Referee's Report on the revised version

Many thanks to the authors for taking many of my numerous suggestions into account and for making substantial changes to the manuscript. It has improved considerably compared with the first version. Many parts of the manuscript, such as that on particle forcings, are well balanced, well written and a pleasure to read. I do not agree with some of the arguments/answers put forward by the authors, but am willing to let them pass in the interests of time. However, there are still two important issues to be clarified/dealt with before the manuscript can be published. There are also a few minor points that the authors may care to look at. Let me begin with the two major issues still present in the revised version of the manuscript.

1. Description of the SOLID composite and Fig. 3: The authors are to be lauded for substantially extending the description of the SOLID composite and for introducing a table describing which data sets have been used. They have now also given a reference to Haberreiter et al. (2017) in which a more detailed description is promised. This is a significant improvement over the first version of the manuscript. All the same, there are still a number of unanswered questions.

For example, the manuscript states that the significant deviations of the SOLID composite from the observations shown in Fig. 3 are explained by the existence of other datasets listed in Table 1, but not shown in Fig. 3. Coincidently, these other, not shown, datasets almost exactly compensate the difference between the data that are plotted and CMIP6. If this is indeed the case, then the authors are to be congratulated, but from the brief description in the paper, it does not become clear how this happens. For instance, the disagreement between SOLID and SORCE/SIM in the 200-400 nm range is explained by averaging with SORCE/SOLSTICE. However, SORCE/SOLSTICE data are only available shortward of 309 nm (Table 1 in the manuscript), whereas much of the flux variability measured by SIM in the 200-400 nm range comes from the 309-400 nm domain not covered by SOLSTICE (see Fig. 1 from Haigh et al. 2010). According to the manuscript this flux should enter SOLID with weight 1 so that it remains unclear why SOLID is so different from the SIM data.

Likely this and other questions that I still have will be reslved by the more detailed description in the manuscript submitted to JGR by Haberreiter et al. (2017), which unfortunately is not accessible. I therefore ask the authors to make this paper available to me via the editor, to enable evaluating the strengths and weaknesses of this approach. The information likely provided there will hopefully quickly clarify the issues and questions I still have regarding the SOLID composite.

2. Future scenarios: Section 3 has improved, but is still substantially weaker than most of the rest of the paper. It also remains misleading, maybe partly because the authors seem to have misunderstood my comments in the last report. I apologize if they were unclear. Let me try to be clearer this time. In spite of the "scenarios" in the title, large parts of Sect. 3 give a false sense that in effect some kind of prediction of future solar activity is being made. True, in the revised version "predict" has been replaced by "forecast" in a few places, but dictionaries (e.g., Collins English dictionary) use the two words synonymously (although they do have somewhat different meanings in meteorology). Also, in spite of these changes in name, a lot of trouble is taken (and quite some text has been added) to argue that the two provided time series are reliable forecasts and not just scenarios.

However, uncertainties in the adopted approach are likely strongly underestimated.

For me the claimed ability to predict or forecast solar activity on longer timescales, although even the amplitude of the next cycle cannot be predicted reliably by dynamo models (or by other means except partly from the strength of the polar fields during the previous minimum), is an assumption, and not a statement of fact, as the current version purports it to be. Assumptions should be clearly identified as such to the reader. On the basis of this assumption and various purely statistical models the authors then “forecast” future solar activity, while the assumption itself remains completely untested in this paper and very poorly tested otherwise. E.g., given the many unknowns affecting the solar dynamo and consequently the very wide variety of dynamo models that have found their way into the recent literature, it may be questioned whether the predictions of any one current dynamo model will stand the test of time.

Even if this basic assumption of Sect. 3 were correct, it is unclear that any of the statistical methods used by the authors, or an ensemble of such methods, will be any good in predicting solar activity in 50, 100 or 300 years. I cannot find a reliable reason for the belief apparently held by the authors that purely statistical “more elaborate reconstruction schemes” will give the correct behavior a number of cycles down the line, while failing already at the next cycle.

The trouble with all the methods used by the authors is that they have not been tested on solar activity on the time scales of interest to the authors (and cannot be properly tested for 50, or 100 years, or even longer periods without waiting that length of time). Therefore, the value of all these extrapolations, be they called “predictions” or “forecasts”, is fairly limited.

Take the argument that the Sun has just left a grand maximum and is now likely to stay at a low level of activity and consider the recent revision of the sunspot number. Until a new consensus is reached a number of different sunspot number data sets vie with each other, some of which do not even agree with the statement that the last 6-7 decades was a grand maximum. E.g., according to Svalgaard and Schatten (2016) there have been at 3 episodes since the Maunder minimum, when the activity was approximately as high as in the last 60-70 years. Therefore, the authors possibly have been extrapolating on the basis of wrong assumptions regarding past solar activity. Please note that I am not arguing that Svalgaard and Schatten are correct, but that currently even the sunspot number record is very much under debate, placing a big question mark behind some of the assumptions made here (which, yet again, are sold for a fact). Consequently, the uncertainties in the extrapolations made in the present paper are much larger than the authors acknowledge.

In summary, I am arguing that the “forecasts” or “scenarios” of future activity proposed here are reasonable-looking courses that future activity may possibly take, but they are not more probable than any number of other such scenarios. This is because we do not have reliable, i.e. thoroughly and successfully tested means of forecasting solar activity on the timescales of interest. To be clear, I am not arguing that the authors remove or change their time series (although they show only part of the possible range of solar activity in the upcoming decades, thus restricting the range of solar forcing in CMIP6 climate runs). Rather I am arguing that the authors acknowledge that these are 2 of many equally possible scenarios.

It is necessary to strongly tone down the claims that actual forecasts of activity are presented, removing the impression that the time series being provided are better than many other time series that could be proposed. Ideally, much of the introductory part of Sect. 3, as well as almost the entire Sect. 3.1 should be removed. This could become part of a separate paper in an appropriate journal. In such a paper the authors could provide more details, much more stringent tests etc. It could allow readers to estimate the reliability of the employed statistical methods when applied to solar data.

However, as this means a major rewrite of this section, I would be satisfied if the authors changed the text to reflect the fact that what they are proposing really are to some extent random scenarios and should not be taken as reliable forecasts or predictions.

Minor issues

1. page 2. lines 1-2: The manuscript still contains the confusing sentence about CMIP6 “statistically indistinguishable from available observations”. The answer to the referee explains what the authors mean by this. Such an explanation should also be added to the manuscript to avoid confusing readers

2. page 3. lines 30-31. The sentence it is still not fully correct. Please note that Usoskin et al. (2014) do not discuss TSI reconstruction. Possibly the authors mean reconstructions of solar activity?

3. SATIRE-TS has been converted everywhere to SATIRE, but SATIRE-TS is still mentioned in the legends of Figures 1, 2, 4, 6 (see previous report).

4. page 11, lines 25-28. “In both models, the historical reconstructions are sensitive to the assumptions made when constraining them to direct (satellite era) observations that suffer from large uncertainties. This mainly explains why these models differ before 1990 by an offset.” Which assumptions are meant? Why do they lead to an offset? Please explain in the manuscript.

5. page 8. line 16. “a disk-integrated ratio of the core to the-wings of the Mg II emission line at 280 nm”. This is not how the Mg II index is defined (a disk integrated ratio is not the same as the ratio of disk integrated quantities); see the paper by Viereck et al. (2001), cited in the same sentence of the revised manuscript. Please correct.

6. Figure 4 may be misleading to some readers, as the sum of the SSI variability of the various models does not seem to be the same (and hence cannot all be equal to the TSI variability). This is likely because the IR is missing. I would propose to (ideally) add the IR band to the figure (as done by Ermolli et al. 2013, although a larger wavelength coverage would be better), or to at least point this out in the text.