

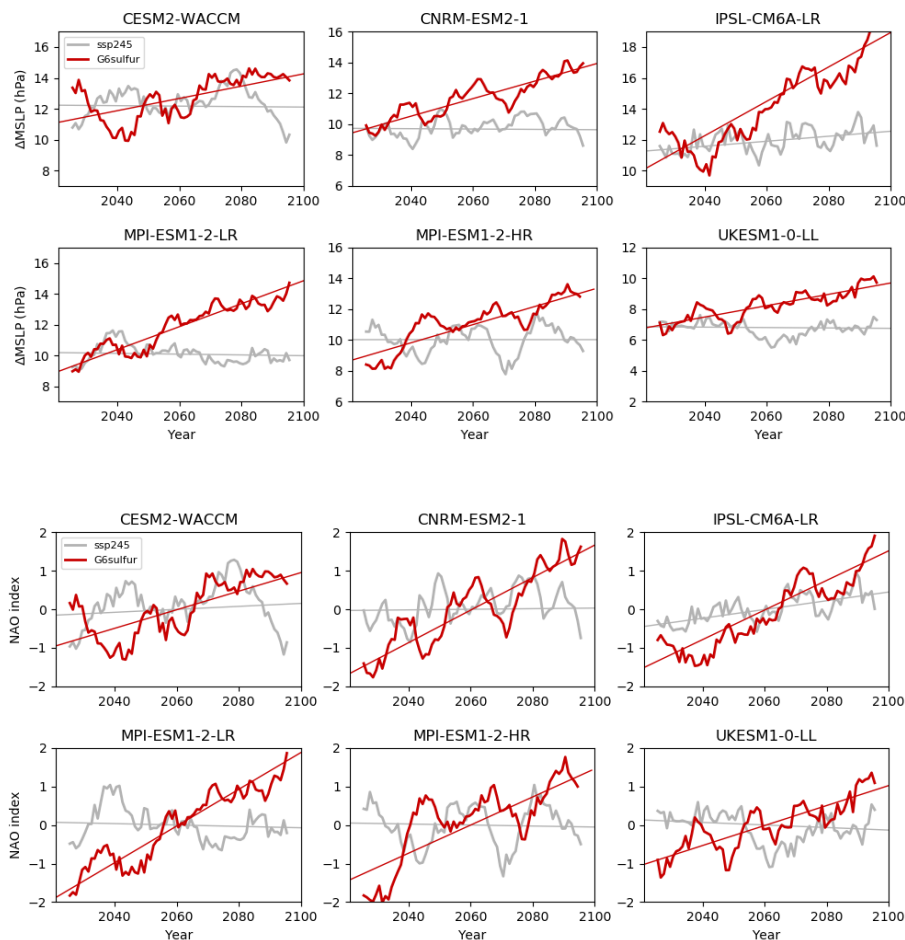
Response to Anonymous Referee #1

We thank the Referee for their comments which have helped improve the paper. We quote the Referee's comments below (in **bold**) and then provide our responses to them; revised text is shown in red.

Major comments

- 1. The definition of the NAO is unusual. Normally the NAO is defined as the difference of normalized (to unit variance) pressure in the two regions giving a dimension-less index. Here it is defined without the normalization. I see that the same definition is used in Stephenson et al. 2006 and Baker et al. 2018, but I did not find any motivation in these papers. The authors should describe the background for choosing a non-standard definition. They should also describe if the conclusions differ when using the more standard method.**

We follow Stephenson *et al.* (2006) and Baker *et al.* (2018)'s definition of the NAO because of its simplicity: being just the pressure difference between two regions makes it extremely easy to understand. Using a more complex definition of the NAO does not, however, affect our conclusions. Below we first show Fig. 2 from the manuscript using the simple definition of Stephenson *et al.* and Baker *et al.*, followed by another version of the figure using a more complicated approach to define a dimensionless NAO index (e.g., Tsanis and Tapoglou, 2019; Hurrell *et al.*, 2020):



The more complex definition first requires calculating the long-term mean and standard deviation for each region's DJF pressure timeseries, then the normalization of each series by first subtracting the long-term mean from each and then dividing by its standard deviation. The difference between the resulting series is a dimensionless NAO

index (which can have values outside the [-1, 1] range). Both versions of this figure show that the NAO is essentially unchanged in ssp245 over the period while exhibiting a clear positive trend in G6sulfur.

We have modified the text (lines 153-154) to state why we use the simple NAO definition and note that this does not affect our conclusions:

We use this definition of the NAO for its simplicity, but our conclusions are not affected by the use of a more complex NAO definition (e.g., Tsanis and Tapoglou, 2019; Hurrell *et al.* 2020).

References

Hurrell, J., and National Center for Atmospheric Research Staff (Eds): *The Climate Data Guide: Hurrell North Atlantic Oscillation (NAO) Index (station-based)*, last modified 24 Apr 2020, available at: <https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based> (accessed: 21 December 2021), 2020.

Tsanis, I., and Tapoglou, E.: Winter North Atlantic Oscillation impact on European precipitation and drought under climate change, *Theor. & App. Clim.*, **135**, 323-330, <https://doi.org/10.1007/s00704-018-2379-7>, 2019.

- 2. In Figs. 5 and 6 the multi-model means are shown. The intermodel consistency -- significance of a non-zero signal? -- is calculated as where at least 4 out of the 6 models agree on the sign. If I understand this correctly, then even for random signs this agreement will happen with a probability of more than 50 %. The authors should estimate the significance with a more strict method.**

Figures 5 and 6 have been recreated using a two-tailed Student's *t*-test to define areas where differences are significant at the 5% level; the figure captions have been changed accordingly.

Minor comments

- 1. I163: In order to assess .. I find this sentence rather convoluted. Why not compare directly to the ssp585 experiments?**

Experiment ssp585 cannot be used for comparison as it is much warmer than G6sulfur and this has been found to affect the NAO as noted in the manuscript. We have re-written the text (lines 164-170) to make this clearer:

In order to assess whether the NAO changes seen in G6sulfur are due to SAI the results from G6sulfur need to be compared against those from a similar experiment which does not include SAI and which also follows the same temperature evolution to G6sulfur, thus ruling out a straightforward comparison against ssp585. The latter condition is required because studies of scenarios with warming levels similar to ssp585 have been found to affect the NAO (e.g., Tsanis and Tapoglou, 2019). Both conditions are satisfied by GeoMIP experiment G6solar (Kravitz *et al.*, 2015) which is parallel to G6sulfur but achieves the cooling from ssp585 to ssp245 levels by the highly idealised method of reducing the specified solar output.

- 2. I183: Christiansen 2018 (10.1175/JCLI-D-17-0197.1) explains why the model mean is better than individual models and could be cited here. As mentioned above, I don't think the inter-model consistency shows much.**

We thank the Referee for the reference and have included the citation at line 186.

- 3. 1225: In general the QBO will probably more or less vanish in an average of many experiments as its phase is almost random.**

We agree and is why we show the results from a single ensemble member and not an ensemble average.

- 4. Given the length of the paper it contains many figures. Perhaps a few of them could be discarded (maybe Fig. 10).**

We would prefer to keep the current figures as we believe they are useful.