Referee report

Development of a semi-parametric PAR partitioning model for the Contiguous United States

Kathilankal et al are presenting a semi-parametric PAR partitioning model for the US subcontinent. The main idea for this model is to use commonly measured climatic variables, which are known to correlate with diffuse fraction, as variables in a logistic model. The relative weights of these variables are determined by fitting measured diffuse fraction data. It is shown that the model outperforms the simple cubic polynomial model. The article is well written and contains a nice overview of the development of these types of models. I would recommend publication, provided that the following comments are addressed.

Scientific significance: Good

Scientific quality: Good

Scientific reproducibility: Excellent

Presentation Quality: Excellent

Specific comments

P1649. Title: I would recommend to fully write out the acronym PAR in the title to make the study clear to newcomers and outsiders to the field.

P1652. Line 27: 'removed outliers'. In the remaining of the paragraph some criteria are given. Is this the what you meant with removing outliers, or is this in addition to that? If the former, then please make the clear in the text and if it is in addition, then please clarify what the criteria are to remove outliers.

P1653. Line 10: I would like to see somewhere how many years of measurements are available for each station. Maybe this could be included in Table 1.

P1653 Eq 1: R_E is not mentioned in the text.

P1653, line 22 and Eq. 2: In line 15 the sine of the solar elevation angle is set as sin β. I would use the same variables here: thus not αS (in the text, or is this actually the albedo?) or just β (in the equation). Also in the equation, k_t appears twice.

P1653, line 3: just to be sure; k_t refers to the daily clearness index and k_tp to the PAR clearness index, correct?

P1654, line 9 and line 18. The figures are discussed here, but I am not sure I can read the figures fully. There are so many points, that some points may be overlapped by others. So, the order of printing actually matters. I believe that the conclusions in this paragraph are correct, but I am not sure if I can easily read this from the figures.

P1655, line 20-23. In Fig 2 the measured and modeled values are depicted in one plot and it seems that the modeled values lie nicely in the center of the range of measured values. However in Fig. 3 the differences show only negative values. I am not sure I understand this. I would expect that fitting would result in some positive and some negative values.

P1656 line 16. I enjoyed this interpretation of the findings, but I was wondering if there is a physical explanation why it is harder to model these conditions. If so, then please insert that here.

P1657 line 3-5: Does one of the parameters show significant different values for Sep-Dec wrt. the rest of the year?

P1662, Table 1: Indicate the total amount of data (measurement points and/or range of years). You might want to consider to remove the sensor column.

P1663, Table 2: The two sets of parameters differ wildly from each other. I presume this is because of strong correlations between certain parameters. Because of these correlations, the 95% confidence interval is of limited value, as you may not freely change all parameters within the corresponding ranges. If the authors have some insight in these correlations, then it may be worthwhile to discuss this in the text.

P1665: Fig 1: I have the impression that I do not see all the points making it hard to interpret these figures. Could the data be presented in a different way, or with a reduced data set?

P1666: Fig 2: Please explain the difference between panel a and b in the caption.

Technical corrections

P1664: Caption: estiamte ➔ estimate