"This review was prepared as part of graduate program course work at Wageningen University, and has been produced under supervision of Prof Wouter Peters. The review has been posted because of its good quality, and likely usefulness to the authors and editor. This review was not solicited by the journal."

HEADER

The Tibetan Plateau is a region which undergoes significant climate change. Air temperatures have increased with 1.39 K since 1850 while the amount of incident solar radiation decreased. The consequences of this solar dimming phenomenon on surface warming are still unclear. Previous research shows contradictory conclusions regarding the proper attribution of solar dimming. Therefore, the roles of clouds and aerosols will be investigated in this study to provide more clarity regarding the causes and impacts of solar dimming.

The paper is well written and the different sub-sections improve the readability and enable the reader to search for specific sections. I feel confident about the data analysis and interpretation done by the authors. However, there are some important remarks regarding certain assumptions, significance of results and data visualisation. I would strongly recommend considering and including these remarks in the manuscript before publication. I will come back to these remarks in more depth in the remainder of this review. Firstly, I want to emphasise what I thought to be very good and interesting about this research. To start with the introduction which describes in a clear and convincing way why this research is relevant. The current controversy regarding the proper attribution of solar dimming is a driving force for this research to introduce new knowledge and provide a conclusive answer. In order to generate this new knowledge, multiple high-quality data sources have been used: model simulations, remote sensing products and ground measurements. The methods applied seem quite advanced and are well-documented in previous literature which makes the methods trustworthy because it can be checked and compared with other research. Especially the improved accuracy of the generated downward surface radiation datasets by applying the NNLS method is a very strong aspect of this research. The solar dimming phenomenon has a large effect on local but also on global climate change. It turns out that humans are largely responsible for the increase of air pollution which turns out to be the main driving factor of solar dimming. The role of human activities in remote areas is discussed and emphasises the societal relevance of the topic.

MAJOR ARGUMENTS

Major argument 1:
The method which is used for the attribution analysis of solar dimming is the optimal fingerprinting method. It's is based on a linear relationship between driving variables and a responding variable, in this case downward shortwave radiation (DSR). When the scaling factor is larger than zero at a certain significance level, the variable has a positive contribution towards the responding variable. My concern regarding this method is that no value of the significance level is given in the manuscript. The results of the attribution analysis indicate that anthropogenic aerosols (AA) are the main cause of solar dimming. However, it's not clearly described or listed if other variables were tested with optimal fingerprinting method besides the noAA simulation and if there were variables which didn't reach the required significance level and are consequently left out of the analysis. The CMIP5 simulations with and without AAs have uncertainties which are indicated by the shaded area in figure 6a. Zhou et al. (2018) calculated the 5%-95% confidence intervals using Monte Carlo simulations. Do these shaded areas and errors bars represent the same confidence intervals and are they calculated in a similar manner? It is stated that the overall variation is of significance tested but the outcomes of these statistical tests or thresholds (p-values, r-values, etc.) are not included. The time over which the method is applied is divided in two periods: 1950-2005 and 1970-2005. Is the selection of these periods linked with the respective increase and decrease of downward longwave and shortwave radiation? Do you believe that two periods are enough to describe the trend in the data? Yao et al. (2018) described for example that the heating of the Tibetan Plateau began in the 1960s but reached the highest levels in the last 30 years which indicates that significant changes in the climate have occurred within the selected periods.

The results show that the scaling factor is positive for the AA simulation and negative for the noAA simulation, which supports the conclusion that AAs are the main driver of solar dimming. Especially the scaling factor for the AA simulation for 1970-2005 seems convincing with small error bars and a mean value close to 1. If other variables would have been included in the analysis, the scaling factors could be compared with the scaling factor of the AA simulation. This would show the relative contribution of other factors and possibly strengthen the assumption that AAs are indeed the main driving factor. It can be observed that the scaling factors become more positive and more negative for the shorter time period. The error for noAA (1970-2005) is quite substantial in my opinion because the total length of the error bars covers approximately 1/3 of the length of the y-axis of the graph. The negative scaling factor for noAA is attributed to the decrease in cloud cover. The evidence for this statement is obtained from figure 5c where the temporal variation in cloud cover from three satellite data sources is shown. However, the satellite data only covers the period 1980-2005/2015. Thus, from the period 1950-1980 there is no data available to support this conclusion. In addition, the trend of the ISCCP data shows a slight increase of cloud cover which doesn't support the statement that the cloud fraction decreased over time.

I would recommend providing the value(s) of the significance level in the methodology section and indicate if certain variables were left out of the analysis. Could these variables be included when the significance level would change and would this make a difference for the outcomes of the analysis in your opinion? The results of the analysis would be more robust if the values of statistical tests and thresholds are included with the results and figures in the manuscript. Currently I have to believe that the variation is of significance tested without this statement being supported by numbers. Could you elaborate a bit more the selection of the two different time periods in the methodology section, why did you choose for these periods? Am I correctly assuming that it's related to the turning-point of the increase of longwave radiation and the decrease of shortwave radiation? The negative scaling factor of the noAA simulation, with the largest uncertainty, is completely attributed to the decrease in cloud cover which is supported by 2 out of 3 data sources whereas the third data source indicates a slight increase in cloud cover. What is your opinion on the controversy regarding these results and do you have possible suggestions for other factors besides cloud cover which could play a role? Perhaps it would be nice to add a paragraph of discussion concerning the remarks related to this argument in the manuscript.

Major argument 2:
Shortwave and longwave radiative effects are separated in order to quantify the depressing effect of aerosols on surface warming. It is assumed that the change in air temperature is dominated by the change in surface skin temperature interacting with the air temperature through radiative and thermal processes and the change in atmospheric circulation. Consequently, the variable f is calculated which represents the sensitivity of air temperature to 1 W/m^2 radiative forcing. For this analysis I'm wondering whether it's valid to employ values of α, ε_y and S which are calculated by taking the mean values of satellite products for several years. Is there a substantial variation between different products and how large or small is the error estimate of this mean value? From the introduction and other studies, it becomes clear that this region undergoes significant climate change which is supported for example by the analysis done by Yao et al. (2018) regarding oxygen isotopes in ice cores collected at glaciers at various locations. The Tibetan Plateau contains large amounts of snow and ice and is called the Third Pole for a reason. Warming and consequent melting of snow and ice could substantially change the albedo. The positive snow-albedo feedback could accelerate the change in albedo and warming over the Tibetan Plateau (Zhang et al. 2003). In addition, other studies indicate that black carbon (BC) and dust are responsible for about a 20% reduction of the albedo (Schmale et al. 2017). However, the results in this study show that the amount of dust decreased over time. Is the amount of BC somehow related to the amount of PM2.5 and could this be responsible for the decrease in albedo besides the decrease due to snowmelt? My main concern regarding this method is whether it's valid to assume that a mean value can represent the rapid changes caused by a positive feedback mechanism in combination with other factors like dust and BC. Additionally, it's not clearly stated over how many years this average is taken, if multiple averages were used for different time periods and which satellite products were used.

The results show that the aerosol radiative forcing has been increasing by 8.08 W/m^2 between the first and last 30 years of climatology. However, the supplementary figure S5 shows a negative forcing anomaly which implies a decrease of the radiative forcing. The depressing effect of aerosols on air temperature is calculated using two methods: first-order approximations of the direct near-surface air temperature response to each radiative and thermodynamic component (α, ε_y and S are included using this method) and multiple noAA simulations. It can be observed that the methods show similar depressing magnitudes in the supplementary figure S6. If the albedo is overestimated because the effects of the snow-albedo feedback can't be captured by taking the mean value, the temperature anomaly could start to deviate and will likely result in a larger value. Consequently, this will have an effect on the mean of the two methods which is represented in figure 8.

Could you elaborate a bit more on the thought of reasoning behind the assumption of employing the mean values of satellite products for these variables (especially concerning the albedo). What are the exact values and sources of these variables which were used for the analysis and do they correspond with previous studies or observations? Perhaps an analysis of the albedo from the downward shortwave radiation products could be included to visualise the temporal variation of the albedo. It's stated in other research that dust and BC can be responsible for a reduction in the albedo besides the snow-albedo feedback. In this research it's shown that dust shows a decreasing trend since 2000 whereas PM2.5, which is related to air pollution, shows an increasing trend. Could there be a possible relationship between PM2.5 and BC which could also contribute to the change in albedo and would you perhaps consider this in the manuscript or future research? Related to the suggestion of the previous major argument, could you include the statistical information regarding the shaded areas in figure 8, S5 and S6.

Major argument 3:
The final aerosol depressing effect on the Tibetan Plateau climate warming is calculated by taking the average over two data sources where one included and the other ignored the heat exchange with the surroundings. The first-order approximation which consists mainly of remote sensing products ignored the heat exchange with the surroundings. The CMIP5 noAA simulations are assumed to be less reliable but did compute the influence of the interaction with other regions. Thus, it is stated that the remote sensing products had more reliable input than the model calculations but this is not supported by numbers/ statistical tests/ previous literature. Furthermore, it seems counterintuitive because the accuracy of the CMIP5 datasets is improved by the NNLS method. Is it a sound methodology to lump these two sources of data together for the final depressing effect and assume that the exchange is considered to a certain extent? Personally, I'm not convinced by the assumption that the interaction with the surroundings can largely be ignored. In the introduction it is stated that the Tibetan Plateau is a weak heat sink in winter but a strong heat source in summer which is already indicative for differences between the seasons. Also, it's mentioned that the large-scale orography is crucial for water and heat exchange between the Pacific Ocean and Eurasia.

This assumption focusses on the exchange of heat with the surroundings but what about other types of exchanges? Aerosols resulting from air pollution in surrounding areas enter the Asian tropopause aerosol layer by deep convection. From here they are consequently transported to other locations. This is an important pathway for anthropogenic aerosols to enter the Tibetan Plateau, which is thought to be the main cause for solar dimming in this region (Lau et al. 2018). Furthermore, the depressing effect calculation is assuming that the change in air temperature is mainly driven by radiative and thermal processes and the change in atmospheric circulation: advection of cold and warm air masses. Again, related to an interaction with the surroundings. Are these interactions included in the results? Could you elaborate a bit more on the points mentioned above in the reply-to-the-reviewer?

I would like to see the supporting material in the manuscript regarding the statement that remote sensing products have a more reliable input than model calculations. A follow-up point of discussion is then related to taking the mean value of these data sources. Figure S5 shows the mean value of the two datasets (with and without interaction). When the two sources of data are separately added to the figure, it enables a visualisation of how they differ/ relate to each other and what their magnitude is in comparison to the mean value. Furthermore, can you justify why the heat exchange is ignored while substantial differences between seasons are found? The final depressing effect propagates in the calculation of the air temperature anomaly which plays a key role in the interpretation and attribution of the solar dimming phenomenon and its effects on surface warming over the Tibetan Plateau.

MINOR ARGUMENTS AND ISSUES

MINOR ARGUMENTS

Minor issue 1:
There is a difference in the validation of shortwave and longwave radiation due to a system bias at the GAME and CAMP networks caused by disparate instruments. The manuscript states that it's a "minor validation difference" but could you please provide a quantification of the difference?

Minor issue 2:
Can the spatial mismatch between radiation datasets and site observations be ignored, even though this is in line with former studies? Especially because the results of this study focus on spatiotemporal variation over the Tibetan Plateau it seems a bit counterintuitive to accept a spatial mismatch in data validation.

Minor issue 3:
Firstly, it is stated that deep convective clouds have little influence over the Tibetan Plateau whereas further in the text it is described that aerosols enter the Asian tropopause aerosol layer by deep convection. I assume that deep convection occurs in the regions surrounding the Tibetan Plateau and the aerosols are consequently transported. Thus, deep convective clouds are important but in an indirect pathway.

Minor issue 4:
The overall variation of multiple models (AA and noAA simulations) is of significance tested in temporal analysis and optimal fingerprinting method. However, no values of a statistical test are given.

MINOR ISSUES

Page 1, line 12: missing "a" before "higher accuracy"
Page 1, line 15: "the fastest decrease in DSR is in the southeastern TP". Maybe it looks nicer to write that the fastest decreases occurs/ can be found in the southeastern TP.
Page 2, lines 31 and 33: Firstly, increased surface air temperature is mentioned in line 31. Afterwards in line 33 this suddenly becomes surface temperatures. Is the same variable meant here or are we talking about two different things?
Page 2, line 41: Please be careful with the word "significantly" when it's not supported by a value or reference.
Page 2, lines 42 and 43: I would recommend being consistent with terminology. In the abstract and in line 36 the term is introduced as solar dimming whereas in these lines it's mentioned as TP dimming.
Page 2, line 47: missing "the" before "TP"
Page 2, line 48: "spatial temporal variation" is used whereas elsewhere in the paper the word spatiotemporal variation is used. Or say: "spatial and temporal variation".
Page 3, line 72: It's stated that datasets are chosen which have a spatial resolution less than 2º. However, in Table 1 there are two datasets which have a resolution of 2.50º x 1.88º and one with 2.50º x 1.26º. Are these datasets not used in the analysis? If they are not used it might be better to remove them from the table.
Page 3, line 84: Perhaps it's better to move the link to the reference section of the manuscript. It seems out of place here.
Page 4, line 103: Is spatiotemporal resolution meant, or spatial and temporal resolution?
Page 5, line 130: I think that "lack" should become lacking. Or "because the sensor calibaration lacks long-term stability".
Page 5, line 150: missing "a" before "comparable accuracy"
Page 6, line 166: Perhaps it's better to move the link to the reference section of the manuscript. It seems out of place here.
Page 6, line 168: Perhaps it's better to move the link to the reference section of the manuscript. It seems out of place here.
Page 6, line 171: I would phrase the beginning of this sentence slightly different. Perhaps "collected data from 5 GEBA sites" or "included 5 sites from the GEBA network".
Page 6, line 173: I would phrase this sentence slightly different. Perhaps "even though the number of sites is not large enough…".
Page 7, line 194: "Given that radiative fluxes are always positive, ….". What kind of sign convention is used here? Usually downward directed fluxes are positive whereas upward direction fluxes are negative (a loss for the surface).

Page 10, line 278: compressing does not seem like the right word in this context. Perhaps counteracting or diminishing the greenhouse effect?
Page 10, line 279: missing "the" before "TP"
Page 11, line 319: missing "a" before **"**different conclusion"
Page 12, line 339: missing "the" before "TP"
Page 12, line 340: missing "the" before "TP"
Page 12, line 343: missing "the" before "TP"
Page 12, line 357: "causing **a lower** elevation in the model than **in** reality"

Page 26, Figure 1: The elevation map which is plotted as background has a scale ranging between 0 and 9000 m. It's difficult to figure out at which location the individual ground networks are located. Could you please add a scale which is better to read?
Page 26, Figure 1: In the central Tibetan Plateau, the network is quite dense and the symbols overlap each other. Could it be possible to provide a zoom-in on this specific area?
Page 26, Figure 1: The caption became very long because all the projects are mentioned by their full name instead of the abbreviation.
Page 28, Figure 3: The Shi and Liang data covers a relatively small amount of time in comparison with the CMIP5 data. Therefore, I think that adding the regression (for the short period only) is not adding a lot of extra surprising information because the trend is already quite obvious from the time-series. In addition, no statistics regarding the regression are mentioned.
Page 28, Figure 4: The caption doesn't mention which data is used for Surface DSR, DLR and Mean Air Temperature. This is mentioned for the air temperature data obtained from ground measurements.
Page 28, Figure 4: In the caption of the figure suddenly a p-value of <0.01 shows up which is not clearly mentioned in the manuscript.
Page 29, Figure 5: In all four panels are linear regression lines added, again without any extra statistical information.
Page 29, Figure 6: The shaded area is not explained in the caption. Are these confidence intervals? For the second panel, the link with the methodology can be a bit stronger so it becomes clear that this figure belongs to the optimal fingerprinting method.
Page 30, Figure 7: There is a regression line plotted but there are only four points in the figure, and again no statistical significance mentioned.
Page 30, Figure 7: I would have phrased the first line in the caption different because now it seems that the changes are variable instead of variables which are changing.
Page 7, Figure 7: In the manuscript, only an explanation is given for the summer season while the other three seasons are plotted as well. Why is it useful to leave them in the figure when nothing is mentioned about them?
Page 7, Figure 8: The data source is not completely clear to me from the figure caption. Additionally, extra lines which are not represented in the legend are present in the figure (yellow, light-blue and orange). What do these lines represent? The caption and legend should have provided this information. Finally, the shaded area is not explained in the caption. Are these confidence intervals?

Supplementary material:
Page 4, Figure S2 and S3: The statistical quantification is lacking for the regression (e.g. R^2-values).
Page 5, Figure S5 and S6: The shaded areas represent uncertainties but it's not mentioned how large these uncertainties are. Is it the 5%-95% confidence interval? The light-red colour in Figure S6 is difficult to see.