Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-91-RC2, 2016

Geoscientific Model Development Discussions
Open Access

Interactive comment on "Solar Forcing for CMIP6 (v3.1)" *by* Katja Matthes et al.

Anonymous Referee #2

Received and published: 13 July 2016

This is an important paper as it is meant to be a standard reference for the climate models taking part in the CMIP6 exercise. It is rather comprehensive, covering many topics. Some parts are well written, easy to understand and, as far as I can see, are internally consistent. However, other parts are inconsistent or unclear. Below I list the points where the paper could be improved. Many of these are minor issues such as typos etc. that should be easy to take care of, but there are a few points that are more fundamental and which the authors must deal with carefully. They might require bigger changes to the manuscript.

1) page 1, line 10 (abstract): "The TSI and SSI time series are defined as averages of two (semi-) empirical solar irradiance models, namely the NRLTSI2/NRLSSI2 and SATIRE-TS." Two comments: a) NRLTSI2/NRLSSI2 are empirical, not semi-empirical models. This is correctly explained later in the text of the manuscript but the statement in the abstract is misleading. b) More importantly, could the authors describe why they

have taken recourse to this unusual step of averaging two independent models? Both models are constructed differently, and both have to some extent been tested against data. What the authors do is to provide a new model that is largely untested with regard to solar data. More comments on this aspect will be made later.

2) page 2, lines 1-2: "The slight negative trend in TSI during the last three solar cycles in CMIP6 is statistically indistinguishable from available observations". What do the authors mean by this? Different TSI composites indicate different trends. There are implicit assumptions underlying this statement that need to be spelt out and the reasoning clarified.

3) page 2, line 5: CMIP6 cannot be tested against CMIP 5. It can only be compared to CMIP5.

4) page 2, line 6: The expression "background SSI" is neither clear nor used in the literature.

5) page3, line 5: "Because of its prominent 11-year cycle, solar variability may offer a degree of predictability for regional climate and could therefore help reduce uncertainties in decadal climate predictions." However, the solar cycle itself is notoriously hard to predict and predictions of upcoming solar minima have not been particularly successful, so that the statement does seem too optimistic. But possibly the authors are not concerned with such niceties here?

6) page 3, lines 23-24: "The quantitative assessment of radiative solar forcing has been systematically hampered so far by the large uncertainties and the instrumental artifacts that plague TSI and SSI observations" The TSI observations are signiïňĄcantly more precise than those of SSI (especially SSI in the near UV and visible spectral domains). This makes the quoted sentence misleading.

7) page 3, line 28: The IAU resolution (Mamajek et al. 2015) adopted the result of Kopp & Lean (2011). The latter is a well-known and well-cited paper in a refereed journal,

while the former is not a scientific paper at all, is not properly published, and the text in the reference is not really useful. Please replace the reference to IAU resolution with the Kopp & Lean paper everywhere in the text.

8) page 3, line 29: The proxy reconstructions of the TSI do not exhibit "occasional phases of unusually low nor high activity", the Sun does. Also, Usoskin et al. 2014 did not mention TSI reconstructions at all.

9) page 3, line 32: "There is growing evidence for the Sun to enter a phase of low activity near 2050, after a grand maximum that peaked during the 20th century." The authors should explain what they base this claim on. The solar dynamo is chaotic, possibly even stochastic and it is even difficult to predict the next solar cycle much before the previous minimum. Going beyond the next cycle is even tougher. See e.g. * J. Jiang, R.H. Cameron, M. Schüssler, 2015: The cause of the weak solar cycle 24; ApJL 808 L28 * Cameron et al. 2013: Limits to solar cycle predictability: Cross-equatorial flux plumes, A&A 557, A141

10) page 4, line 12: " . . . with respect to the CMIP5 solar forcing recommendation." Is the CMIP5 paper supplying the solar input cited? At least so far "CMIP5" is not associated with such a paper.

10a) page 4, line 20: "Lockwood et al., 2019" This is likely "Lockwood et al., 2010"; also to be corrected on page 74

11) page 6, line 15: " . . . and one observational estimate (SOLID), . . ." Is this correct? From the evidence given in the paper, SOLID is not an "observational estimate" but an empirical model (see below).

12) page 6, lines 26-27: The sentence simplifies things too much. Faculae are not "the Mg II index" and dark sunspots are not "sunspot area". These proxies are used to represent the contributions from faculae and sunspots to the irradiance. This should be written more carefully to reflect the actual relationships.

13) page 7, line 8: ". . ..SORCE SSI observations Lean and DeLand (2012) the wavelength dependent scaling coefficients . . ." Change to: ". . ..SORCE SSI observations (Lean and DeLand, 2012) the wavelength dependent scaling coefficients . . ."

14) page 7, lines 5-15: The description of the adjustments in the NRLSSI2 model is not clear. The authors should explain better how the adjustments are done.

15) page 7, line 19: What is SATIRE-TS? It does not become clear from reading the paper, as references are given only to SATIRE-T and SATIRE-S. TS is used multiple times, so that it cannot be a typo. Is TS some combination of the two models? What kind of combination?

16) page 7, line 28: The reconstruction made by Dasi-Espuig et al. (2014) is not based on sunspot number.

17) page 8, line 3: "In NRLSSI2, this internal consistency also applies to the integral of the facular and sunspot contributions to SSI to their respective counterparts in TSI (see Coddington et al. (2015) for more details). Why are the authors stressing this point? This sounds rather trivial. Doesn't every model that distinguishes between spots and faculae do that? Or are they implying that SATIRE does not fulfil such internal consistency? They should either explain why this is such an achievement and is unique to NRLSSI, or they should remove this unnecessary sentence.

18) page 8, lines 5-6: "The controversial out-of-phase behavior of SORCE/SIM observations in that band (Harder et al., 2009) are likely to be an instrumental artefact (Lean and DeLand, 2012; Ermolli et al., 2013) but this has not yet been corrected in the SOLID composite." If SORCE is wrong, is then the SOLID composite the best one to use to test the SSI models? If models are chosen by how close they are to SOLID, this could have unwanted effects for the atmospheric chemistry and finally climate modelling. As shown by Haigh et al. (2010, An influence of solar spectral variations on radiative forcing of climate, Nature, 467, 696), the SSI variability found by SORCE has a major effect on atmospheric ozone, changing its concentration in a way contrary to

expectations.

19) page 8, line 14 “In NRLSSI2, the proxy index for sunspot darkening is the sunspot area as recorded by ground-based observatories in white light images since 1882 (Lean et al., 1998). Values prior to that are estimated from the sunspot number.” Does not also the SATIRE model rely on sunspot areas over the period of time covered by Greenwich observatory? It seems to according to * Krivova et al. 2010: Reconstruction of solar spectral irradiance since the Maunder minimum; JGRA, DOI: 10.1029/2010JA015431 Please comment.

20) Sect. SOLID composite After reading this section the exact nature of the SOLID composite is still very unclear. A single paper (Schöll et al., 2016) is cited, which explains only the preprocessing of the data. As long as SOLID is not properly documented, it is important to make the description here sufficiently clear so that readers who don’t know the nitty gritty of the SOLID approach can follow and, if necessary make their own judgement. So far, this composite (if it is really a composite at all and not an empirical model, see below) is not widely-known or established in the solar community.

21) page 9, line14: “. . ., the most reasonable approach consists in averaging both reconstructions, weighted by their uncertainty. But this means that yet another model is produced, one that is untested, except for the rudimentary tests briefly discussed in the paper. It is not clear why this is “the most reasonable approach”

22) page 9, line 17: “. . . SATIRE-TS (Yeo et al., 2014).” What is SATIRE-TS? Is it a typo or a combination of SATIRE-T and SATIRE-S? Please define it. In Yeo et al. (2014) I could find a description of SATIRE-S, but no mention of SATIRE-TS, so that this reference seems not to be relevant (unless TS is a combination of T and S).

23) page 9, line 24: “The EUV band (10-121 nm) is required for CMIP6 but is absent from NRLSSI and SATIRE. We added it with spectral bins from 10.5-114.5 nm by using a nonlinear regression from the SSI in the 115.5-188.5 nm band, trained with

TIMED/SEE data from 2002 to 2009.” Indeed, this is a difficult situation and the authors have done the reasonable thing and somehow modelled this wavelength region themselves. The procedure chosen may be fine and may produce a good result, but without being shown anything it is hard for the reader to judge. Please provide a figure and some more explanation.

24) page 9, line 29: “Let us note that while the F10.7 index is a good proxy for EUV variability on daily to yearly time scales, this may not be true anymore on multi-decadal time scales.” This is a good point.

25) page 10, line 3: “. . . observed composite from PMOD” This sounds strange since the composite was not observed. Rather it is based on a set of observations of TSI.

26) page 10, line 7: “All TSI records agree well on daily to yearly time scales, and in some cases (e.g. NRLTSI1 and NRLTSI2) they match as well on multi-decadal time scales.” Aren’t both models part of the same model family and are founded on the same proxies? I am not sure what is remarkable about them being consistent with each other? The authors can leave this – it is not an important point – but I would like to understand why they stress this agreement, which I would have naively thought to be trivial.

27) page 11, Figure 1: The orange curve described as CMIP5 goes till 2015. What exactly is plotted? Are these extrapolated CMIP5 data? Or are the plotted CMIP5 data regularly updated and are available up to 2015?

28) page 11, line 12: “The major difference come from SATIRE-TS ..” “The major difference comes from SATIRE-TS ..”

29) page 12, Figure 2, bottom left diagram Why is NRLSSI2 producing such strange cycles between ∼1940 and 1960? These certainly do not look realistic. How come, these strange cycle shapes are restricted mainly to the visible (although there may be some sign of similar behavior in the IR)? Should not also the UV cycles behave strangely

in order to compensate for the behavior in the visible and produce a reasonable TSI? Has the TSI produced by NRLSSI2 been compared with measured time series? Also, something seems to be going wrong right at the start of the time series. How do the authors explain and compensate for these problems?

30) page 12, line 4: " . . . and higher-quality data from the SORCE mission on the rotational timescale in NRLSSI2, . . ." But isn't SORCE giving wrong trends for SSI? This is what the authors claim multiply elsewhere in the paper.

31) page 13, Figure 3: Fig. 3 is confusing and does not agree with the text of the paper, mainly regarding SOLID. Earlier it was said that SOLID is based entirely on observed data (page 8, line 20: "More specifically it is derived as the weighted mean of all available SSI observations in the satellite era.") If that were true, then the green curve should be a lot closer to the plotted observations. Thus, it is not clear why The SOLID composite departs so strongly from SORCE at a time when that is the only data set used (according to the authors)? Or are other data sets also used after all? The SSI variations shown by SOLID seem to be smaller than of all the instrumental records, except maybe UARS SOLSTICE (it is hard to see - there are too many light colors in this plot). Anyway, SOLSTICE shows a behavior completely inconsistent with an SSI composite of the observations. This is a serious problem that points to a fundamental inconsistency in the paper. Another strange feature of this plot is that SOLID covers also the 1950s and 1960s when there were no SSI data available. How does SOLID produce something at those times if it is purely based on SSI data? The description given in the paper is totally inadequate and obviously seriously misleading. However, Fig. 3 very strongly suggests that the SOLID "composite" uses either a proxy (possibly something like 10.7 cm flux?) or makes really strong changes to the data while processing them. In either case, I would strongly oppose calling it a composite of SSI observations. Rather Fig. 3 clearly shows that it is an empirical model. If the authors want to maintain that SOLID is a composite of observed SSI, then they should provide a detailed explanation that goes far beyond the inadequate one in the current version

of the paper. This should include a list of all data sets that enter into the SOLID "composite" and all the steps that are undertaken to produce it. Also, they should provide a convincing explanation why SOLID differs so strongly from the observational data. Fig. 3 raises an issue regarding fairness and bias in the paper. If SOLID is indeed an empirical model, and I have seen no evidence to counter this in the paper, I see no advantage in using SOLID to "test" the other two models. Indeed, if SOLID is a model (and an unpublished one at that), why is it being discussed ahead of the numerous other (published!) models in the literature. I see only two paths that that the authors can follow: a) Either remove SOLID completely from the publication and instead compare the averaged model that the authors have produced more rigorously with the observations directly, b) or discuss SOLID on an equal footing with the other SSI models that the authors simply ignore in this version of the paper. Irrespective of which of these paths the authors follow, I strongly urge them to use the original SSI observations to test the new model data set obtained by averaging NRLSSI2 and SATIRE (-TS?).

32) page 13, Figure 4 and its discussion in the text: Averaging over one month at activity maximum and minimum does not allow eliminating the rotational cycle in solar variability, so that this ïňAgure mixes information on shorter timescales into the solar cycle variability that the authors want to show. The figure should be redone using at least 81-day averaging. Why are the comparisons in Figures 3 and 4 being done in such broad, seemingly arbitrary wavelength bands, rather than broken up according to the important molecular band listed in Table 1? What is the advantage for the climate community of following the bands used by Ermolli et al. (2013). Also, where would the observations lie in Fig. 4 (to the extent available for exactly these times, which is a limitation of the figure)?

33) page 13, lines 1-2: " . . . the only available measurements are from the SORCE/SIM instrument, which has calibration issues (Lean and DeLand, 2012) . . ." Until now no calibration issues in SORCE/SIM instrument have been reported by the instrumental

team. In particular, the paper by Lean & DeLand does not identify any calibration issue.

34) page 14, line 5: “In the NIR CMIP6 shows slightly larger variability than CMIP5 and remarkable here is the largest variability in NRLSSI2.” Why is this remarkable? Both SATIRE & NRLSSI2 reproduce TSI. NRLSSI2 has a smaller variability in the UV and this must be compensated by NRLSSI2 in the IR. Or is there something more complex at work here that I am missing? Please explain or simplify the text.

35) page 17, line 1: “Solar activity and hence spectral irradiance vary between different solar cycles. However, these differences are small compared to the total 11 year solar cycle amplitude . . .” This has been the case in the second half of the 20th century, but the sizes of cycles vary between zero and the very high amplitude of cycle 19. This is only a minor quibble, however.

36) page 19, line 18: “. . . produces slightly higher SW heating rates than NRLSSI1(CMIP5)” “. . . produces slightly higher SW heating rate differences than NRLSSI1(CMIP5)”; the diagram shows the differences between heating rates for solar minimum and solar maximum.

37) page 20, Figure 5: “Impact of solar forcing according . . .” add: for perpetual solar minimum conditions

38) page 21, Figure 6: “CMIP6 SSI differences in % for perpetual solar minimum conditions . . .” It may not be immediately clear what differences are actually meant here, i.e. differences in which parameter; add e.g.: “CMIP6 SSI differences of the solar irradiance in % . . .”

39) page 21, line 1: “More important for the solar ozone signals seems to be the choice of the CCM (with its specific photolysis scheme, see also Fig. 8), especially for the lower stratosphere (10 hPa and below).” This is true for the lower stratosphere only; above that the dataset-induced differences are larger than the model-induced ones, in particular in the lower Mesosphere.

40) page 21, line 12: “Note that statistically significant irradiance differences between CMIP5 and CMIP6-SSI irradiances are particularly observed between 300 and 350nm . . . (Fig. 8).” In Figure 8 differences in the irradiance amplitude between solar minimum and solar maximum are shown. i.e. “irradiance differences” should be replaced by “differences in the irradiance amplitude”.

41) page 25, line 19: “. . .for mesospheric OH production Fytterer et al. (2015b), and for . . . “. . . for mesospheric OH production (Fytterer et al., 2015b), and for . . .”

42) page 27, Figure 11: SSN scaled by a factor of 0.67 should have larger values (in Figure 13 SSN scaled by a factor of 0.741 has values above 200); a factor of 0.067 appears to be much more reasonable.

43) page 33, Figure 16: There are differences between the caption and the labels in the diagram. caption: 70–90oS (left) and 70–90oN (right); in the diagram: 70–90oS (right); 70–90oN (left) caption: 0.01 hPa (upper panel) and 0.1 hPa (lower panel); in the diagram: 0.1 hPa (upper panel) and 1 hPa (lower panel)

44) page 36, line 13: “Since fast transient solar energetic particle events often occur at the background of enhanced geomagnetic disturbances, straight-forward computation of the particle trajectories in a realistic geomagnetic field is needed” Why are fast transient solar energetic particle events relevant in this context? This paragraph deals with the penetration of GCRs in the Earth’s atmosphere.

45) page 38, line 25: The CMIP6 future solar forcing is different from that of CMIP5. However the manuscript does not demonstrate that it is “more realistic”. The authors need to provide solid arguments for this realism or remove any such claims.

46) page 38, lines 29-31: “We ignore scenarios with high levels of solar activity because the Sun just left such an episode (called grand solar maximum), and several studies suggest that it is very unlikely to return to one in the next 300 years”. As pointed out above (see point 9 of this report), predictions of anything beyond the next cycle are

affected by chance (in the sense that the activity level can be changed significantly by singular events that in turn cannot be predicted). It also seems that statistically, from the record of past solar activity, a grand minimum is equally unlikely as another grand maximum. According to Solanki and Krivova (2011, Science 334, 916), “Half the grand maxima in (6) were followed by one or more subsequent grand maxima before a grand minimum ïňĄnally occurred.” (The reference (6) in this sentence is to Usoskin et al. 2007, Astron. Astrophys. 471, 301.) Consequently, the authors should revise the above statement and find new arguments for why they choose to concentrate on just low values of solar activity for the coming centuries.

47) “Nonetheless, memory effects associated with these periodic reversals play a major role in determining solar variability on multi-decadal time scales, and to some degree are decoupled from the short-term variability. This is our prime motivation for considering predictions on multi-decadal time scales.” As given, this is just a statement without a physical basis. As this is their prime motivation for the predictions, the authors should provide solid evidence for such memory effects and the decoupling of multi-decadal from decadal variability.

48) Sections 3.1-3.4. I have signiïňĄcant doubts about the results presented in these sections. Section 3.1 presents three forecast methods, chosen seemingly arbitrarily from all those that have been proposed. As far as I can tell, they are all in one way or another linear. For a strongly non-linear system such as the solar dynamo, I see little value in using linear forecasting methods. I argue that applying inappropriate, but complex sounding forecasting techniques projects a sense of accuracy where none is present in reality. The performance of the techniques is discussed in Sect. 3.3 and Fig. 21c. From Fig. 21a I get the impression that the errors in the Phi forecast are comparable (and over some periods exceed by a factor of 2-3, e.g. around 2080-2090) the values of Phi. Around 2200 various methods give Phi of about 100 to 400, and the forecast error is 150 for all methods. This essentially means the range of 0 to 550 (with $\sim$600 being the highest value measured during the modern Grand max). In summary,

Fig 21 shows that the three methods often give hugely different results (which is not surprising and simply reinforces that solar activity cannot be reliably predicted using such simple techniques and possibly cannot be predicted at all on these time scales). The authors then consider the mean of these three results, claiming that it represents “the most likely level of solar activity”. Such an approach can hardly be called scientiïňĄc and cannot lead to a “more realistic” forecast than CMIP5. Thus, the mean of 3 more or less random numbers is still a more or less random number of little value. The construction of the “extreme” scenario is also difïňĄcult to follow. A lot of the description is rather opaque. All this seems such a complicated way of computing something that is likely very unreliable anyway and does not provide that much reliable information. For example, the reference scenario seems to be somewhat below present conditions and stays nearly constant at that level, while the extreme scenario drops down to the Maunder minimum and basically stays there. Is that so much different from what was done for CMIP5. I would find it a lot more honest towards the reader to not invoke all these different methods, but rather to make a clear and simple assumption and to show the result it gives. This result may turn out not to be very different from what the authors are proposing now, if the authors make the appropriate assumptions. BTW, what is the meaning of a negative modulation potential (Fig. 21B) and how is it obtained?

49) page 39, line 17: “. . . is certainly presentat some level.” “. . . is certainly present at some level.”

50) page 39, line 35: “. . . using the geomagnetic reconstruction of the open solar flux Lockwood et al. (2014).” “. . . using the geomagnetic reconstruction of the open solar flux (Lockwood et al., 2014).”

51) p. 40, line 11-12: "According to solar dynamo models, the solar-cycle averaged modulation potential (and the sunspot number) cannot be predicted more than a few decades ahead." This statement is not entirely consistent with p. 39, line 7-8 "As of today, even predicting the cycle amplitude one cycle ahead remains a major challenge (Pesnell, 2012)." Which statement do the authors actually support, one cycle ahead or

multiple cycles ahead? According to Cameron et al. and Jie et al. (referred to earlier in this report; see point 9), the statement on p. 39 appears to be the valid one and the statement on p. 40 should be changed accordingly.

52) page 40, line 12-14: “The observed modulation potential has an autocorrelation function that decays exponentially with a characteristic time of 48 $\pm$ 5 years. This quantity can be interpreted as the time beyond which memory is lost.” This is weak und unconvincing evidence of memory. Earlier in the same paragraph, the authors state that they are dealing with 22-year averages of the modulation potential Phi. The decay time of 48 years is hence basically the time resolution of the data (based on the Nyquist frequency). I do not see this as evidence for a memory, just that the true resolution of the data is not very high. In addition to this argument, there may well be hidden connections between data points lying close in time, so that they are not entirely independent. This can be the case in cosmogenic isotope data, so that using this as an argument of memory should be done with considerable care. The authors should first convincingly show that individual data points are completely independent, before making claims of memory.

53) Page 40, line 15-16: “To the best of our knowledge, no existing method has been able to meaningfully predict solar activity more than 60 years ahead.” This seems to imply that there are methods that can predict up to 60 years ahead. I would like to hear more about these. E.g. how do the authors know that they work up to 60 years ahead, without waiting for another 60 years to find out? Unless, of course, they are referring to methods that are at least 60 years old and that I seem to have missed. I have seen many so-called predictions tuned to reproduce past data exceptionally well, but then do rather less well when predicting even the next cycle. This statement also is not consistent with p. 39, line 7-8 and other work.

54) Page 45, line 1: “The resulting SSN time series of both future scenarios have then be used” “The resulting SSN time series of both future scenarios have then been used”

55) page 53, line 5: NRLTSI2/NRLSSI2 are empirical models not semi-empirical ones.

56) page 54, lines 8-10: the statement is not clear (see above);

57) page 54, line 13: The statement is too ambiguous. Which satellite measurements are meant?

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-91, 2016.