Manuscript Summary

The manuscripts presents a comparison of surface downwelling shortwave radiation (SDSR) simulated by CMIP6 models against observations from the GEBA network over the period 1961-2014. The study attempts to provide additional observation constraint on CMIP6 models covering a period prior to the availability of other long-term ground based and satellite records. Evaluating CMIP6 models across Europe and China highlights important differences in the ability of models to simulate changes in the observed SDSR. The study attributes the failure of CMIP6 models to reproduce the temporal changes in SDSR over China to errors in the emissions of aerosols and their precursors and the translation into atmospheric burden. The paper represents an updated analysis using the CMIP6 models of that previously done with CMIP5 models. I think the paper presents analysis on an important topic which would be useful to the community, but I would like to see further details on the following points prior to publication.

General comments

1. It would improve the paper if more background information in the introduction section was provided on the key drivers of SDSR i.e. clouds and greenhouse gases can also influence SDSR in addition to aerosol effects.

2. Throughout the paper there are numerous mentions to the fact that aerosols play a key role in the dimming signal of SDSR observed and simulated across all regions. However, the same cannot be said for the observed brightening signal in more recent years. A key question seems to be why are aerosols a key driver in the dimming but not brightening? If the emission inventories and aerosols were in error throughout the whole period of study then surely the models would not be able to simulate the temporal evolution of both phenomenon across all regions?

3. The paper states that the CMIP6 models are able to represent the observed SDSR signal over Europe relatively well. However, I think there are a few interesting discrepancies which should be discussed further. Prior to 1980 the observations do not show much of a dimming signal (in fact the observed anomaly is slightly positive at times) but the CMIP6 models do show a consistent dimming signal. Is there a specific reason for the absence of a dimming in the observations, when we know there were large concentrations of aerosols over Europe at this time? Contrary to what was mentioned in point 2 above Europe is the only region where there is a simulated brightening signal in both the model and observations, implying that models are able to reproduce brightening signal over certain regions. It would be good to know if there a reason for this over Europe and does it occur over other regions like for example North America.

4. For the analysis over China the paper suggests that the error between the models and observations of SDSR are due to the errors in emission inventories that translate into errors in the calculation of atmospheric burden of aerosols. Are we certain that the errors in the emission inventories are that large to account for the discrepancy in model and observed SDSR? Is there an estimate of the uncertainty for the CMIP6 emission inventory and how does CMIP6 compare to other global and regional emission inventories? Can these differences explain some of the inconsistencies of models with observations? I am not convinced that the observed trend reversal in SDSR over China in ~1990 can be explained by errors in the emission inventories alone. Are we anticipating a slowing down of SO_2 emissions in China from the

1990s onwards? As far as I understand, anthropogenic emissions of aerosols and their precursors (particularly SO_2) have largely been increasing over China up until ~2010 when air pollutant control measures were then implemented to reduce emissions. Therefore, if aerosols were driving the temporal change in SDSR over China a dimming signal should have been observed up until this point, but it isn't. This is present in the observed and simulated change in SDSR over India but not China. How do this discrepancy match with the conclusions of the paper and what else could be driving the SDSR trend over China throughout this period? I think this needs to be explored further in the paper as the assumed underlying trend in emissions (and therefore aerosols) and SDSR do not seem to match over China and from what I can tell it cannot be reconciled by errors in the emission inventories alone.

5. Only limited discussion within the paper is provided on clouds and aerosol-cloud interactions, which needs to be improved throughout the paper. Within section 3.3 a link is made between cloud cover change and SDSR but how much of an influence do clouds have on all-sky SDSR? How reliable are the observed and simulated cloud cover changes and can some uncertainty bounds be placed on them? Is a regional cloud cover change of 1-2% significant in terms of SDSR? In figure 3 the temporal change in observed cloud cover is similar to that in observed SDSR so even if clouds can't explain the magnitude and all of the observed change then surely they must be exerting some influence on SDSR? Is it possible to compare a clear-sky derived observed SDSR to that from model simulations to eliminate any influence of clouds on the signal?

6. The previous comparison of CMIP5 models to observed SDSR by Storelvom et al., (2018) is mentioned throughout this study, with similar results presented here for CMIP6 models. A key question is therefore why has there been no improvement in simulating observed SDSR between CMIP5 and CMIP6 models? This is despite some changes to individual aerosol schemes within models and also different historical aerosol precursor emission datasets being used. Some discussion is needed on what is continually missing from the models and what are the model developments to focus on to improve the future simulation of SDSR.

7. Further details are required, either in Table 1 or a new table, on each of the CMIP6 models used in this study. In particular, it would be useful to know horizontal resolution and some information on the individual chemistry and aerosol schemes in each model. This could provide useful information to the reader of the potential causes of discrepancies between models. In addition, it would be good to have a record somewhere of the actual output used from the ESGF (e.g. temporal period, variant ID, CMIP table ID etc). Furthermore, if there is data now available for additional CMIP6 models then it would be useful to include it, as long as it further informs the current study.

8. The methods section (2.3) appears to lack important details of what model data is being used (see point 7) and how the gridded model data has actually been compared to the observations which are at point locations. In calculating the regional means at observation locations, do the number of sites used change over time period and does this have any impact on the results? Furthermore, in the results section the clear-sky SDSR is discussed but is not mentioned in the methods section. I also think that it is important to use multiple ensemble for meaning purposes when using coupled experiments members from models so that the internal variability in each model can be shown (this would give a range of variability important to show on some of the Figures for certain variables).

9. A General comment on the figures is that they could be improved to make them easier to read by using better colours (I found the light green very bright), tick marks on the axis and line types that are easier to distinguish between different model experiments. Also, if it is possible to include a measure of observational and model uncertainty on any of the figures then this would improve the comparisons. When values from figures are continually referred to in the text it would help the reader if there is reference table containing some of the key numbers included (like the supplementary table).

Minor Comments

Page 1, line 9 – Reword this sentence as mentioning SO2 emissions, which are not aerosols, and then other aerosols relevant to SDSR. Be more precise in this statement.

Page 1, line 13 – Can you say how much error is associated with aerosols and emission inventories that might contribute to error in SDSR?

Page 2, Line 30 – Is this statement true across all regions? What about for Europe?

Page 2, line 35 – For the introduction it would be good to include a bit more detail on what the GEBA observations on their own show before introducing any comparisons to models.

Page 2, line 46 – here the study says that two observational datasets are used but only one has been mentioned in the previous paragraph. Please include details of what is the second dataset used in this study.

Page 2, line 47 – please reword sentence "An explanation of the methods used to obtain and analyse the data complete Section 2."

Page 3, line 57 – it would be good to include the error in the observations on all figures to show the uncertainty in the observations.

Page 3, line 60 – Please clarify if this temporal gap filling technique allows for all 1487 stations to have a complete record of observations over the entire 1961-2014 and how this technique impacts the observations. If the number of stations used changes over the entire time period then it could be important for the analysis.

Page 3, line 74 – insert 'is' between "these the"

Page 4, line 93 – replace 'stales' with "stalls"

Page 4, line 94-95 – "So these experiments will show to what extent the removal of cloud cover change from global warming has an effect on SDSR." – I am sure that this is the case as there will be still be variability in the cloud fields simulated by climate models in these experiments. In addition, as the aerosol fields are changing in these experiments, they will also impact the simulated clouds in the models. Therefore, to make this statement further evidence would be required from each model that the cloud fields are being properly constrained to isolate their impacts on SDSR.

Page 4, line 107 – It would be good to show on a figure the spatial distribution of the GEBA observations within each defined region.

Page 4, line 110-112 – Please clarify exactly how anomalies have been calculated. Are anomalies calculated for each individual observation site within a region first before then calculating a regional mean value?

Page 4, line 112 - Supplementary table number is not shown

Page 4, line 113 – Provide more information on exactly what model data has been obtained from the ESGF (perhaps in a separate table) e.g. CMIP table ID, variant label etc. (see general comment 8)

Page 4, line 115 – I think it would be more prudent to use more ensemble members for coupled experiments and with this an idea of the internal variability for each model could be obtained for variables such as cloud cover and SDSR.

Page 4, line 116 – It is not clear if the 10-year running mean is used for the model data, observation data or both?

Page 5, line 121 – it is hard to see from Figure 1 a) as to whether the global SDSR representation in the models is similar to the observations at all. There is clearly a difference in magnitude but there does not appear to be a strong dimming signal in many of the models. Is this just the scale on the figure or is there not much change in the model at all? Can the Figure be improved in any way to make this easier to see?

Page 5, line 122 – Change “these discrepancy originate” to “this discrepancy originates”

Page 5, line 125 – More discussion on European model observational differences (see general comments point 3)

Page 5, line 135 – I think that this is only true for certain models as others seem to have opposite temporal changes compared to observations e.g. NorESM2.

Page 5, line 138 – It is hard to say without tick marks on the figures as to whether the end points in models are similar to the observations. For example, is a -10 Wm^{-2} anomaly in 2014 from GEBA considered to be similar to a ~-6 Wm^{-2} from NorESM2?

Page 5, line 140 – please explain what “temporal forcing evolution” means in this context.

Page 6, line 156-157 – does this imply that the greenhouse gases impact on SDSR over China throughout this period?

Page 6, line 157-158 – I am not sure this is true for all models. The temporal evolution of SDSR from CanESM5 seems quite different in the historical and piClim-histall but perhaps not so much in MIROC6.

Page 6, line 167 – Aerosols have a key role in dimming but not it appears brightening – why not? (see general comment 2)

Page 6, lines 168-169 – similar to point above in that there are differences between these simulations which don’t appear to be the temporal driver of SDSR but perhaps can influence it? It would be good to show the actual differences between models in these simulations and what influence other factors (like clouds and greenhouse gases) can have on SDSR.

Page 6, line 173 – how has all-sky SDSR been decomposed into clear-sky?

Page 6, line 179-180 – Can the clear-sky and all-sky changes be shown on the same figure to compare differences?

Page 6, line 182-189 –How have the changes in model cloud cover been calculated? This needs to be in the methods section. Also line 183-184 states that cloud cover changes mask the clear-sky SDSR signal. This implies that the clear-sky decrease would have been even larger without changes to clouds indicating that clouds do have an important influence on SDSR in models. I think this needs to be explained more - see general comment section 5 for more details.

Page 7, line 193 – "session" should be "section"

Page 7, line 194 – "In this session we found the clear-sky SDSR to be stronger than all-sky SDSR, indicating the simulated dimming is primarily caused by aerosol-radiation interactions." But also that clouds have an influence on SDSR too.

Page 7, line 205 – "SO_2 burden" is mentioned but should this not be SO_4 burden.

Page 7, line 205-206 – Given that all models have the same SO_2 emissions, do we know why the changes in SO4 burden are so different between NorESM2 and CESM2? Could this indicate some of the potential problems in translating emissions into atmospheric burden or aerosols, which lead to errors in SDSR?

Page 7, line 210 – can a more scientific term be used than "real story".

Page 7, line 210-211 – This sentence makes the assumption that aerosols are the sole driving force in SDSR and that it is only the emissions and removal processes that could be in error. Other potential causes could be mentioned like the model translation of emissions to burden which leads to the larger differences in simulated SO_4 burdens between models. Also see major comments above.

Page 7, line 212 – "the precursor of SO_2", should this not be SO_4?

Page 7, line 215-218 – Should we be expecting a trend reversal in SO_2 emissions over China between 1980 and 1990? At this point in time emissions would have been increasing over China and emissions have only begun to reduce recently (since 2010). See general comment point 4

Page 8, line 235 – Is it possible to include the clear-sky proxy from GEBA here and compare to that from models on Figure 3 to show how well models simulate the aerosol radiation interactions?

Page 8, line 238 – change "shown in Figure displayed" to "(Fig. 2) show"

Page 8, line 242 – But the magnitude of the dimming was not sufficient to reproduce that observed (same as Allen?) and implies emissions are not high enough historically?

Page 8, line 247 - change "burden of SO_2" to "burden of SO_4".

Page 8, Lines 246-249 - The study only shows change in SDSR is opposite to SO_4 burden over Europe and not the case over China so can we really say that the process of translating burden to forcings are ok? What about over other regions? Might not just be due to errors in atmospheric burdens, but other factors combining?

Page 8, Line 250 – "The models of this study …" changed to "The models used in this study …"

Page 9, line 254-255 – Should we expect a reversal of emissions across China over this period?

Page 9, line 256 – Is this referring to Figure 3 in Hoesly et al., (2018)? Make clearer.

Page 9, line 258 – should we expect BC and OC to influences SDSR much? Need to mention these aerosol components earlier in the manuscript if going to mention now as no introduction to them at all. All discussion previously has been made about SO_4 so why suddenly bring them in now?

Page 9, line 259-261 – Do these studies give an uncertainty in emission inventories and can this be used to see if it can account for the differences between model and observed SDSR.

Page 9, line 270 – change "CMIP6 experiment models" to "experiments, CMIP6 models"

Page 9, line 273 – mention that the dimming is underestimated by the models.

Page 9, line 276-279 – Would we not have anticipated the SO_4 burden to have increased across China over this period as SO_2 emissions are anticipated to have also increased? Are the errors in SO_4 burden and SO_2 emissions really that large to account for the observed discrepancy in SDSR? More work to back up this statement and other factors should be included in conclusions. Uncertainty in emission inventories probably do contribute to this but the trend changes in SDSR and anticipated emission changes don't match for China, so this cannot be the sole reason and needs to be expanded on. see general comment point 4.

Page 10, line 285-287 – how would these future investigations improve our understanding of SDSR temporal evolution?

Fig. 2b – why is CanESMS so different in Hist-Nat and does show that other drivers influence the SDSR trend?

Fig. 3b – Can the uncertainty in cloud cover from observations and models be shown?

Fig. 4 – CESM2 seems to show a small change, can you confirm that this model has interact aerosols included? If not then why such a small change compared to others?