Dear Reviewer,

Please find bellow our responses to your valuable comments regarding the manuscript entitled "TEMIS UV product validation using NILU-UV ground-based measurements in Thessaloniki, Greece".

For better readability, your comments are highlighted in bold while our responses are provided in blue.

Sincerely,

Dr. Melina Zempila.

Reviewer #1: Comments and Suggestions

Some rationale should be provided why TEMIS data were evaluated with NILU-UV measurements and not directly with Brewer measurements, which should be the most accurate. While the calibration of NILU-UV measurements against the Brewer measurement with the NN technique is a very interesting novel approach, it involves an extra step leading to an increase in the uncertainty of ground-based measurements.
I realize that that NILU-UV data have much larger temporal resolution than Brewer measurements but it is not clear whether this is important considering that only daily dose data from TEMIS were evaluated. For example, are there large gaps in Brewer measurements, which would favor the NILU-UV data set? Is there an analysis that shows that the high temporal resolution of the NILU-UV data is critical for satellite data validation?

We agree that Brewer data provide higher accuracy since, as the reviewer indicates, NILU data also include the uncertainty of the NN retrieval. However, for this study in order to evaluate the TEMIS daily doses we used data of 10-minute time intervals, the time resolution of the TEMIS UV dose time integration. NILU provides data with the necessary time resolution in order to acquire higher number of coincidences at the exact time of the TEMIS model estimation during a day. Unfortunately, Brewer's time frequency spans from 20 to 40 minutes (page 6 / line 23). Under cloudy conditions, this higher time resolution is considered more beneficiary for the accuracy of the comparisons. Thus, we chose to use NILU data in order to have a daily representative value.
To make this clear we also added a short description on page 8, lines: 3-4. We hope that this is sufficient.

"The B086 provides measurements with a time frequency of 20 to 40 minutes, but atmospheric circumstances can change considerably within this period. It is therefore better to base the evaluation of the TEMIS UV dose rate (available at 10-minute intervals) on the NILU103 data, which have a better temporal resolution; thus they suffer much less from changes in atmospheric conditions (like clouds) during one measurement than the Brewer measurements."

Differences between instruments are often given with a 0.01% precision. Considering that the uncertainties of all datasets are much larger, I suggest to round percentages to 0.1% throughout the paper, including the figures. This would also improve the readability of the text.

We thank you for the suggestion. We updated all pertinent graphs and text accordingly.

Specific comments

P2, L7: The sentence "Furthermore..." is confusing. It implies that the production of Vitamin D is detrimental. Mention the benefits of Vitamin D and then discuss that there may be an ideal UV exposure, which balances the harmful and beneficial effects of UV radiation!

We rephrased the sentence to "On the other hand, the cutaneous production of vitamin D, a 'vitamin' that is proven to be essential for human health, is also activated by spectral UV radiation.

Hence accurate knowledge of 'safe' UV doses for humans is paramount in order to balance the harmful and beneficial effects of UV exposure."

P3, L30: I note that the 1987 CIE norm for the UV index has been updated. See: Webb, Ann R., Harry Slaper, Peter Koepke, and Alois W. Schmalwieser. "Know your standard: clarifying the CIE erythema action spectrum." Photochemistry and photobiology 87, no.2 (2011): 483-486. for details. Considering that TEMIS uses the old (1987) norm, it is OK to use this norm throughout the paper, but the new norm could be mentioned.

We thank the reviewer for pointing us to the updated CIE spectrum. In the forthcoming upgrade of the TEMIS service, the updated CIE spectrum will be used: the expected impact on the UV index values will be small, but we consider that it is important to follow the official standard. We have rephrased the beginning of Sect. 2.2, where UVI-CIE is introduced.

"In the current v1.4 TEMIS service, the UVI is based on the CIE action spectrum described by McKinlay and Diffey (1987). Webb et al. (2011) describe an improved version of that action spectrum adopted by CIE in 1998. The effect of this improvement on the UVI values is small, well below 1% except for high solar zenith angle situations (Webb et al., 2011). The improved CIE erythemal action spectrum will be included in the forthcoming upgrade (v2.0) of the TEMIS service."

P4, L9: I note that the action spectrum for DNA damage suggested by Setlow (1974) is only defined for wavelengths up to 365 nm. The parameterization by Bernhard and Seckmeyer (1997), which was based on a suggestion by the NDSC steering committee (now NDACC), uses 370 as the terminal wavelength. In contrast, the spectrum drawn in Figure 1 goes up to 400 nm. The difference between the longest wavelength (365, 370, or 400 nm) is not negligible because additional contributions from the UV-A decrease the sensitivity to ozone considerably. The authors should ensure that the definition used by TEMIS is identical to that used in their work. Because the list of authors also includes colleagues that are involved in creating new versions of TEMIS products, I suggest that they carefully consider the latest definitions of the erythemal, DNA-damage, and Vitamin D action spectra when preparing a new TEMIS version.

We appreciate the comment. For this study we used the exact same action spectra with the ones that TEMIS uses to avoid discrepancies due to different applied spectra as you indicate.

P4, L13: Please specify the wavelength shift!

We now provide this information as stated bellow:

"The difference, which includes a wavelength shift of 3 nm (the applied action spectrum peaks at 295 nm and not at 298 nm as proposed by CIE), …"

P5, L6: No. Equation (1) already defines the UV Index. So either delete this sentence or define Eq. (1) and the subsequent descriptions at erythemally weighted irradiance.

In the TEMIS processing the UVI(t) is computed in W/m2 as indicated in Eq. (1), with a time dependent SZA(t), and as such it is used in the integration over time t to determine the daily UVD. Only when reporting the UV index at local solar noon UVI(t=12h) the scaling to dimensionless units is performed, which is why this sentence is present. Describing UVI(t) in Eq.(1) as "erythemally weighted irradiance" is a good idea, thank you – the idea has been implemented, but without "erythemally", as it is valid for all action spectra.

In the following sentence, UVD should be calculated by integrating the erythemally weighted irradiance instead of integrating the UVI.

Yes, the UVD is an integration over UVI(t) over time t from sunrise to sunset, with SZA(t) dependent on time, where UVI(t) is the UV index at time t. It sounds a little confusing perhaps, but calling the UV index at local solar noon (the quantity communicated to the public) just "UV index" is actually the confusing part of this.

We rephrased the whole description trying to convey this message.

P5, L9: If a cloud fraction within a 0.5◦x 0.5◦ grid cell is defined, the resolution of the satellite must be much better than 0.5◦x 0.5◦. What is it?

The cloud fraction is derived from the MSG cloud mask. The resolution of the MSG measurements varies with latitude/longitude: along longitude 0 the resolution at latitude 30N is about 0.04 degrees, and at latitude 60N it is about 0.08 degrees.

Eq. (4) is curious. If Ag is zero, fA should be 1. Yet it is 0.9775. When Ag is 1 (e.g., pristine new snow), it should be about 1.5 for erythemally weighted irradiance, yet it is only 1.3. Because Eq. (4) is part of the TEMIS code, it cannot be changed, however, it should be pointed out that the equation (which was empirically derived from measurements at two urban sites) may not be a good parameterization for large parts of the area relevant to the TEMIS UV product, which includes Scandinavia.

Eq. (4) is correct because it is emperically based on the (average) ground albedo Ag of the measurement sites used for the parameterization. This means that the albedo correction factor fA equals 1 for Ag=0.09.

Many factors determine the actual enhancement of the UV due to upward diffused radiation backscattered to the surface. We do not see why this should lead to a factor equal to 1.5.

P6, L25: "are less than 5.6%". Delete "less than". (The concept of "uncertainty" defines

a distribution (typically normal) and 5.6% defines the width of that distribution.)

We deleted it.

L6, L33: According to the text, only UVB-1 data were corrected for the degradation of the instrument's absolute spectral response. According to my knowledge, also NILU-UV instruments are subject to drifts. If the NILU-UV channels have drifted, as I suspect, a paragraph should be included in the manuscript describing how these drifts were corrected. How often was their calibration adjusted based on comparison with the Brewer? When comparing with the Brewer, did you take into consideration that the time associated with the Brewer measurements is different for every wavelength and did you interpolate NILU-UV measurements to the times of Brewer measurements?

Thank you for the comment. On the same page, line 26 we mentioned that NILU103 measurements were calibrated with coincident B086 measured irradiances.
We also added a paragraph, as suggested, to make sure that all the details are conveyed through the manuscript:
"Specifically, for the calibration of NILU103 raw data, cloud free response weighted irradiances were derived from B086's measured spectra. Since B086 scans the UV solar spectrum within approximately 7 minutes, the time period needed to scan the spectral range of each NILU103's channel spectral response, is approximately 3 minutes. The coincidences of NILU103's raw data to B086's weighted spectra, were performed based on the time that B086 measured the wavelength at which each channel peaks. Subsequently, the time difference that can be introduced between the two datasets is normally less than ±1 minute. To account for this time window, the mean values of 3 consecutive NILU103 measurements were analyzed, with the central one chosen to be the closest to B086's time scan of the peak wavelength of each channel. Then, NILU103's data were corrected for possible drifts in time via a time dependent smoothing spline fit. Furthermore, the drifts of the channels were monitored through monthly lamp measurements. Both methods resulted in the same patterns for the drifted channels.
After correcting for time drifts, a time independent absolute calibration factor is derived through scatter plots based on linear regression through origin. To evaluate the validity of the calibration procedures, the NILU103 calibrated data were compared once again with B086 response weighted irradiances and the timeseries were checked for time drifts and SZA dependence. By calibrating the NILU103 measurements with the B086 coincident response weighted irradiances, we estimate that the uncertainties of the NILU103 measurements used in this study are 5.6% (Zempila et a., 2016a)."

Section 3.2.1: Follow-up to the previous comment: what was the time associated with a effective doses calculated from the Brewer measurements? Since a Brewer spectrum takes several minutes to record, the time is ambiguous.

For the Brewer's effective doses, we considered as measuring time, the time when brewer scanned the wavelength of the peak for each action spectrum. For the DNA damage dose the starting time of the scan was taken into account. Since we used only cloud free cases, we consider that this approach doesn't introduce uncertainties larger than those of the NILU measurements themselves, even for larger SZAs.

The text was updated accordingly (page 7, lines: 27-31).

"The corresponding effective doses have been calculated by integrating the weighted spectra over the nominal wavelength range, while the time of measured doses was matched to the time that B086 scanned the wavelength where the highest sensitivity of each action spectrum is found. Since DNA damage action spectrum peaks at the lower measured wavelengths, the correspondent time was chosen to be the starting point of the scan. It appears that in most cases the 3 doses have time differences less than 1 minute."

Figure 2: Replace "mu" in legend with "Average"

Thank you. We have changed the first sentence in the caption from: "Model selection. (Top) The z-scores of the input variables and the erythemal UV dose (CIE)."

to:

"Model selection. (Top) Boxplots of the z-scores of the input variables and the erythemal UV dose (CIE) with mean values denoted by μ."

P9, L3: What is the variable "n"?. Line 12 suggests that n is the total number of data records. However, if log(n)ˆ1.5 = 36, n would be about 8E10 or 80 billion. This number must greatly exceed the number of NILU-UV data records!

n = 47,908 is the number of co-located input-output vectors (Page 8, Line 15). To help the reader, we have put log(n) inside brackets so that the expression now reads → (log(n))^1.5. Precisely, this gives 35.3793 which we rounded to the next multiple of 2 to get the value of 36 quoted in the manuscript (page 8, line: 27).

First paragraph Section 3.2.3: The description of the calculation of effective Vitamin D dose could be improved. For example: (1) Calculate effective dose for the response function of the UVB-1 (2) Convert this instrument response function weighted dose to erythemal dose taking into account SZA and total ozone (e.g., as described on page 6, line 29). (3) Convert erythemal dose to Vitamin D dose using the parameterization suggested by Fioletov et al. (2009). (4) Apply correction (Eq. (5)).

Thank you, we enriched this section.

P11, L9: The empirical relationship for DNA-damage effective dose is indeed very complex. What was the idea behind this complicated parameterization?

The initial idea was to get the DNA-damage effective dose from the CIE and the main factors that determine its levels, i.e. the TOC and the SZA in a relatively simple way (without involving look-up tables). Although the specific quantity could be directly derived from the Brewer spectra, getting it from the YES UVB-1 radiometer provides higher temporal resolution and more accurate calculation of the daily doses. We could not find a simpler parameterization than the one provided in the paper, for which the calculated quantities are accurate and unbiased from the dependent variables (TOC, SZA). Thus, although the parameterization is very complex we used it for the purposes of the present study. We provide it in the document since it might either be directly used by other people, or help them to find an improved, more general parameterization.

Eq. (6): In the second term, replace UVI with CIE.

Thank you, we did.

Eq. (8): The term CIEˆ3 appears twice, with coefficient a4 and with coefficient a6. This makes little sense. CIEˆ3 should only appear one with the coefficient a4+a6= -0.0354.

Thank you. It was a typo. The equation has been corrected properly.

P13, L10: Delete "exact"

Thank you, we did.

Figure 5: The seasonal variation in 2011 appears to be much stronger than in other years. What is the reason? Wildfires? Perhaps there is something interesting that could be learned!

The reviewer is correct. In early summer 2011, wildfires took place at the suburbs of the city, while data logging of the NILU's data was interrupted due to power failures resulting in less data points.

P15, L10 "...respectively." > "...respectively (Figure 7)."

Thank you, we did.

P17, L7 and Figure 8: The right side of Figure 8 only shows 5 discontinuities. I would expect many more if cloud information is updated every half hour, as the text indicates.

We think that discontinuities should be seen when the cloud information changes "significantly" within the 30 minutes steps. Based on the cloud information update frequency, a set of 3 points onto the graph corresponds to data with the same cloud information. If the cloud information does not change or changes slightly, discontinuities are absent of hard to be seen. Please check the changing point at around 800 (minutes) on the right hand plot of figure 8. We expect the changes to be seen within days with rapidly changing cloudiness conditions.

P17, L13: How were cloud-free data characterized? What dataset was used to determine sky condition?

The filter we are using for defining the cloud free cases stated on page 13, line 22 is the same for all comparisons, apart from figure 9 were we evaluate the cloud influence on the TEMIS-NILU comparisons. We added the following sentence in order to clarify this selection criterion (Page 13, lines: 23-25)

"This cloud classification criterion according to which days with more than 70% abundance of cloud free measurements are characterized as cloud free, is used throughout the study, unless stated otherwise."

Again on page 17, lines: 10-11, we also emphasize on this detail.

"At this point it should be mentioned that for the characterization of the cloud free one-minute data, the cloud screening detector proposed by Zempila et al. (2016a) was applied on the NILU103 Photosynthetically Active Radiation (PAR) measurements"

P19, L8 and Figure 10, and P22, L6: I don't see much difference in the slope for AOD < 0.4 and > 0.4. Perhaps the difference would become more obvious if the symbol size in Figure 10 were to be reduced.

Unfortunately resizing the marker size ends to a faint and hard to read figure. To support this statement, the linear fits of each dataset were calculated, one for AOD<=0.4 and one for AOD>0.4. For all three daily doses, CIE, DNA damage and vitamin D, the slopes are significantly larger for AOD<=0.4 than those calculated for the cases where AOD was higher than 0.4. An additional paragraph provides this information into the text (page 20, lines: 5-9).

"To further testify on this aspect, linear fits were conducted for two datasets, one that comprised data with AOD≤0.4 and the second with data with corresponding AOD>0.4. It was found that for all three UV effective doses, the slopes for the first imposed limitation on AOD were higher than those calculated for the second dataset. Specifically, the slopes for the two AOD limitations were found to be 44.5% and 11.7% for the CIE, 50.6% and 8.5% for the DNA damage, 46.1% and 8.3% for the vitamin D doses respectively."

Appendix A:

Please specify the numbers of s1 and s2 (or the range if the numbers are not constant).

Thank you. We have explicitly stated the values of s1 and s2 in the appropriate sentence in the Appendix (Page 23, Lines 17-18) as follows:

“Layer 1 (the “hidden” layer) contains s1 = 13 neurons each having a nonlinear activation function f1 = tanh and Layer 2 (the “output" layer) contains s2 = 3 neurons each having a linear activation function f2.”

Technical corrections:

While the quality of the language is generally good, many sentences are too long and this affects the readability. Whenever possible and appropriate, the authors should reduce the length of sentences and split them in two.

We agree with the reviewer, thus, we shortened the sentences where possible.

P2, L5: Change “UV sunlight” to “solar radiation in the UV range”. By definition, “light” should only be used to describe wavelengths visible to the human eye.

Thank you for the information. It was changed to solar UV radiation.

P2, L6: Delete “extreme”. Mutations can technically be triggered by only one photon.

Thank you, we did.

P3, L5: “...product services started in the 2003 and...”

Thank you, we did change the sentence accordingly.

P3, L13: “following for example changes in the operationally assimilated...2003) which were initially based on the....and later on GOME-2...”

Thank you, we did change the sentence accordingly.

P3, L20: “... SEVIRI instruments that have been operational...”

Thank you, we did change the sentence accordingly.

P3, L32: “The UVI-CIE is given as a dimensionless number...”

Thank you, we did change the “UVI-CIE” to “UVI” in order to be consistent.

P4, L16: 'bare'? Finding a better word is indeed challenging. Perhaps: raw, uncorrected, approximate, first-guess...

We changed this to "first guess of the UV index".

P4, L17: "... is then calculated from UVI' by..."

Thank you, we did change the sentence accordingly.

P6, L10: "...triangular-like slit resulting in a bandwidth of 0.55 nm FWHM.

Thank you, we did change the sentence accordingly.

L6, L13: higher SZA > larger SZA (so not to confuse with "higher Sun")

Thank you, we did change the sentence accordingly.

P9, L21 and figure 3: I don't see any change in the colors of between a training fraction of 50% and 90%, consistent with the text. So if the proportion of training data has almost no effect, why is it so important to discuss this and include a figure? Is your point to illustrate that that your results are basically independent of t/n? The left figure could be simplified by plotting MSE versus the number of neurons.

Thank you, you are correct. As we describe in the text on Page 10, Lines 6-7, and as you note, the training MSE is not sensitive to the training fraction for large numbers of input-output vectors – rather it is sensitive to the number of neurons. While we agree that the same conclusion can be drawn by plotting MSE versus neurons, there would be a loss of information on the lack of sensitivity to training fraction. The left figure embraces both concepts in one go and is why we decided against doing this.

P9, L33: "ballpark" > "rough" or "approximate"

Thank you, we did change that to rough.

P15, L10: datasets are > datasets is

Thank you, we changed all occurrences.

P18, L7: either of the > all

Thank you, we changed this point.

P18, L11: Move “on average” to end of sentence.

Thank you, we changed this point.

P19, L24: “in the” > of

Thank you, we changed this point.

P21, L25: moments > periods

Thank you, we changed this point.

P21, L28: “limits the dataset by almost 75%” > “make up only 25% of the dataset” (if that’s what you want to say)

Changed to “The number of cloud-free days limits the dataset to one fourth of the original, while …”