D}_{obs}\right\}$ within the given window *win_size* and then calculate the mean y value of the subset as the smoothed observation at x_{*_i}. The process is repeated for all possible x in $\mathcal{D}_*$. The parameter *win_size* of MA is set at 20 for all applicable cases in this study.

2.3 UVMRP operational algorithm (OPER)

UVMRP operational algorithm (OPER) was specially designed for smoothing its in-situ calibration factor sequences (http://uvb.nrel.colostate.edu/UVB/dataProcessingInfo/VnaughtsDataProcessing.jsf). OPER is included as an additional source for methods comparison. The algorithm has three steps. In the first step, a 12-count running mean and the corresponding standard deviation are maintained to detect outliers (i.e. points outside half of the running mean or two standard deviation). During the process, if three consecutive points are determined to be outliers, visual examination is performed to determine if a permanent change in the instrument responsivity has occurred. If such a change is confirmed, calculation of a new running mean begins on the three points. In the second step, a moving linear regression is used to smooth the values at the center of each moving window. The moving window size is ±3 months. If visual examination finds significant value changes on a date of interest (the center of a moving window), the regression is not performed on that date. In the final step, the regression results from step two are used as input into a weighted means algorithm to generate continuous smooth in-situ calibration factors. The inverse of year fraction between the current date of interest and the date of each participating point is used to calculate the weights. The weighting window is also ±3 months from the date of interest.

2.4 Validation method for 368-nm in-situ calibration factors

~~Since there are no other ground radiometers that measure the irradiance at the exact 368 nm channel as the UVMRP UV-MFRSR does, the estimated mean normalized Vo (Vo_norm) values from the Gaussian Process regression and the other two comparison methods (i.e. MA and OPER) are validated indirectly in terms of aerosol optical depth (AOD) against those obtained at the collocated AERONET sites. AERONET sunphotometers are routinely calibrated with the uncertainty of AOD around 0.002 to 0.005 in the visible and up to 0.01 in the UV region (Eck et al., 1999;Holben et al., 2001) and are therefore considered a reliable source for AOD intercomparison and radiometer validation [e.g. (Alexandrov et al., 2002, 2008;Augustine et al., 2003;Krotkov et al., 2005a;Krotkov et al., 2005b;Kassianov et al., 2007;Tang et al., 2013;Yin et al., 2015;Zhang et al., 2016)].~~ Ideally, to avoid additional uncertainties caused by the interpolation between wavelengths, the calibration factors should be validated via a direct comparison of direct sun signals from the to-be-calibrated UV-MFRSR and a reference instrument measuring at the 368 nm channel [e.g. the standard precision Filter radiometer (PFR) operated by the Physikalisches-Meteorologisches Observatorium Davos, World Optical Depth Research Calibration Center (WORCC)]. However, such reference measurements are not available at most UVMRP stations. Therefore, the estimated mean normalized Vo (Vo_norm) values from the Gaussian Process regression and the other two comparison methods (i.e. MA and OPER) are validated indirectly in terms of aerosol optical depth (AOD) against those obtained at the collocated AERONET sites. We admit that the uncertainty of UV-MFRSR AODs could exceed the World Meteorological Organization (WMO) U95 criterion [e.g. 95% of the measured data have uncertainty in the range of 0.005 ± 0.01 / airmass, (Kazadzis et al., 2018)] at many UVMRP sites because the stability assumption of the Langley method may not be strictly fulfilled. Therefore, the AOD comparison in this study can only serves as an indirect evidence to verify whether the calibration of UV-MFRSR is reasonably accurate.

AERONET sunphotometers are routinely calibrated with the uncertainty of AOD around 0.002 to 0.005 in the visible and up to 0.01 in the UV region (Eck et al., 1999;Holben et al., 2001) and are therefore considered a reliable source for AOD intercomparison and radiometer validation [e.g. (Alexandrov et al., 2002, 2008;Augustine et al., 2003;Krotkov et al., 2005a;Krotkov et al., 2005b;Kassianov et al., 2007;Tang et al., 2013;Yin et al., 2015;Zhang et al., 2016)]. During the recent Fourth Filter Radiometer Comparison held in Davos, Switzerland (between 28 September and 16 October 2015), most AOD values derived from the three AERONET CIMEL sunphotometers are within the ±0.01 range compared with the PFR triad standard (Kazadzis et al., 2018). This includes those determined at 368nm from the extrapolation of AERONET AODs at 340nm and 380nm. The 2015 Davos campaign also included four MFRSR instruments. Overall, the results showed good agreement between the four MFRSRs and the PFR triad standard, though one instrument exhibited a positive bias and low precision compared to the sun-pointing instruments (Kazadzis et al., 2018). However, such errors were likely explained by instrument-specific uncertainties (e.g. angular response correction, responsivity calibration, and shadowband position issues) and do not suggest inherent error in MFRSR AODs (Kazadzis et al., 2018).

Previous comparison between (UV-) MFRSRs and AERONET AODs have generally showed good agreement and demonstrate the utility of AERONET as an effective standard for field-based measurements.~~Augustine et al. (2003) compared SURFRAD MFRSR AODs at the Table Mountain station in Colorado with UVMRP MFRSR AODs at the Pawnee station (85 km northeast of Table Mountain) and with National Renewable Energy Laboratory (NREL) sun-photometer-derived AODs at Golden station (50 km to the south). The AOD difference on the test cases showed a magnitude of 0.1 to 0.2 and was variable over time even for the same comparison site.~~ For instance, Krotkov et al. (2005a);Krotkov et al. (2005b) validated the UVMRP UV-MFRSR AODs with the interpolated AERONET AODs at 368 nm at the National Aeronautics and Space Administration/Goddard Space Flight Center (NASA/GSFC) site in Greenbelt, Maryland. They found that the UV-MFRSR AODs at 368-nm channel on cloud-free days had a daily RMSE less than 0.01 when calibrated using AERONET measurements and increased to approximately 0.02-0.05 (depending on the season) when calibrated using standard Langley method (Harrison and Michalsky, 1994;Slusser et al., 2000). Alexandrov et al. (2002) developed an comprehensive calibration method for the VIS-MFRSR and validated the calibration at the four channels (i.e. 440, 500, 670, and 870 nm) by comparing the derived AOD values with interpolated AERONET values at the ARM Cloud and Radiation Testbed (CART) site. The results showed small AOD difference (i.e. <0.005) at 440, 500, and 870 nm channels for a variety of atmospheric conditions with AODs ranging from 0.03 to 0.4 (at 500-nm). Alexandrov et al. (2008) considered optical depth of NO_2 and Ozone during the MFRSR AOD calculation, although they were small enough to be ignored (i.e. 0.008 NO_2 optical depth at 415 nm and 0.005 ozone optical depth at 615 nm) at their test location at the ARM Southern Great Plains (SGP) site. The long-term intercomparison showed a good agreement (i.e. difference between them <0.01) between the MFRSR and AERONET AODs at 440, 675, and 870 nm channels. Kassianov et al. (2007) validated the MFRSR-retrieved optical properties and reported small RMSE values (i.e. 0.0043-0.0075) among MFRSR, AERONET, and Normal incidence multifilter radiometer (NIMFR) derived AODs at 500 and 870 nm channels during the ARM Program's Aerosol Intensive Operational Period (IOP) in 2003.

In this study, for the UV-MFRSR at 368 nm channel, aerosol optical depth ($AOD_{368nm,UVMRP}$) is calculated by subtracting Rayleigh optical depth ($RLOD_{368nm,UVMRP}$) from total optical depth ($TOD_{368nm,UVMRP}$) under cloud-free conditions. The absorption of ~~ozone~~O_3, NO_2, and other trace gases are very small at the 368 nm channel (e.g. NO_2 optical depth is around 0.002 to 0.003 at AERONET Cart_Site), so they are ignored during the calculation of $AOD_{368nm,UVMRP}$:

$$AOD_{368nm,UVMRP} \approx TOD_{368nm,UVMRP} - RLOD_{368nm,UVMRP}\ . \quad (8)$$

TOD is calculated using Beer's Law (e.g. (Slusser et al., 2000)), where the actual calibration factor at top of atmosphere (Vo_raw) is restored from GP estimated mean Vo_norm. The cosine corrected voltage and airmass are obtained from the UVMRP webpage (https://uvb.nrel.colostate.edu/UVB/da_queryCosCorrected.jsf). *RLOD* is calculated by following the equations in Bodhaine et al. (1999). The site latitude and height for *RLOD* calculation are from the UVMRP webpage (https://uvb.nrel.colostate.edu/UVB/uvb-siteinfo.jsf), and the instantaneous site-level surface pressure for *RLOD* calculation is obtained from the collocated AERONET sites (https://aeronet.gsfc.nasa.gov/cgi-bin/webtool_opera_v2_new).

To obtain reliable AOD values, UV-MFRSR measurements with quality concerns or cloud contamination are excluded in the following comparison. More specifically, (1) any measurements with UVMRP-provided quality control flag(s) relevant to the data quality of the direct beam at 368 nm channel are excluded; (2) data with small (direct beam) measurements at 368 nm are also excluded because they are more sensitive to noise or errors introduced during various calibration steps; and (3) a simple variation check is performed to reduce the potential of mixing cloud and aerosol optical depth. If the ratio between the standard deviation of TODs and the mean TOD value in the 15-minute time window exceeds 0.05, they are excluded from further analyses.

AERONET (v2.0) provides AOD at 340 and 380 nm channels. These values are interpolated to the effective wavelength of the UV-MFRSR 368 nm channel for comparison using the Ångström exponent as follows. Note that in the log transformed coordinate system [i.e. log(AOD) vs. log(wavelength)], log(AOD) is generally linear between 340 and 380 nm (Krotkov et al., 2005a). First, the AERONET AOD spectrum between the two wavelengths is derived by linear interpolation of AERONET AODs at 340 and 380 nm in the log transformed coordinate system. Next, since the UV-MFRSR AOD at 368 nm is a bandpass value over a narrow band (i.e 2 nm FHMW), the equivalent AERONET AOD at that channel is derived by

$$AOD_{368nm,AERONET} = \frac{\int_{340nm}^{380nm} AOD_{\lambda} F_{\lambda} d\lambda}{\int_{340nm}^{380nm} F_{\lambda} d\lambda}\ , \quad (9)$$

where AOD_λ is the interpolated AERONET AOD spectrum; F_λ is the spectral response function of the UV-MFRSR at 368 nm channel (http://uvb.nrel.colostate.edu/UVB/da_queryFilterFunctions.jsf);~~ S_λ is the solar irradiance spectrum at top of atmosphere (Chance and Kurucz, 2010)~~ and the wavelength interval for the integral is 0.05 nm. Note that negative AERONET AOD measurements are excluded from the validation because of using log transform.

Since AERONET and UV-MFRSR AOD values at 368 nm are derived from measurements involving different instruments and wavelengths, the uncertainties when comparing these AOD values should be noted. Some important sources of uncertainties include:

1) AERONET calibration error – At the time of calibration at MLO, AERONET reference instruments have an uncertainty of ~0.2 to 0.5%, which is equivalent to a 0.002 to 0.005 uncertainty in AERONET AOD (Holben et al.,

2001). These calibration factors are likely to shift within the year following calibration, which may result in a total AOD uncertainty of ~0.01 to 0.02 (wavelength dependent, higher in the UV) (Holben et al., 2001).

2) Instrument Field of View (FOV) – AERONET CIMELs have a FOV of 1.2° while the UV-MFRSR has a larger FOV [e.g. ~6.5°, reported by (Kazadzis et al., 2018)]. AODs obtained from instruments with larger FOVs are associated with greater AOD uncertainty due to larger contributions of scattered light to the direct irradiance measurement (Kim et al., 2005).

3) Instrument maintenance – Periodic soiling and cleaning of the UV-MFRSR diffuser can result in spurious increases and decreases in AOD, respectively. The frequency of on-site maintenance (e.g. cleaning of the UV-MFRSR dome) as well as rainfall events may therefore account for some of the AOD difference (Kim et al., 2005;Kim et al., 2008).

4) Trace gases – As mentioned above, AERONET AOD accounts for NO_2 optical depth (e.g. ~0.002-0.003 at OK02) while UV-MFRSR AOD does not.

2.5 Datasets

2.5.1 Synthetic Case

We generate a synthetic time series that is composed of six segments with a varying base function and noise levels [Figure 2 (a)]. The base function [Eq. (10)], including linear, quadratic, and cubic functions, simulates a wide variety of functions for which the proposed technique is applicable. The noise levels are the same within each segment but different across segments. The noise at segment i is sampled from a fixed normal distribution $\mathcal{N}\left(0,\sigma_i^2\right)$, where σ_i is equal to 4, 8, 6, 15, 7, and 3 from left to right segments, respectively. Each segment originally contains 200 points. Their x coordinates are sampled randomly from six uniform distributions within their domains. Points with x coordinates in the three designated windows (i.e. [64.2, 69.2], [80.8, 85.8], and [122.5, 127.5]) are removed to simulate data gaps in reality.

$$y=\begin{cases} 1.5x-30, & 0\le x<50 \\ -1.2(x-50)+45, & 50\le x<100 \\ -0.02(x-100)^2+2.3(x-100)-15, & 100\le x<150 \\ -0.02(x-150)^2-0.5(x-150)+50, & 150\le x<200 \\ 0.0004(x-200)^3+0.012(x-200)^2+0.4(x-200)-25, & 200\le x<250 \\ 0.002(x-250)^3-0.1(x-250)^2-2.5(x-250)+75, & 250\le x<300 \end{cases} \quad (10)$$

2.5.2 Application Cases: In-situ calibration factors

In this study, the in-situ calibration factors of UVMRP UV-MFRSRs are used as application cases to test the performance of the three smoothing methods (i.e. GP, MA, and OPER). These UV-MFRSR in-situ calibration factors over several months or years are obtained through the Langley method on clear days. Their varying uncertainties are mainly attributed to two aspects.

One is the optical stability of atmospheric constituents (e.g., the aerosol, ozone, and thin clouds) when the in-situ calibration factor is derived (Chen et al., 2015), and the other is the aging status of the radiometer. UVMRP publish its in-situ calibration factors on their website (http://uvb.nrel.colostate.edu/UVB/da_queryVoIntercepts.jsf). To reduce the chances of abrupt changes in the sequences, the data associated with the same instrument (i.e. UV-MFRSR) at the same UVMRP site (denoted as a deployment period) are processed together. Three UVMRP sites with collocated ~~AEROENT~~ AERONET sites (for validation) were selected (Table 1). The in-situ calibration factors at these UVMRP sites represent time series with contrasting densities, noisiness, and slopes (Table 1). Appendix B uses the Oklahoma site (OK02) to show that the UV-MFRSR 368-nm in-situ calibration factors obey normal distribution.

Table 1. The three UVMRP 368-nm UV-MFRSR in-situ calibration factor time series for test.

UVMRP Site Name	UVMRP Site Location	Collocated AERONET Site	Deployment Start and End Dates	Figure (original time series)	Time Series Characteristics
HI02	19.54° N, 155.58° W, 3409 m	Mauna_Loa	17 September 2015 to 1 July 2018	Figure 3(a1)	dense, low noise, variable slope
IL02	40.05° N, 88.37° W, 213 m	BONDVILLE	21 March 2017 to 29 May 2018	Figure 3(b1)	sparse, high noise, sharper slope
OK02	36.60° N, 97.49° W, 317 m	Cart_Site	17 January 2007 to 11 June, 2011	Figure 3(c1)	medium density, medium noise, variable slope

3 Results and Discussion

3.1 Synthetic Case

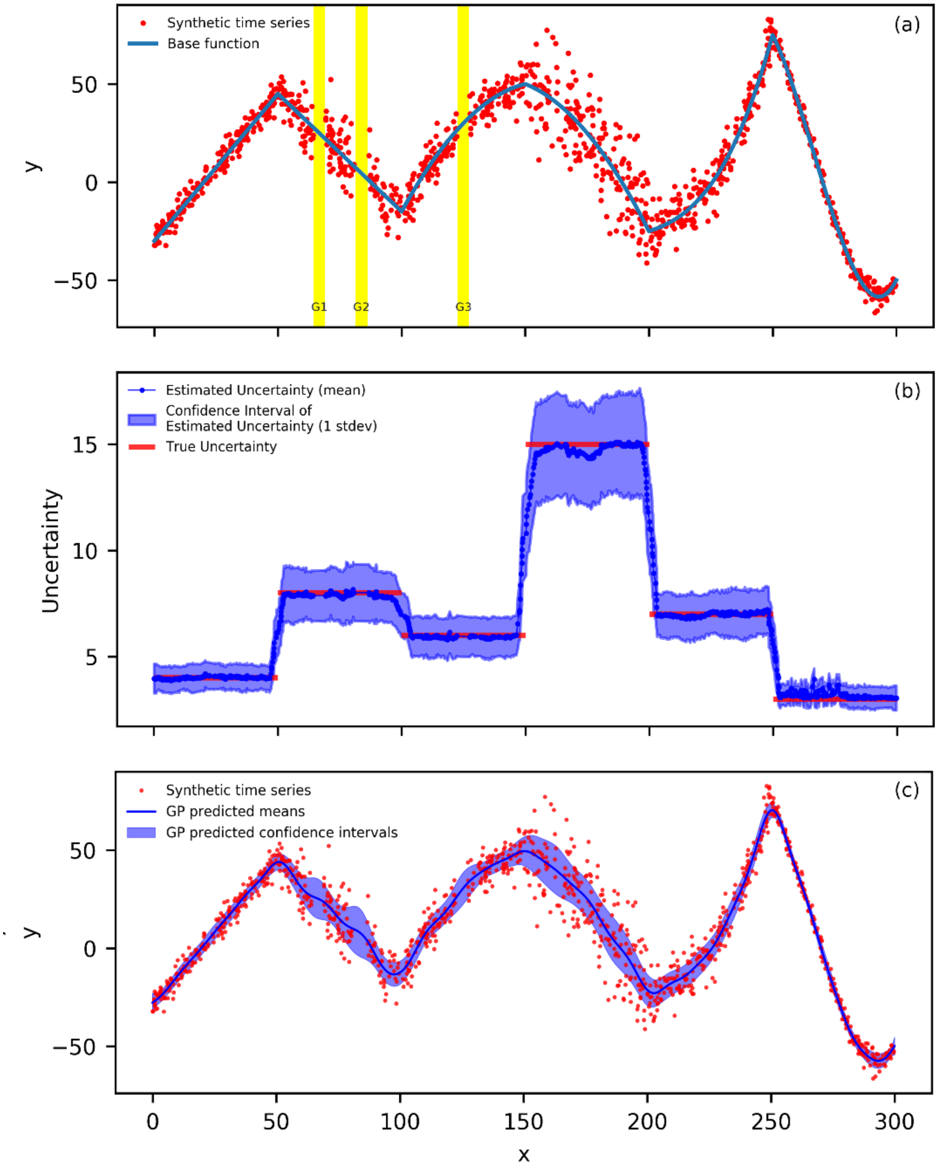

Figure 2. (a) The synthetic time series based on Eq. (10) (the blue line) with varying noise levels. There are originally 200 samples within every 50-wide interval (or segment) in the x coordinate, but points between [64.2, 69.2] (highlighted in yellow, G1), [80.8, 85.8] (highlighted in yellow, G2), and [122.5, 127.5] (highlighted in yellow, G3) are removed to simulate data gaps in practice. The final number of points in the sequence is 1140. (b) The means (dark blue circles) and confidence intervals (light blue area) of the estimated uncertainty for the 200 synthetic sequences [all sampled from the distribution of (a) but with different random noise]. The true uncertainty (red line segments) is also displayed. (c) The Gaussian Process regression results on the synthetic time series from (a). The dark blue line is the predicted mean function and the light blue area is the corresponding confidence intervals.

3.1.1 Estimation of Input Uncertainty for Gaussian Process

The proposed "Dynamic Input Uncertainty Estimation" method is first applied to the synthetic case. To observe the statistical properties/characteristics of the estimated input uncertainty, this procedure was applied on 200 synthetic time series, each of which is generated by adding random noise into the base function [Eq. (10)] following the procedures discussed in section "2.5.1 Synthetic Case".

Figure 2(b) shows the means (dark blue circles) and confidence intervals (light blue area) of estimated uncertainty of the 200 estimated input uncertainty sequences. The mean of the estimated uncertainty is close to the true uncertainty (RMSE = 0.6321) for the entire synthetic case as demonstrated by a linear regression between estimated and true uncertainty with a slope close to one (i.e. 1.0332) and a high R^2 of 0.9759 (Table 2). Most true uncertainty (red line segments) is covered by the confidence intervals except for the areas near the ends of the six segments. In these areas, the method averaged the uncertainty from the adjacent segments and presented a smooth transition between segments. This small RMSE value suggests that using smaller subgroup size (e.g., 3~6 points) does not significantly influence the estimation of uncertainty [Figure 2(a)]. Therefore, smaller subgroups are preferred over larger ones as larger subgroups are more likely to have gap(s) with large variation, which tends to increase its estimated standard deviation [Eq. (7)].

To demonstrate the improvements in the GP resulting from the dynamic input uncertainty estimation, the GP is also run with three typical constant input uncertainties: overall standard deviation of the synthetic time series (30.95), minimum true uncertainty of the synthetic time series (2.00), and maximum true uncertainty of the synthetic time series (15.00). The results from all three constant input uncertainties are less accurate than the estimated input uncertainty generated by the proposed method (Table 2). The proposed method has significant smaller RMSE (i.e. 0.6321) compared with the three constant input uncertainties (i.e. 24.1152, 6.5226, and 8.7921, respectively). Similarly, the linear regression between the estimated and true uncertainties shows that the proposed method has the slope and the R^2 values both close to one (i.e. 1.0332 and 0.9759) while the three constant uncertainties shows no (linear) correlation with true uncertainties (i.e. the slope and R^2 values close to zero).

3.1.2 Estimation of Means and Confidence Interval and Its Validation

The kernel function in the Gaussian Process regression used in this study is the rational quadratic (RQ) kernel, with two parameters: length scale and alpha [Eq. (2)]. To use RQ with Gaussian Process regression, we need to provide the initial (estimated) values for these two parameters. First, we round the original data points [red points in Figure 2(a)] to the nearest 0.25 interval grids. Then, we calculate the autocorrelation on these rounded data points from lags of 0.25 to 22.25 (approximately equivalent to lags of 1 to 90 points). Next, we perform curve fitting on autocorrelation results and obtain 9.80 and 1.05 as initial length scale and alpha estimates, respectively. With these initial RQ parameters and the estimated inputs uncertainty (from the proposed method or using three representative constant input uncertainties), Gaussian Process regression predicts the mean and uncertainty functions. Figure 2(c) shows the GP results for the proposed method: dark blue line for the mean function and the light blue area for the confidence intervals (4.42 times of the GP predicted uncertainty function).

In terms of the GP predicted mean function vs. the base function [Eq. (10)], the proposed input uncertainty estimation method shows a 12.0% to 15.7% improvement on RMSE over the three constant input uncertainties (i.e. 1.1785 vs. 1.3146, 1.3976, and 1.3146) (Table 2). Similarly, the slope of the linear regression between the two functions is closer to one for the proposed uncertainty estimation method (i.e. 1.0082) than the three constant uncertainties (i.e. 1.0228). In addition, the predicted mean function from the proposed method is close to the base function even near the gaps [G1, G2, and G3 in Figure 2(a)] [Figure 2(c)].Additionally, the proposed method's predicted uncertainty function (or confidence intervals) shows better agreement with the true uncertainty of the synthetic time series [Figure 2(c)] while the three constant input uncertainties' results show consistent over- or under-estimated pattern over the entire time series (figures not shown). It is noted that the predicted confidence intervals from the proposed method are wider near the three gaps [G1, G2, and G3 in Figure 2(a)] than nearby locations with similar uncertainty. This is anticipated because the constraint in the gaps are from distant points where the RQ kernel gives low correlation.

Table 2 Validation of the input uncertainty and mean of GP prediction using four input uncertainties: the input uncertainty estimated by the proposed method (Section 2.1.2), overall standard deviation of the synthetic time series (30.95), minimum true uncertainty of the synthetic time series (2.00), and maximum true uncertainty of the synthetic time series (15.00). RMSE stands for root mean square error. LR stands for linear regression. R^2 stands for the coefficient of determination for linear regression. Note: † y_1 represents the true input uncertainty of the synthetic time series, x_1 represents the estimated input uncertainties. ‡ y_2 represents the true values on the base function [Eq.(10)], x_2 represents the GP estimated mean values using the respective input uncertainty.

			Constant input uncertainty		
	Metrics	Proposed input uncertainty estimation method	Overall standard deviation (30.95)	Minimum synthetic time series uncertainty (2.00)	Maximum synthetic time series uncertainty (15.00)
input uncertainty	RMSE	0.6321	24.1152	6.5226	8.7921
	LR†	$y_1=1.0332x_1-0.2277$	$y_1=-0.1962x_1-0.2277$	$y_1=0.0x_1+7.1632$	$y_1= 0.0x_1+7.1632$
	R^2	0.9759	0.0	0.0	0.0
mean of GP prediction	RMSE	1.1785	1.3146	1.3976	1.3146
	LR‡	$y_2=1.0082x_2-0.3865$	$y_2=1.0228x_2-0.5351$	$y_2=1.0228x_2-0.5636$	$y_2=1.0228x_2-0.5351$
	R^2	0.9986	0.9986	0.9983	0.9986

3.2 In-Situ Calibration Factors Cases

3.2.1 Applications

The same GP procedure is applied on three in-situ calibration factor (Vo_norm, sun-earth distance normalized) sequences from three UVMRP deployment periods (Figure 3) at three different UVMRP locations previously described in Table 1.The Hawaii site (HI02) sits at a clean, high altitude location, which means its atmospheric condition is more stable than other UVMRP sites and its Vo_norm has the lowest variation [Figure 3(a1)]. The Illinois site (IL02) is surrounded by croplands/rangelands with the closest city (Champaign) located 12 km northwest [Figure 3(b1)]. The Oklahoma site (OK02) is also surrounded by croplands/rangelands with the closest city (Oklahoma City) located about 96 km south [Figure 3(c1)]. Both wildfires and agricultural activities (e.g. cultivation and harvest) at IL02 and OK02 contribute to the relatively hazy and unstable atmosphere condition for Langley regression. As ~~the~~ a result, Vo_norms at IL02 and OK02 have larger variation compared with HI02. The dynamic input uncertainty estimation results confirm that the uncertainty at HI02 [15–40, Figure 3(a2)] is also lower than the other two sites [100–300, Figure 3(b2), (c2)]. Generally, the proposed method gives lower uncertainty values for time windows with more clustered points (e.g. December 2008 and April 2010 at OK02 [Figure 3(c2)], and February 2017 at HI02 [Figure 3(a2)]). There are no obvious temporal patterns of uncertainty at any of the three sites.

Figure 3(a3), (b3), and (c3) show the estimated means (dark blue line) and confidence intervals (light blue area) after the initial pass through GP. The length scale parameter of the RQ kernel for the HI02, IL02, and OK02 sites are 6.091, 6.369, and 6.228 (days), respectively. Their corresponding alpha parameters of the RQ kernel function are all close to 1.0 (i.e. 0.948, 0.862, and 0.944, respectively). As expected, the confidence interval is narrower near time windows with more data points, and the confidence intervals are wider near gaps [Figure 3(b3)].

As depicted in Figure 1, the outlier removal and GP are repeated following the initial GP regression, giving the final GP results shown in Figure 3(a4), (b4), and (c4). After this final pass, the length scale parameter of the RQ kernel function for the HI02, IL02, and OK02 sites are 6.091, 11.149, and 6.907 (days), respectively. Compared with the first round, all length scale parameters increase as more outliers are removed (except for HI02). At HI02, the average ratio between GP means and standard deviations is lower than the threshold (i.e. 0.01) after the first round and the iteration stops. The corresponding alpha parameters of the RQ kernel function are still all close to 1.0 (i.e. 0.948, 1.010, and 1.110, respectively). Because of outlier removal, compared with the first-round results, GP generates smoother mean functions and narrower confidence intervals at the last round.

The other two methods (i.e. MA and OPER) are applied on the same in-situ calibration time series. They can provide mean functions but not confidence intervals. The MA (win_size=20) results [Figure 3(a5), (b5), and (c5)] are generally smoother than OPER [Figure 3 (a6), (b6), and (c6)] but both are more responsive to noisy points than GP. In addition, since OPER is scheduled to run once per month on active deployments, there may be some lags at the end of those deployments [e.g. Figure 3(a6)].

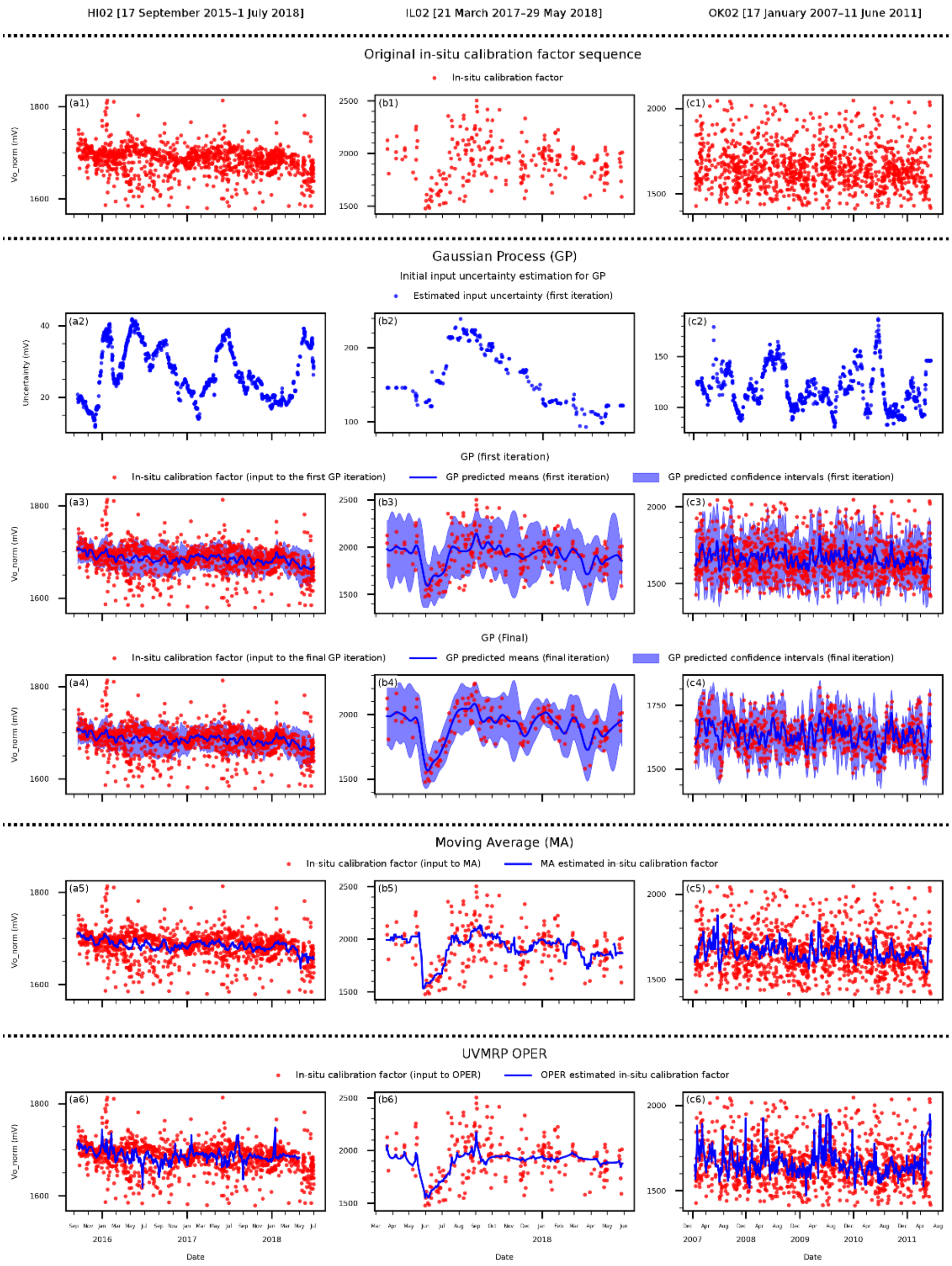

Figure 3. The results of the three smoothing methods (i.e. GP – Gaussian Process, MA – Moving Average, OPER – UVMRP operational algorithm) on the three UVMRP in-situ calibration factor sequences: (a) HI02 (17 September 2015 to 1 July 2018), (b) IL02 (21 March 2017 to 29 May 2018), and (c) OK02 (17 January 2007 to 11 June, 2011). The first row (a1, b1, c1) displays the original in-situ calibration factor (Vo_norm) sequence. The second row (a2, b2, c2) shows the initial input uncertainty estimated for GP. The third row (a3, b3, c3) presents the predicted daily mean and confidence interval from the first iteration of GP. The fourth row (a4, b4, c4) shows the final results of GP after iterations. The fifth row (a5, b5, c5) shows the results of MA. The sixth row (a6, b6, c6) shows the results of OPER. .

3.2.2 Validation

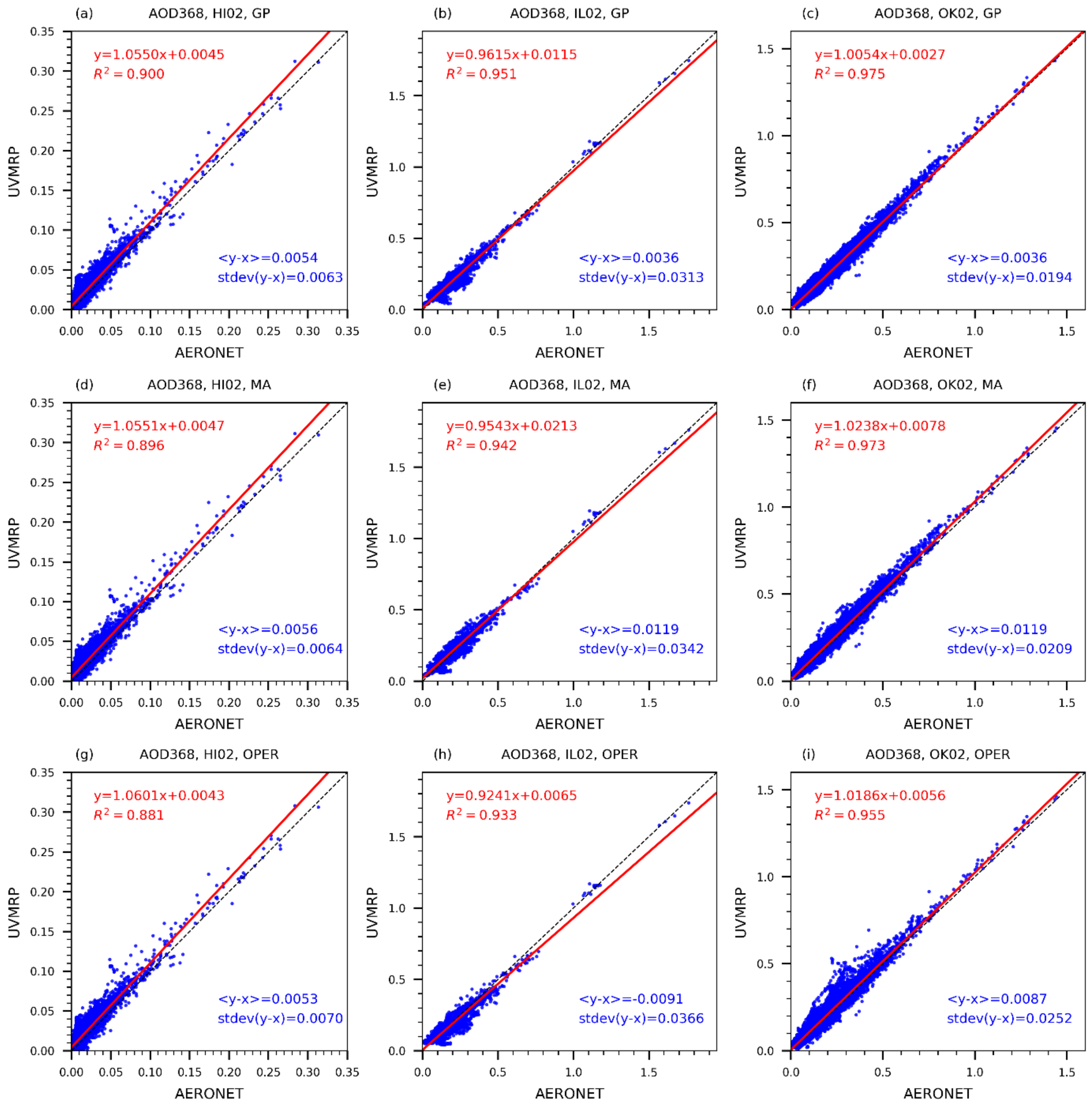

Figure 4. 368-nm AOD scatter plots between UVMRP (*y* axis) and AERONET (*x* axis). The UVMRP 368-nm AODs are calculated from UV-MFRSR direct normal voltages using calibration factors estimated by the three methods (i.e. from top to bottom: GP, MA, OPER) at the three sites (i.e. from left to right: HI02, IL02, OK02). The AERONET 368-nm AODs are derived from collocated (i.e. Mauna_Loa, BONDVILLE, Cart_Site) AERONET AODs on the 340- and 380-nm channels. The linear regression line (solid, red) and the 1-by-1 line (dashed, black) are also plotted. "<*y*-*x*>" means the average difference between AERONET and UVMRP AOD at 368 nm channel. "stdev(*y*-*x*)" means the standard deviation of their difference.

Following the procedures described in section 2.4, the UVMRP AODs at 368 nm channel generated by GP, MA, and OPER are validated against the corresponding AERONET AODs at the three collocated sites (i.e. HI02 – Mauna_Loa, IL02 – BONDVILLE, OK02 – Cart_Site). The scatter plots between these UVMRP and ~~AEROENT~~ AERONET AODs are displayed in Figure 4. The performance of all three methods at HI02 [Figure 4(a), (d), (g)] are similar. For example, the average bias "<y-x>" is approximately 0.0054 and standard deviation of the difference "stdev(y-x)" is approximately 0.0066. For IL02 [Figure 4(b), (e), (h)] and OK02 [Figure 4(c), (f), (i)], GP shows superior agreements with AERONET than the other two methods. For example, at IL02, the absolute value of GP's average bias (0.0036) is about 3.3 to 2.5 times lower than that of MA (0.0119) and OPER (0.0091). Similarly, at OK02, the average bias for GP (0.0032) is much lower than those for MA (0.0119) and OPER (0.0087). The validation results for GP at OK02 are similar to the previous comparison results between AERONET and MFRSR AODs at 415 and 440 nm (Tang et al., 2013;Alexandrov et al., 2008).

Table 3 Statistical metrics (average absolute difference, average absolute relative difference, and linear regression) on comparing 368-nm AOD between UVMRP ($AOD_{368,UVMRP}$) and ~~AEROENT~~AERONET ($AOD_{368,AE}$). The UVMRP 368-nm AODs at the three sites (i.e. HI02, IL02, and OK02) are calculated using calibration factors estimated by the three methods (i.e. GP, MA, and OPER). The AERONET 368-nm AODs are derived from collocated (i.e. Mauna_Loa, BONDVILLE, and Cart_Site) AERONET AODs on the 340- and 380-nm channels. LR stands for linear regression. R^2 stands for the coefficient of determination for linear regression. The "x" and "y" in "Linear regression equation" refer to $AOD_{368,AE}$ and $AOD_{368,UVMRP}$ of the respective methods.

Site	Metrics	Method		
		GP	MA	OPER
HI02	$Avg(\lvert AOD_{368,UVMRP}-AOD_{368,AE}\rvert)$	0.0062	0.0065	0.0067
	$Avg\left(\frac{\lvert AOD_{368,UVMRP}-AOD_{368,AE}\rvert}{AOD_{368,AE}}\right)$	0.5803	0.6078	0.6261
	LR	y=1.0550x+0.0045	y=1.0551x+0.0047	y=1.0601x+0.0043
	R^2	0.9000	0.8957	0.8812
IL02	$Avg(\lvert AOD_{368,UVMRP}-AOD_{368,AE}\rvert)$	0.0228	0.0291	0.0270
	$Avg\left(\frac{\lvert AOD_{368,UVMRP}-AOD_{368,AE}\rvert}{AOD_{368,AE}}\right)$	0.1669	0.2087	0.1930
	LR	y=0.9615x+0.0115	y=0.9543x+0.0213	y=0.9241x+0.0065
	R^2	0.9514	0.9420	0.9332
OK02	$Avg(\lvert AOD_{368,UVMRP}-AOD_{368,AE}\rvert)$	0.0150	0.01785	0.01847
	$Avg\left(\frac{\lvert AOD_{368,UVMRP}-AOD_{368,AE}\rvert}{AOD_{368,AE}}\right)$	0.1714	0.2067	0.1939
	LR	y=1.0054x+0.0027	y=1.0238x+0.0078	y=1.0186x+0.0056
	R^2	0.9749	0.9726	0.9554

Table 3 shows two additional statistical metrics for validation: “Avg(|$AOD_{368,UVMRP}$-$AOD_{368,AE}$|)”, a measure of absolute difference between the two quantities and “Avg(|$AOD_{368,UVMRP}$-$AOD_{368,AE}$|/$AOD_{368,AE}$)” a measure of relative difference between the two quantities. For HI02, the GP Vo_norm values improves both the absolute ~~(~4.5%)~~ and relative ~~(~7.5%)~~ differences between $AOD_{368,UVMRP}$ and $AOD_{368,AE}$ when compared to MA (by ~4.5%) and OPER AODs (by ~7.5%), respectively. Results from linear regressions (LR) performed between $AOD_{368,UVMRP}$ and $AOD_{368,AE}$ are also reported in Table 3. The LR results are similar between GP and MA, but GP has closer-to-one LR slope (1.0550) and higher R^2 (0.9000) than those of OPER (1.0601 and 0.8812) for HI02. For IL02, GP shows 21.6% smaller absolute difference and 20.0% smaller relative difference to AERONET than MA; GP shows 15.6% smaller absolute difference and 13.5% smaller relative difference to AERONET than OPER. Similarly, for OK02, GP shows 16.0% smaller absolute difference and 17.1% smaller relative difference to AERONET than MA; GP shows 18.8% smaller absolute difference and 11.6% smaller relative difference to AERONET than OPER.

Overall, the ~~368-nm~~368 nm AODs by GP shows higher correlation, closer-to-one slopes, and lower absolute and relative biases compared to AERONET AODs than MA and OPER at all three sites. The improvement of GP over MA and OPER at IL02 and OK02 are more significant than at HI02. The main reason may be that HI02 is the least polluted site among the three sites. Both of its maximum and mean 368-nm AOD values are low: 0.35 and 0.016, respectively. As a result, higher accuracy of Rayleigh and other optical depth components is required to discern small improvement on AOD for HI02. Since the AERONET’s sun photometer is routinely calibrated, the agreement on AOD values suggests that the calibration factors mean function generated by GP are more accurate than MA and OPER.

~~In addition,~~ Figure 5 shows the 368-nm AOD time series calculated using GP generated in-situ calibration factors at the three UVMRP sites. The blue solid line represents the AODs calculated using the GP means, and the green and red dotted lines represent the AODs calculated using the GP confidence intervals. We first note that, unlike the obvious seasonal changes in AOD difference reported in the previous study at the NASA/GSFC site by Krotkov et al. (2005a), this study (Figure 5) shows no discernible seasonal pattern in the AOD differences at all three sites. It is seen that the AOD confidence intervals are approximately ±0.0095, ±0.0480, and ±0.0273 at HI02, IL02, and OK02, respectively. The corresponding AERONET AOD time series are also plotted (i.e. purple lines in Figure 5). The insets in Figure 5 show comparison details at HI02, IL02, and OK02. For most of the AOD time series, AERONET results are within the GP confidence intervals. The average absolute differences of daily AOD values between GP and AERONET are ~0.006 for HI02, ~0.024 for IL02, and ~0.014 for OK02. These values are also close to or within the AERONET AOD uncertainty level (i.e. 0.01), demonstrating the potential for the development of a reasonably accurate UVMRP AOD product. ~~suggesting the high-quality of the potential UVMRP AOD product. In addition, unlike the obvious seasonal changes in AOD difference reported in the previous study at the NASA/GSFC site by Krotkov et al. (2005a), this study (Figure 5) shows no discernible seasonal pattern in the AOD differences at all three sites.~~

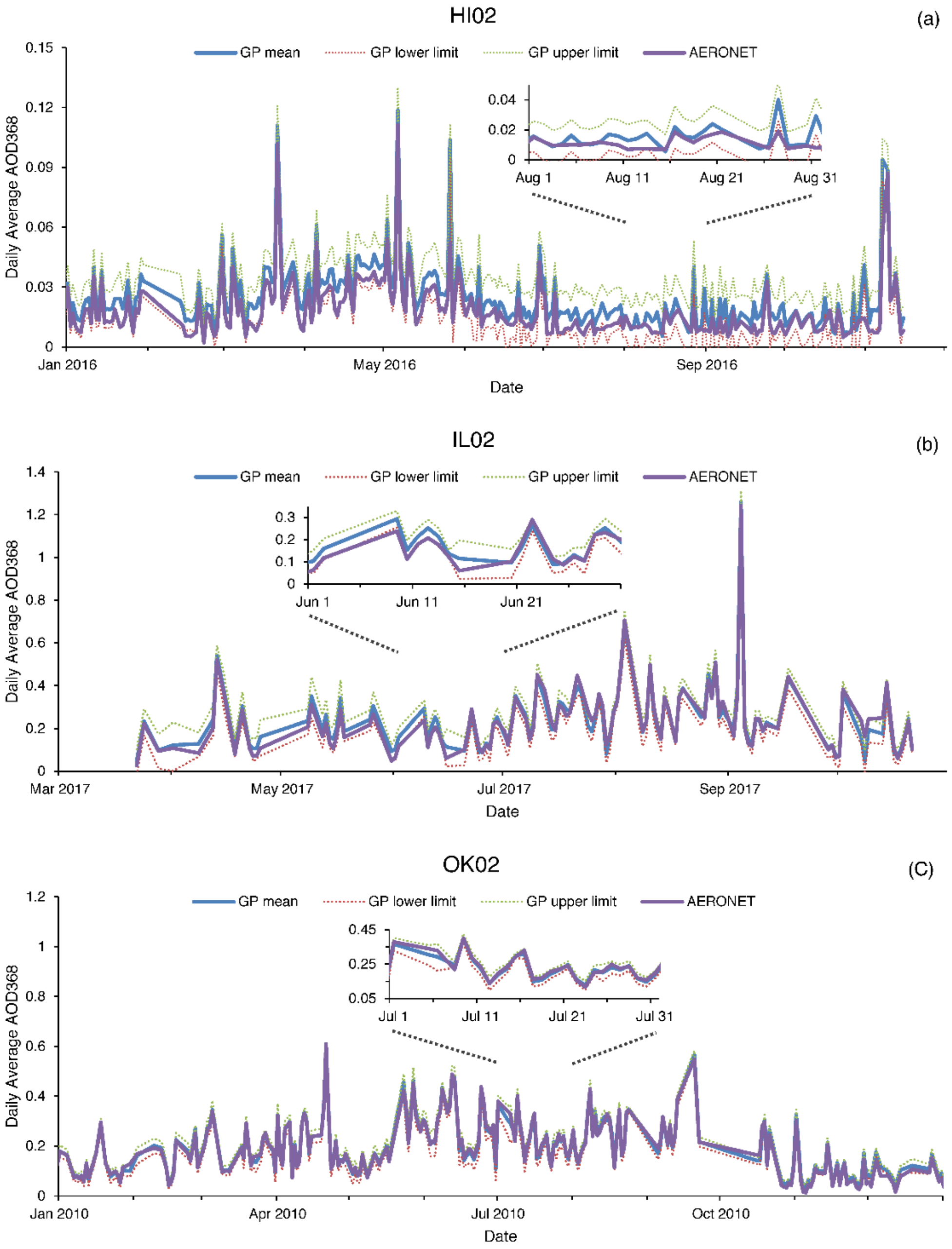

Figure 5. Time series of UVMRP and AERONET 368-nm daily average AOD at HI02, IL02, and OK02 sites. The daily AOD mean values derived from the GP mean in-situ calibration factor (V_o) functions (blue) and the corresponding AERONET values (purple) are shown in blue solid lines. The corresponding lower and upper limits of AOD derived from the GP V_o confidence intervals are shown in dotted red and green lines, respectively. The insets for HI02 (August 2016), IL02 (June 2017), and OK02 (July 2010) are also included in the respective subplots to show the comparison details.

4 Conclusions

A new dynamic uncertainty estimation method for noisy time series is developed in this study. Combining this method with Gaussian Process regression, we provide a solution to estimate the underlying mean and uncertainty functions of time series with variable mean, noise, sampling density, and length of gaps. For the synthetic case with linear, quadratic, and cubic base functions, noise level varying from 2 to 15, and noticeable gaps, the proposed solution returns a mean function with the RMSE of 1.1785 (linear regression R^2 of 0.9986), which is at least 12.0% lower than RMSEs associated with the three constant input uncertainties. ~~Its~~ The estimated input uncertainties determined by this method are close to the true uncertainty levels except for the transitional region between segments. The solution also gives accurate mean values at the three gaps. The proposed GP solution as well as the other two comparison methods (i.e. MA and OPER) were then applied on three in-situ calibration factor time series of UV-MFRSR (368 nm) at three UVMRP sites. The GP solution handles the variation in slope, noise, sampling density, and length of gap in the three cases as expected. Since irradiance at 368 nm is not measured by a collocated (and calibrated) radiometer, the performance of the three methods is validated against the collocated AERONET sites in terms of AOD. The results show that AODs calculated using GP-derived UV-MFRSR calibration factors (Vo_norm) have consistently better agreement with AERONET AODs than MA and OPER in terms of average absolute and relative differences, and linear regression R^2 values. These results suggest that the proposed GP solution is a robust method for time series analyses of data with variable mean, noise, sampling density, and length of gap, and has potential for application across disciplines.~~.~~

Appendix A. The formulation between the overall standard deviation and the subgroup standard deviation

Given a time series $\{x_i\}$, its total N points are divided into J groups $\{x_k^j\}$, and the number of points in group j is N_j (j=1,2,…,J; k=1,2,…, N_j). For data points in each group, their sample mean and standard deviation are μ_j and s_j. For the entire time series, its sample mean is $\mu=\frac{1}{N}\sum_{i=1}^{N}x_i$, and the sample variance is

$$\begin{aligned}
s^2 &= \frac{1}{N-1}\sum_{j=1}^{J}\sum_{k=1}^{N_j}(x_k^j-\mu)^2 \\
&= \frac{1}{N-1}\sum_{j=1}^{J}\sum_{k=1}^{N_j}[(x_k^j-\mu_j)+(\mu_j-\mu)]^2 \\
&= \frac{1}{N-1}\sum_{j=1}^{J}\sum_{k=1}^{N_j}(x_k^j-\mu_j)^2+\frac{1}{N-1}\sum_{j=1}^{J}\sum_{k=1}^{N_j}(\mu_j-\mu)^2+\frac{2}{N-1}\sum_{j=1}^{J}\sum_{k=1}^{N_j}(x_k^j-\mu_j)(\mu_j-\mu) \\
&= \frac{1}{N-1}\sum_{j=1}^{J}N_j s_j^2+\frac{1}{N-1}\sum_{j=1}^{J}N_j(\mu_j-\mu)^2
\end{aligned},$$

where the third term on the right-hand side is equal to zero because $\sum_{k=1}^{N_j}(x_k^j-\mu_j)=0$ (i.e. $\mu_j=\frac{1}{N_j}\sum_{k=1}^{N_j}x_k^j$). If assume that the sample standard deviation of each data point is invariant (i.e. $s_1=s_2=\cdots=s_J=\hat{s}$), then

$$s^2=\frac{N-J}{N-1}\hat{s}^2+\frac{1}{N-1}\sum_{j=1}^{J}N_j(\mu_j-\mu)^2 \quad .$$

Appendix B. The distribution of the 368-nm in-situ calibration factors of UV-MFRSR.

Since the true 368-nm in-situ calibration factors are not available, their distribution is derived using the ~~AEROENT~~ AERONET 368-nm AOD distribution via Beer's Law (transformed Langley regression).

Beer's Law links the irradiance [or voltage (V)] at top of atmosphere with the one reaches ground at time t with the equation: $V_t=V_o e^{-TOD_t\cdot m_t}$, where m_t is the airmass at time t and TOD_t is the corresponding total optical depth. For the 368-nm channel, *AOD* is the main contributor for the *TOD* variation. Therefore, for a short time period, TOD_t can be expressed as the sum of a constant optical depth ($\bar{P}$) and variable residual aerosol optical depth ($\Delta AOD_t=AOD_t-\overline{AOD}$): $TOD_t=\bar{P}+\Delta AOD_t$. To derive an unbiased V_o, Langley regression (in the transformed $\ln V\cdot m^{-1}$ vs. m^{-1} coordinate system), it requires the participating measurements have a constant TOD_t over the calibration period and $\ln Vo$ is the slope of the regression. ΔAOD_t biases to the regression slope with the component varying linearly with m_t^{-1} (Chen et al., 2014). Therefore, we decompose ΔAOD_t as the sum of a constant term (α) and a m_t^{-1} term (βm_t^{-1}), where α and β are obtained from daily AERONET 368-nm AOD measurements via linear regression. With the TOD_t components expanded, the original Beer's Law equation is expressed as $\ln V_t\cdot m_t^{-1}=-\left(\bar{P}+\alpha\right)+\left(\ln V_o-\beta\right)m_t^{-1}$ and the (transformed) Langley regression obtains the slope ($\ln V_o=\ln V_o-\beta$) via linear regression. The disturbed distribution of V_o is the same as the distribution of $\exp(\ln V_o-\beta)$. Assuming the true V_o is 1500 mV (a typical value at OK02) and using a long-term set of β values from AERONET at Cart_Site (17 January 2007 to 11 June, 2011), a set of V_o is obtained. Removing the tails on the distribution of V_o (i.e. V_o <1200 or V_o >1800), the normal test of the V_o set (using the Python function scipy.stats.normaltest (D'Agostino and Pearson, 1973)) returns the p value of 0.4689, which is greater than the threshold (10^{-3}), suggesting that the V_o set comes from a normal distribution.

Data availability

The in-situ calibration factors (sun-earth distance normalized) used in this study were downloaded from the UVMRP website: http://uvb.nrel.colostate.edu/UVB/da_queryVoIntercepts.jsf. The cosine corrected voltage and airmass were obtained from https://uvb.nrel.colostate.edu/UVB/da_queryCosCorrected.jsf. The spectral response functions of the UV-MFRSRs were obtained from http://uvb.nrel.colostate.edu/UVB/da_queryFilterFunctions.jsf. The site latitudes and heights of the three UMVRP sites tested in this study were obtained from https://uvb.nrel.colostate.edu/UVB/uvb-siteinfo.jsf. ~~The solar irradiance spectrum at top of atmosphere was obtained from Chance and Kurucz, 2010.~~ The ~~AEROENT~~ AERONET (v2.0) data (i.e. aerosol optical depth and surface pressure) used in this study were downloaded from https://aeronet.gsfc.nasa.gov/cgi-bin/webtool_opera_v2_new.

Author Contributions

Authors Maosi Chen and Zhibin Sun are equally significant contributors to the research. Methodology, M. Chen and Z. Sun; Software, M. Chen; Analysis, M. Chen, Z. Sun, C., C. Corr, and J.M. Davis; Validation, M. Chen, Z. Sun and Y.-A. Liu; Writing-Original Draft Preparation, M. Chen and Z. Sun; Writing-Review & Editing, M. Chen, Z. Sun, C. Corr, J. M. Davis, Y.-A. Liu and W. Gao; Supervision, J.M. Davis and W. Gao; Project Administration, W. Gao; Funding Acquisition, W. Gao.

Acknowledgments

This work is supported by the US Department of Agriculture (USDA) UV-B Monitoring and Research Program, Colorado State University, under USDA National Institute of Food and Agriculture Grant 2016-34263-25763. We thank Rick Wagener (PI) and the team for the effort in establishing and maintaining the U.S. Southern Great Plains (SGP) Cloud and Radiation Testbed (CART) Site. We thank Brent Holben (PI) and the team for the effort in establishing and maintaining the AERONET Mauna_Loa site. We thank Brent Holben, Christopher M.B. Lehmann, and their team in establishing and maintaining the AERONET BONDVILLE site.

References

Alexandrov, M. D., Lacis, A. A., Carlson, B. E., and Cairns, B.: Remote Sensing of Atmospheric Aerosols and Trace Gases by Means of Multifilter Rotating Shadowband Radiometer. Part I: Retrieval Algorithm, Journal of the atmospheric sciences, 59, 524-543, 10.1175/1520-0469(2002)059<0524:Rsoaaa>2.0.Co;2, 2002.

Alexandrov, M. D., Marshak, A., Cairns, B., Lacis, A. A., and Carlson, B. E.: Automated cloud screening algorithm for MFRSR data, Geophysical Research Letters, 31, 10.1029/2003GL019105, 2004.

Alexandrov, M. D., Kiedron, P., Michalsky, J. J., Hodges, G., Flynn, C. J., and Lacis, A. A.: Optical depth measurements by shadow-band radiometers and their uncertainties, Appl. Opt., 46, 8027-8038, 10.1364/AO.46.008027, 2007.

Alexandrov, M. D., Lacis, A. A., Carlson, B. E., and Cairns, B.: Characterization of atmospheric aerosols using MFRSR measurements, Journal of Geophysical Research: Atmospheres, 113, D08204, 10.1029/2007JD009388, 2008.

Augustine, J. A., Cornwall, C. R., Hodges, G. B., Long, C. N., Medina, C. I., and DeLuisi, J. J.: An Automated Method of MFRSR Calibration for Aerosol Optical Depth Analysis with Application to an Asian Dust Outbreak over the United States, Journal of Applied Meteorology and Climatology, 42, 266-278, 10.1175/1520-0450(2003)042<0266:Aamomc>2.0.Co;2, 2003.

Augustine, J. A., Hodges, G. B., Cornwall, C. R., Michalsky, J. J., and Medina, C. I.: An Update on SURFRAD—The GCOS Surface Radiation Budget Network for the Continental United States, Journal of Atmospheric and Oceanic Technology, 22, 1460-1472, 10.1175/jtech1806.1, 2005.

Bodhaine, B. A., Wood, N. B., Dutton, E. G., and Slusser, J. R.: On Rayleigh Optical Depth Calculations, Journal of Atmospheric and Oceanic Technology, 16, 1854-1861, 10.1175/1520-0426(1999)016<1854:orodc>2.0.co;2, 1999.

Burša, M., Raděj, K., Šima, Z., True, S. A., and Vatrt, V.: Determination of the Geopotential Scale Factor from TOPEX/POSEIDON Satellite Altimetry, Studia Geophysica et Geodaetica, 41, 203-216, 10.1023/a:1023313614618, 1997.

Cardema, J. C., Rausch, K. W., Lei, N., Moyer, D. I., and Luccia, F. J. D.: Operational calibration of VIIRS reflective solar band sensor data records, SPIE Optical Engineering + Applications, San Diego, CA, 2012, 6,

Cebula, R. P., DeLand, M. T., and Schlesinger, B. M.: Estimates of solar variability using the solar backscatter ultraviolet (SBUV) 2 Mg II index from the NOAA 9 satellite, Journal of Geophysical Research: Atmospheres, 97, 11613-11620, 10.1029/92JD00893, 1992.

Chandorkar, M., Camporeale, E., and Wing, S.: Probabilistic forecasting of the disturbance storm time index: An autoregressive Gaussian process approach, Space Weather, 15, 1004-1019, 10.1002/2017SW001627, 2017.

Chen, M., Davis, J., Tang, H., Ownby, C., and Gao, W.: The calibration methods for multi-filter rotating shadowband radiometer: a review, Frontiers of Earth Science, 7, 257-270, 10.1007/s11707-013-0368-9, 2013.

Chen, M., Davis, J., and Gao, W.: A new cloud screening algorithm for ground-based direct-beam solar radiation, Journal of Atmospheric and Oceanic Technology, 31, 2591-2605, 10.1175/JTECH-D-14-00095.1, 2014.

Chen, M., Davis, J., Sun, Z., and Gao, W.: Two-stage reference channel calibration for collocated UV and VIS Multi-Filter Rotating Shadowband Radiometers, SPIE Optical Engineering + Applications, San Diego, CA, 2015, 96100L,

Chen, M., Zempila, M.-M., Davis, J. M., King, R. W., and Gao, W.: In-situ calibration of the water vapor channel for multi-filter rotating shadowband radiometer using collocated GPS, AERONET and meteorology data, SPIE Optical Engineering + Applications, San Diego, CA, 2016, 14,

D'Agostino, R., and Pearson, E. S.: Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$, Biometrika, 60, 613-622, 10.2307/2335012, 1973.

Eaton, M. L.: Multivariate statistics : a vector space approach, Wiley, New York, 1983.

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, Journal of Geophysical Research: Atmospheres, 104, 31333-31349, doi:10.1029/1999JD900923, 1999.

Gao, W., Davis, J. M., Tree, R., Slusser, J. R., and Schmoldt, D.: An Ultraviolet Radiation Monitoring and Research Program for Agriculture, in: UV Radiation in Global Climate Change: Measurements, Modeling and Effects on Ecosystems, edited by: Gao, W., Slusser, J., and Schmoldt, D., Springer Berlin Heidelberg, Berlin, Heidelberg, 205–243, 2010.

Harrison, L., and Michalsky, J.: Objective algorithms for the retrieval of optical depths from ground-based measurements, Appl. Opt., 33, 5126-5132, 10.1364/AO.33.005126, 1994.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 66, 1-16, 10.1016/S0034-4257(98)00031-5, 1998.

Holben, B. N., Tanré, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J. S., Chatenet, B., Lavenu, F., Kaufman, Y. J., Castle, J. V., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N. T., Pietras, C., Pinker, R. T., Voss, K., and Zibordi, G.: An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, Journal of Geophysical Research: Atmospheres, 106, 12067-12097, doi:10.1029/2001JD900014, 2001.

Hyndman, R. J.: Moving Averages, in: International Encyclopedia of Statistical Science, edited by: Lovric, M., Springer Berlin Heidelberg, Berlin, Heidelberg, 866-869, 2011.

Kassianov, E. I., Flynn, C. J., Ackerman, T. P., and Barnard, J. C.: Aerosol single-scattering albedo and asymmetry parameter from MFRSR observations during the ARM Aerosol IOP 2003, Atmospheric Chemistry and Physics, 7, 3341-3351, 10.5194/acp-7-3341-2007, 2007.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resources Research, 42, 10.1029/2005WR004376, 2006a.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resources Research, 42, 10.1029/2005WR004368, 2006b.

Kazadzis, S., Kouremeti, N., Diémoz, H., Gröbner, J., Forgan, B. W., Campanelli, M., Estellés, V., Lantz, K., Michalsky, J., Carlund, T., Cuevas, E., Toledano, C., Becker, R., Nyeki, S., Kosmopoulos, P. G., Tatsiankou, V., Vuilleumier, L., Denn, F. M., Ohkawara, N., Ijima, O., Goloub, P., Raptis, P. I., Milner, M., Behrens, K., Barreto, A., Martucci, G., Hall, E., Wendell, J., Fabbri, B. E., and Wehrli, C.: Results from the Fourth WMO Filter Radiometer Comparison for aerosol optical depth measurements, Atmos. Chem. Phys., 18, 3185-3201, 10.5194/acp-18-3185-2018, 2018.

Kim, S.-W., Jefferson, A., Yoon, S.-C., Dutton, E. G., Ogren, J. A., Valero, F. P. J., Kim, J., and Holben, B. N.: Comparisons of aerosol optical depth and surface shortwave irradiance and their effect on the aerosol surface radiative forcing estimation, Journal of Geophysical Research: Atmospheres, 110, doi:10.1029/2004JD004989, 2005.

Kim, S.-W., Yoon, S.-C., Dutton, E. G., Kim, J., Wehrli, C., and Holben, B. N.: Global Surface-Based Sun Photometer Network for Long-Term Observations of Column Aerosol Optical Properties: Intercomparison of Aerosol Optical Depth, Aerosol Science and Technology, 42, 1-9, 10.1080/02786820701699743, 2008.

Krotkov, N. A., Bhartia, P. K., Herman, J. R., Slusser, J. R., Labow, G. J., Scott, G. R., Janson, G. T., Eck, T., and Holben, B. N.: Aerosol ultraviolet absorption experiment (2002 to 2004), part 1: ultraviolet multifilter rotating shadowband radiometer calibration and intercomparison with CIMEL sunphotometers, Optical Engineering, 44, 041004, 10.1117/1.1886818, 2005a.

Krotkov, N. A., Bhartia, P. K., Herman, J. R., Slusser, J. R., Scott, G. R., Labow, G. J., Vasilkov, A. P., Eck, T., Doubovik, O., and Holben, B. N.: Aerosol ultraviolet absorption experiment (2002 to 2004), part 2: absorption optical thickness, refractive index, and single scattering albedo, Optical Engineering, 44, 041005-041001–041005-041017, 2005b.

Kupilik, M., and Witmer, F.: Spatio-temporal violent event prediction using Gaussian process regression, Journal of Computational Social Science, 10.1007/s42001-018-0024-y, 2018.

Mather, J. H., and Voyles, J. W.: The Arm Climate Research Facility: A Review of Structure and Capabilities, Bulletin of the American Meteorological Society, 94, 377-392, 10.1175/bams-d-11-00218.1, 2013.

Press, W. H.: Numerical recipes in C : the art of scientific computing, 2nd ed., Cambridge University Press, Cambridge Cambridgeshire, New York, 1992.

Proietti, T.: Trend Estimation, in: International Encyclopedia of Statistical Science, edited by: Lovric, M., Springer Berlin Heidelberg, Berlin, Heidelberg, 1613-1616, 2011.

Rasmussen, C. E., and Williams, C. K. I.: Gaussian processes for machine learning, Cambridge, MA: MIT Press, Cambridge, MA, USA, 266 pp., 2006.

Richter, P., and Toledano-Ayala, M.: Revisiting Gaussian Process Regression Modeling for Localization in Wireless Sensor Networks, Sensors, 15, 22587, 2015.

Shaw, G. E.: Error analysis of multi-wavelength sun photometry, pure and applied geophysics, 114, 1-14, 10.1007/bf00875487, 1976.

Slusser, J., Gibson, J., Bigelow, D., Kolinski, D., Disterhoft, P., Lantz, K., and Beaubien, A.: Langley method of calibrating UV filter radiometers, Journal of Geophysical Research: Atmospheres, 105, 4841-4849, 10.1029/1999JD900451, 2000.

Tang, H., Chen, M., Davis, J., and Gao, W.: Comparison of aerosol optical depth of UV-B monitoring and research program (UVMRP), AERONET and MODIS over continental united states, Frontiers of Earth Science, 7, 129-140, 10.1007/s11707-013-0376-9, 2013.

Viereck, R. A., Floyd, L. E., Crane, P. C., Woods, T. N., Knapp, B. G., Rottman, G., Weber, M., Puga, L. C., and DeLand, M. T.: A composite Mg II index spanning from 1978 to 2003, Space Weather, 2, 1-13, 10.1029/2004SW000084, 2004.

Wahba, G.: Smoothing Splines, in: International Encyclopedia of Statistical Science, edited by: Lovric, M., Springer Berlin Heidelberg, Berlin, Heidelberg, 1349-1353, 2011.

Wu, R., and Wang, B.: Gaussian process regression method for forecasting of mortality rates, Neurocomputing, 10.1016/j.neucom.2018.08.001, 2018.

Yin, B., Min, Q., and Joseph, E.: Retrievals and uncertainty analysis of aerosol single scattering albedo from MFRSR measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 150, 95-106, 10.1016/j.jqsrt.2014.08.012, 2015.

Zhang, M., Gong, W., Ma, Y., Wang, L., and Chen, Z.: Transmission and division of total optical depth method: A universal calibration method for Sun photometric measurements, Geophysical Research Letters, 43, 2974-2980, 10.1002/2016GL068031, 2016.