General comments:
This paper presents an evaluation of ocean DMS in CMIP6 models over the historical period, and discusses their projected changes by the late 21st century under SSP585. To my knowledge, no previous work on DMS has been done using CMIP6 models. Therefore, this paper provides useful insights into the current state of DMS represented in the latest generation of ESMs. I recommend publication after major revisions, addressing my general and specific comments below.

The historical evaluation is very extensive, but maybe a bit too extensive to be included in the main text. I do not suggest to delete anything, but I do suggest to move some content into Supplementary Information (SI). One suggestion is to move Section 3.1.2 and associated figure/table (Figure 5 & Table 5) into SI. I particularly pick on this section because: (1) this section compares the models with L11 only, which is now considered to be outdated (i.e. G18 and W20 are better replacements); and (2) this section compares over small biogeographical regions, in which global models are not necessarily expected to perform well. I think Figure 3 is just sufficient for regional evaluation of these coarse-resolution models.

In addition to the environmental variables considered in the paper, I suggest to consider three additional variables for analysis: pH, MLD, and SST. Projected changes in these variables might play a substantial or additive role, as they influence directly or indirectly the DMS concentration and flux, as parameterised in the models. The Arctic might have experienced greater changes in these variables, so it is worthwhile checking these variables.

Specific comments:

L48: replace “last” with “latest”.

L49: I’m not sure if “unprecedented” is an appropriate term here, considering that: (1) Tesdal et al. (2016) have incorporated more products (measurement/empirical/prognostic approaches) in their assessment; and (2) there are only 4 ESMs in CMIP6 that simulated ocean DMS. Has this number increased/decreased from CMIP5?

Sec.2.1.1: Given the dependence of DMSP (DMS) production rate on phytoplankton species, I suggest to list the types of phytoplankton and cellular quota (sulfur to
carbon/CHL ratios) specified in all of the 4 models.

L98&L101: I’m confused about pH dependency. In L98, it says DMS release is computed as a function of pH. In L101, it says it has not been activated in CMIP6 runs. So which one is used for this paper? If it has not been activated, the word “pH” should be removed from L98.

L142: Is there plan to publish the DMS data for MIROC-ES2L on ESGF nodes in the near future?

L158: What about mixed layer depth (MLD)? Given its direct effect on DMS in diagnostic models, I think MLD should be assessed in addition to Chl.

L159: In addition to these, I suggest to show pH of the models whose DMS depends on pH. The parameterisation of Six et al. (2016) has quite strong pH effect, so this might play an important role in some regions like the Arctic. Figures can go into SI.

L163: I am not sure if I get this correct. Is MMM calculated by averaging the ensemble means of the 4 models? Or is it calculated by averaging the ensembles of the 4 models (11 + 10 + 3 + 16 for historical)? I think it is the former, but it is not clear from this sentence.

Figure 2: For readability, indicate in the caption whether these differences represent model-minus-obs or obs-minus-model.

L365: why is it “striking” that models do well in these regions?

L392: Instead of text, it might be helpful to visualise the different wind-speed-based parameterisations used by these models. Consider creating a simple plot like Figure 2 of Ho et al. (2006) (but do this for Schmidt number for DMS).


L410: In addition to wind, would temperature bias play a role in modifying flux via solubility/diffusivity? SST figures could be added to SI.

Figures 6&8: I suggest to add a subplot showing the results of Wang et al. (2020), which I assume are better obs-based products than L11/CAMS19? Without them, it just gives an impression that model are performing badly compared to the obs (L11/CAMS19). I think they compare better with Wang et al. (2020), and this point should be made clear in these figures.

L465: The 4 CMIP6 models differ in the flux parameterisation, so the finding here does not confirm the conclusion of Tesdal et al. (2016) that global emission is roughly linearly dependent upon global mean concentration for “a given flux parameterisation”.

L546: I recommend two papers from Wang et al. (2018), which incorporates perhaps more DMS producers than the 4 CMIP6 models, including Phaeocystis.


Figures 13, 14, 15: For understanding what each colour represents easily, could you plot a legend in one of the subplots? I know the colours are described in figure caption, but it is easier with a legend.

L601: I think this paragraph deserves a bit more discussion. The strong relationship between DMS and Chl/NPP is probably true for a given phytoplankton species (and therefore, this relationship holds for in situ observations of a particular phytoplankton bloom or relatively simple-complexity phytoplankton models). However, should this really be the case at global scale where different phytoplankton species dominate in different regions and phytoplankton have a wide range of DMS production rates (i.e. cellular quota; Stefels et al. 2007)? I understand that this point leads to the conclusion in the subsequent paragraph, but I think the reality of the DMS-Chl/NPP relationship is highly variable regionally due to the diversity of phytoplankton species, which should be acknowledged.

L622: Briefly state what the conclusions are.

L650: I don’t really understand the latter part of this sentence: “the specific role ... are clearly visible.” I think this latter part can be deleted, and combine the earlier part with the previous sentence, i.e. “Comparing the time series ... variables, especially when considering ... (dashed lines).”

Figure 15: DMS emissions at 100 % are indicated only for two models?

Section 5: I think this section should be named as “Discussion and Conclusions”, as it is quite extensive for just Conclusions.

L694: I understand L11 has sampling biases. However, should W20 have similar sampling biases because it also relies on the same dataset (well, twice more) for both training and evaluation (L196-204)? So unless W20 accounts for a preferential sampling of DMS-productive conditions incorporated into the dataset (L189), how can we conclude that W20 does not suffer from similar sampling biases as in L11?

L719: Briefly state what the conclusions are.

L725: I don’t think the word “overcome” is appropriate here. Overcome suggests one effect counteracts and defeats another effect. The trend of DMS concentration is neutral (neither increasing/decreasing; Figure 14 bottom panel), so it’s just that the positive trend of ice-free extent drives the trend of DMS emission.

L750: Data availability for the CMIP6 models should also be mentioned here.