# Response to reviewers for the manuscript on NorESM2, Seland et al., GMD

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-378 , 2020.

**We want to sincerely thank the reviewers for their thorough and insightful comments. Our paper has clearly benefited from this process and we appreciate the time and effort of the reviewers spent on helping document NorESM2.**

Our responses are

> **Indented and in bold.**

## Reply to anonymous Referee #1

Received and published: 16 April 2020

Review of “The Norwegian Earth System Model, NorESM2 – Evaluation of the CMIP6 DECK and historical simulations” by Seland et al.

General Comments: This manuscript documents the NorESM2 model, which is being used in CMIP6. The model is based on the CESM2 model with some notable differences in for instance, the representation of aerosols and their climate interactions; convection parameterizations as well as using different physical ocean and ocean bio-geochemistry models. A description of the model and its differences from CESM and CAM6 models are detailed and an overview of the development and tuning of the fully coupled configuration in a pre-industrial climate is given. An assessment of some key climate responses is provided, including the equilibrium climate sensitivity and transient climate response as well as future climate projections. An overview of the present-day performance of the historical simulations against observations is also provided.

This is a useful description and overview paper of the NorESM2 model and its climate characteristics and will be a very useful reference for the on-going and future CMIP6 analysis work. It is suitable for publication in GMD although I find it in its current form very long with 30 figures. The authors have included a lot in this one manuscript, covering a wide range of the model assessment with the result that the analysis of the individual components feels quite “light touch while I feel it is still missing evaluations of important parts of the fully coupled Earth system. I would recommend publication in GMD after my recommended revisions and additions to the analysis are made which I outline below.

> **We see the point that there is sometimes a mismatch between a too detailed analysis on some parts of the model and not much information on other parts. We propose to move some of the figures and relevant analysis into a supplement. Details are given in the answers to the specific comments. We have also made an attempt to include more information on the carbon cycle and aerosol parameterisations in the model but this had to be done in a way that does not take away points from the specific papers planned / or published on these topics.** Specific Comments:

There is no evaluation of aerosols (even though this is a key difference from the CESM2 model) or ocean or terrestrial biogeochemistry provided. While the land model is essentially CLM5 and documented elsewhere the authors note the important implications of the updated nitrogen-carbon limitation on the carbon cycle in the model but no assessment of the carbon cycle is provided. The ocean biogeochemistry is a totally new component compared to CESM but again no evaluation of this important component of the ES model is given. Overall given this is an overview documentation paper of NorESM2 I feel an assessment of the global carbon cycle at the very least is warranted.

**A full description and evaluation of the ocean biogeochemistry in NorESM2 has been documented in Tjiputra et al. (2020). In the revised manuscript, we have included the following paragraph summarizing the key performance.**

**"Due to the identical ocean component between NorESM2-LM and NorESM2-MM, the performance in ocean biogeochemistry is very similar in both model versions. Compared to NorESM1, the climatological interior concentrations of oxygen, nutrients, and dissolved inorganic carbon have improved considerably in NorESM2. This is mainly due to the improvement in the particulate organic carbon sinking scheme, allowing more efficient transport and remineralization of organic materials in the deep ocean. The seasonal cycle of air-sea gas exchange and biological production at extratropical regions was improved through tuning of the ecosystem parameterizations. The simulated long-term mean of sea-air $CO_2$ fluxes under the pre-industrial condition in NorESM2-LM is -0.126 ± 0.067 Pg C $yr^{-1}$. Under the transient historical simulation, the ocean carbon sink increases to 1.80 and 2.04 Pg C $yr^{-1}$ in the 1980s and 1990s, which is well within the present day estimates. A more detailed evaluation of iHAMOCC performance in NorESM2 is available in Tjiputra et al. (2020)."**

**In the land model description the following information will be added**

**"An overview of gross primary productivity (GPP) and soil and vegetation carbon pools are provided in Table 3, showing a substantially better agreement with observations for both resolutions of NorESM2 than NorESM1. There is consistency between observations and model simulations at different resolutions for GPP and vegetation carbon, whereas both NorESM2 model versions have a negative bias in soil carbon. These results broadly agree with results from offline (land only) simulations with CLM described by Lawrence et al. (2019), who also describe the individual model updates from CLM4 (used in NorESM1) to CLM5."**

Similarly for aerosols, I note the authors cite other papers that are in preparation, however again as a top-level documentation paper and the importance of aerosol-climate interactions for the climate response of the model some overview of the performance of aerosols in the model is needed. In particular, both sea salt and DMS emissions have been used to tune the final coupled model but no detail of this tuning nor impact on the aerosol simulation is provided.

**We will add two figures in the Supplementary material to document the performance of aerosols. The first figure shows a comparison of aerosol optical depth (AOD) at 550 nm between NorESM and observations for the period 2005--2014. The second figure shows global mean time series over the historical period of AOD and ERF.**

**The following will be added to the manuscript:**

**The cooling over the period 1930–1970 in NorESM2 is probably caused by the combination of a low climate sensitivity (see Sect. 4.3) and a strong negative aerosol forcing. Atmosphere-only simulations with NorESM2-LM (see Olivié et al., in prep.) show that the aerosol effective radiativeforcing (ERF) strengthens from around -0.3 W $m^{-2}$around 1930 to -1.5 W $m^{-2}$in the period 1970–1980, becoming slightly weaker again in 2014 with a value of -1.36 W $m^{-2}$. On a global scale anthropogenic $SO_2$ emissions have risen strongly in the period 1950–1980, and these are assumed to contribute most to the anthropogenic aerosol forcing. The ERF are quite similar in both model versions. We find an ERF of -1.36±0.05 W $m^{-2}$in NorESM2-LM and -1.26±0.05 W $m^{-2}$in NorESM2-MM for the year 2014**

**(compared to 1850). Figure S3b shows the time evolution of ERF for the first ensemble member ofNorESM2-LM. Given that the ERF is not an observable quantity, we have also included time series of aerosol optical depth which can be related to measurements (Fig. S3a) along with a comparison of aerosol optical depth with observations (Fig. S4).Detailed analysis of the aerosol properties is done in Olivié et al. (in prep.)**

**A more detailed description of the tuning / modifications of sea-salt and DMS is provided in the relevant sections. Details under the specific comments.**

Tuning: More detail and clarity is needed in some aspects of the tuning description. It is evident from Section 3 that a number of variables have been used to tune the low and higher resolution models (LM and MM) but the tunings differ in a number of places between the two resolutions. A table summarizing the primary and secondary tuning parameters as well as the untuned/tuned values chosen for each configuration would be very beneficial.

**A table of all the tuning parameters for both resolution versions will be included. The table will also include the corresponding values from CESM2 where applicable.**

It should include the impact of the tuned values on an appropriate metric which ideally would be constrained by observations eg: RESTROM or the SW cloud forcing in the case of the tuning of gamma parameter (paragraph beginning 266).

**Time series of RESTOM, SW and LW cloud forcing, temperatures, AMOC, sea salt and DMS emissions are now included in the supplement for both model versions.**

The tuning was carried out in a pre-industrial climate , yet the authors set their tuning targets in order “to maintain values of mean atmospheric and ocean temperatures close to observations” (L254) , given that the observations are predominantly in the present- day the authors should comment on the limitations of any such comparison. Were any present-day simulations done in parallel to validate this tuning and evaluation?

You state (Line 224) that present-day year 2000 AMIP timeslices were used for the general development of CAM6-Nor. Given that a continuous year 2000 forcing is not a realistic representation of present-day climate or of observations over recent decades, can the authors comment on the decision to use year 2000 forcing instead of a timeseries forcing and implications this may have on the model development and evaluation.

**The entire section will be rewritten in order to give a better description of the development with respect to present and pre-industrial conditions. This is included in the manuscript.**

**“Similar to CESM, NorESM2 adjusted towards its coupled climatology with an initial phase of strong cooling in the high latitudes of the northern hemisphere, after which an intensification of ocean heat advection stabilised the simulation. After that point, the climatology tended to settle to a steady drift. During major tuning steps, the coupled model had to be restarted from the initial state several times. In order to save computer resources, minor tuning, especially toward reducing RESTOM, was performed on the best-candidate simulation after this initial, large adjustment. Alongside the final tuning, the CESM components were updated to the versions found in CESM2.1. In this second phase of coupled spin-up, it was found that the sensitivity of some aspects of the simulated coupled climatology to small changes in parameters or parameterisations could be different than that found in stand-alone simulations of the individual components with prescribed boundary conditions.The**

**coupled response could be both amplified or damped with respect to single-component simulations. As a result, some of the final parameter tuning of the model had to be performed in coupled mode."**

**The main goal of the coupled tuning process was to create an energy balanced pre-industrial control simulation with a reasonably stable, adjusted equilibrium state. The simulation can produce a steady climatology only if the time average of the radiative imbalance on the top of the model vanishes. In practice, a commonly used target is to bring RESTOM within ±0.1 W m$^{-2}$ while maintaining values of mean atmospheric and ocean temperatures close to observations. To achieve this, each change in the coupled model was tested in parallel in atmosphere-only (AMIP) and ocean-only (OMIP) mode. Because ocean heat gain and tropospheric air temperature, humidity and cloudiness are strongly associated with the top of the atmosphere fluxes, improving the state in the coupled simulation, and reducing RESTOM and drift in AMIP and OMIP simulations, are closely connected goals. On the other hand, fine tuning of the coupled state should not significantly degrade important climatological variables such as temperature, precipitation, cloud, or the main mode of coupled variability, i.e. ENSO. Our parallel testing procedure ensured that the model simulation maintained a degree of consistency both with the present-day, observed climatology, and with a steady pre-industrial climate. Where available, notably in SST and sea-ice, observational estimates of the state of Earth's pre-industrial climate were also considered against the coupled integrations.**

Overall the different choice of tuning parameters for the two resolutions will impact the models evolution and therefore limits the assessment of the role of resolution on the model performance and any potential benefits of the higher resolution. Can the authors comment on this ?

**We believe that a similar and balanced pre-industrial state is more important than the tuning of a limited number of parameters, but hard to prove otherwise. Our evaluation against observations showed more realism for simulations with a higher resolution, showing its benefit. We will change the text as follows:**

**"Each tuning step was performed in isolation, and an effort was made to ensure the greatest possible similarities in the two model configurations LM and MM. No tuning was performed that attempted to target other modes of variability beside ENSO, or a particular climate response to external forcings, e.g. from changes in greenhouse gas concentration, anthropogenic aerosol emissions, or volcanic or solar forcing.**

**[...]**

**We give a concise summary of the parameters that were used for tuning NorESM2, with their final value and a comparison with CESM2, in Table 2."**

Line 250: What were the main changes from CESM2 –> CESM2.1? These should be documented, perhaps in Section 2 and here in the Tuning section document the impact of these developments on the tuning and development of NorESM2.

**We will include some information on this update. As the change was done quite early in the spin-up we do not think this had any impact on the control period:**

**"The changes from CESM2.0 to CESM2.1 are, according to Danabasoglu et al. (2020), mostly technical although with minor bug fixes and updated forcing fields. The update was done after an initial adjustment, but early in both spin-ups, approximately 1000 model years before the start of the**

**control, at both resolutions. The impact on the global fields are quite small as can be seen in the spin-up figure in the supplementary material."**

Some sort of schematic or diagram (even if included in an appendix) would be useful to aid understanding of the spin-up process, detailing offline spin-up and fully coupled spin-up. How long was the total spin-up period of the final tuned model before the official piControl for the DECK began? Later in Section 4.1 you state the abrupt 4xCO2 and 1pctCO2 runs were started at year 1 of the control -presumably you mean here the piControl but there's no indication of the full length of spin-up and how you determined that the model was fully spun-up.

**A schematic of the tuning time-line is included in the supplement.**

Ensemble size: It is very hard to assess the robustness of the NorESM2-MM model given there is only a single historical member used in this analysis. The authors need to acknowledge such limits in the text. Indeed, the historical evolution of global mean surface temperature is outside of the LM model range but its impossible to say if this is meaningful. Furthermore, it's not clear if only single ensemble members were run for the future projections in both models, it looks like this is the case but again it needs to be clearly stated and limitations on conclusions drawn need to be discussed.

**We have now updated the evaluation by including all three ensemble members of MM.**

It is very interesting how the ECS is so remarkably different in NorESM2 compared to CESM2. Can the authors expand on the detail given here? Have additional sensitivity experiments been done to pick apart the role of the differences between the two models on the response eg, ocean model, aerosol-cloud representation? It seems a very relevant investigation to understand potentially more generally the multi-model differences in ECS, in particular given the tendency of CMIP6 models to move towards higher ECS it's all the more interesting that NorESM2 has gone the other way.

**We agree with the reviewer that this is a very interesting difference. We have changed the text and included further details from the paper Gjermundsen et al. (submitted).**

**"An extensive analysis of the low ECS value in NorESM2 is given in Gjermundsen et al. (submitted, 2020). Note that the aerosol forcing is not very different between NorESM2 and CESM2 and can not explain the discrepancy in ECS values. Several sensitivity experiments have been conducted and are reported in Gjermundsen et al. in order to investigate the importance of different ice cloud schemes, CLUBB and interactive DMS. However, these NorESM2 experiments exhibit similar ECS values. The main reason for the low ECS in NorESM2 compared to CESM2 is, how the ocean models respond to GHG forcing. The behaviour of the BLOM ocean model (compared to the POP ocean model used in CESM2), contributes to a slower surface warming in NorESM2 compared to CESM2. Using the Gregory et al. (2004) method on the first 150 years leads to an ECS estimate which is considerably lower than for CESM2. However, if 500 years are included in the analysis, NorESM2 shows a sustained warming similar to CESM2. This suggests that the actual equilibrium temperature response to a large GHG forcing (the value one finds when the model is run for many hundred years) in NorESM2 and CESM2 is not very different, but that the Gregory et al. (2004) method based on the first 150 years does not give a good estimate of ECS for models."**

Line 402: What is NorESM1-Happi? Given the prevalence of its usage in the analysis of the climate response and present day performance of the NorESM2, an appropriate reference and a brief description of how it's different from NorESM1 is warranted Overall, I don't get the motivation for including NorESM1-Happi in the analysis and find it often confuses the analysis. Also, Presumably the NorESM1 models are not driven with the updated CMIP6 forcing for instance?

> **NorESM1-Happi is now defined in the beginning of section 4. We also mention the differences between NorESM1-M and NorESM1-Happi and the motivation for including both versions of NorESM1 in the present study. By including both versions we can compare against NorESM1 versions and simulations having corresponding MM and LM resolutions. We will add the following to the manuscript:**
>
> **“The motivation for including NorESM1-Happi in the present paper is to present results from a low-resolution (-M) and medium-resolution version (-Happi) of NorESM1 alongside the results from the low-resolution (-LM) and medium-resolution versions (-MM) of NorESM2.”**

The final “Summary and Discussion” reads really just solely as a Summary. Can the authors draw some overarching Conclusions from their analysis, which might include for instance the overall improved performance of the NorESM2 models compared to its predecessors; the motivation of 2 resolutions for NorESM2, role of resolution on model performance and potential benefits or applications of the different resolutions to different aspects of wider CMIP6 analysis. The would be beneficial to a reader and future potential user of NorESM2 data.

> **We agree that this should be a short section containing only the essential points of the paper; accordingly, we have renamed it “Summary and Conclusions”. We also agree that the conclusions may have been expressed a bit too implicitly to give clear recommendations on the use of the two different model versions. We will partly give this information in section 2.1 we will include explicit timing information for the two resolutions and extended the sentence “Due to the generally high computational cost of NorESM2” with “and relatively limited amount of computing power”**
>
> **We will add recommendations for the use of the two model resolutions in the summary and discussion section:**
>
> **Improvements of NorESM2 compared to NorESM1 are already mentioned in the summary and discussion, but we will complement and clarify this in the revised manuscript.**

Length of Paper: This paper is very long. It appears to me that many of the figures are surplus to the requirements of the main thread of the analysis and don’t get discussed much in the main text I would recommend moving some of them to a supplementary material. The ones that strike me are Figures 15, 21, 22, and a reduction in the number of ENSO plots (currently Figures 26-30).

> **The suggested figures 15, 21 and 22 are moved to supplementary material. We also moved four more Figures to the SM. The discussion of the MJO was shortened, summarising the main points only (a further evaluation of the MJO in NorESM will be given in a separate paper) and the respective Figures (24,25) moved to the SM. We also followed the reviewer’s suggestion with regard to ENSO by moving Figures 26, 30 to the supplement; however we left the discussion of ENSO unchanged in the text. This will be the only documentation of ENSO in this model for a while, so it is important to keep it there.**

Technical Comments:

The Title should reflect that use is made use of ScenarioMIP simulations also.

> **Included Title reads now: “Overview of The Norwegian Earth System Model, NorESM2, and Key Climate Response of CMIP6 DECK, Historical, and Scenario Simulations”.**

Line 46: participates –> participate

**Corrected.**

Line 80: It would be nice to inform the reader why the decision was made not to include land-ice model.

**We have included this additional information.**

**“Our tests with an interactive ice-sheet model over Greenland show that the model does not maintain a realistic mass balance, indicating that further development is needed. For CESM, specific tuning was carried out in order to achieve a better Greenland ice-sheet mass balance. Although NorESM2 inherited such tunings, its warmer regional climate would have required additional, dedicated effort. Due to resource limitations we have postponed this until after CMIP6.”**

Line115: Please include an appropriate reference for the prescribed optical properties used for the stratospheric aerosol.”

**Included: "Monthly distributions of stratospheric sulfate aerosols follow the CMIP6 recommendations: Concentration, surface area density, and volume density are based on the work of Thomason et al. (2018)."**

line 121: Can the authors quantify the impact of tuning both the sea salt (and later the DMS) on the total emissions of these natural sources and perhaps AOD and cloud droplet numbers? Were the aerosol tunings applied in the same way in both LM and MM?

**As described in the tuning section, the aerosol parameters are the same for both model resolutions.**

**With respect to sea-salt tuning the word tuning in this connection is imprecise since this change was done prior to the tuning process and followed the recommendations in published sea-salt articles.**

**Rewritten as:**

**The equation for sea-salt emissions has been modified by changing their dependence on 10-meter wind speed. NorESM2adopts the value recommended by Salter et al. (2015), 3.74, for the exponential factor, instead of 3.41 in NorESM1. This change was partly justified as an early tuning prior to the start of the spin-up simulations, in order to reduce the large positive top of the model radiative imbalance of the model before temperature equilibration. Even with the lower exponential factor however the model already produced excessive sea-salt aerosol optical depth (Glíß et al., 2020) and surface mass concentrations (Olivié etal., in prep.) compared to in-situ observations. Thus the change results in an even larger overestimate. Since the emission flux of oceanic primary organic aerosols is proportional to that of fine sea-salt aerosols (Kirkevåg et al., 2018), this specific change also has an impact on the natural oceanic organic matter emissions.**

**Additional text.**

**“The sea-salt emission changes were tested in a predecessor model version, NorESM1.2 (Kirkevåg et al., 2018). Annual and globally averaged, this lead to increases from 99.5 to 228.3 ng/m2/s (129%) in sea-salt emissions, from 7.8 to 17.2 mg/m2 (121%) in sea-salt column burdens, with**

**corresponding changes in total clear-sky AOD from 0.086 to 0.119 (38%), and cloud droplet numbers at top of the cloud (using the method of Kirkevåg et al., 2018) changed from 31.3 to 32.7 cm-3 (4.5%).**

**For a more detailed analysis of the impact of changes in the natural emissions please see Olivié et al. (2020).”**

Sect 2.4: does the dust aerosol impact the iron fertilisation of the ocean?

**No it doesn’t. This is specified in the text**

**“Currently the atmospheric deposition into the ocean is decoupled. The ocean biogeochemistry uses the monthly climatological aerial dust (iron) deposition of Mahowald et al. (2005).”**

**Mahowald, N., Baker, A., Bergametti, G., Brooks, N., Duce, R., Jickells, T., Kubilay, N., Prospero, J., and Tegen, I.: Atmospheric global dust cycle and iron inputs to the ocean, Global Biogeochem. Cycles, 19, 4025, https://doi.org/10.1029/2004GB002402, 2005.**

Line 224: I don’t understand what is meant by a “data-atmosphere” for the offline forcing of the ocean components – presumably the atmospheric forcing came from the CAM6-Nor simulations?

**With “forced by a data-atmosphere” we meant “forced with prescribed atmosphere and runoff of the OMIP1 protocol”, so not using forcing from CAM6-Nor. This is made clear in the text.**

Line 246: “towards its climatology” –> towards its own climatology

**Corrected**

Line 252: steady –> steady-state?

**Replaced and slightly reformulated.**

Line 266: Define the acronym CLUBB

**Defined**

Line 281: “detail here” appears to be inserted here in error. Some quantification of the change in seasalt and DMS emissions should be given.

**Included as in the comment related to “line 121”.**

Line 300: The lower biological production of DMS is said to agree better with observations but leads to an underestimation in DMS emissions. What observations are used here? How do the authors explain this apparent discrepancy?

**The ‘lower biological production here refers to the primary production (PP) through phytoplankton photosynthesis, and not the DMS production. NorESM1 simulates too strong bias in its spring bloom PP in the high latitude, and is now alleviated in NorESM2. The statement**

**“Compared to Schwinger et al. (2017), NorESM2 has doubled the diatom-mediated DMS production parameter in order to maintain the observed high DMS concentration at high latitudes. This tuning is necessary due to the lower biological production simulated in NorESM2 (relative to NorESM1), which**

**is a better representation to the observations, during spring bloom in both hemispheres (Tjiputra et al., 2019).”**

**Rephrased to**

**“Compared to Schwinger et al. (2017), in NorESM2 the parameter controlling DMS production by diatoms was doubled, which allowed to maintain high DMS concentration at high latitudes during spring and summer seasons in both hemispheres, as in observations (Lana et al., 2011). This tuning compensates for the reduced primary production simulated in NorESM2 compared to that in NorESM1 (Tjiputra et al., 2020). “**

Line 305: greenhouse gas climate scenarios → greenhouse gas future climate scenarios.

**done**

Lines 308âAT310: Please make sure you include the appropriate references here for the CMIP6 DECK and ScenarioMIP experimental protocols.

**Added**

Line 330: were there any particular criterion used for the choice of initialisation years ?

**No. The section has been made simpler only including the interval between the ensemble members.**

**“Following CMIP6 guidelines, for this experiment we carried out a small ensemble of integrations, consisting of 3 members. This helps isolate the forced signal from internal climate variability. The three model integrations of the ensemble differ only in their initial conditions, which were obtained from model states late in the spin-up at intervals of 30 model years apart. This is analogous to the historical ensemble of NorESM1 produced for CMIP5..**

Line 331: it seems a great shame that there is only one ensemble member of NorESM2-MM included in this analysis and severely limits any conclusions drawn about the performance of this model compared to the LM equivalent. If additional members can be added this would significantly add to the value of the analysis. If not, the limitations of having only 1 ensemble member should be highlighted in the text.

**The other ensemble members have been run and we have been able to update the figures and text in the revised paper with the two new HIST members of NorESM-MM. Since both resolutions have the same number of ensemble members the sentence about MM is removed here.**

Line 334: future climate development –> future climate change?

**done**

Line 338: RCPs – define acronym

**done**

Line 354: please define what you mean by a “sufficiently long spin-up”

**The sentence is deleted as it is superfluous. If the control simulation is stable enough fulfilling the requirements, the length of the spin-up is irrelevant.**

Line 371: abrupt4xCO2 –> abrupt-4xCO2

**Done**

Line 420: Does the model have a representation of nitrate aerosol?

**No. A sentence stating “Nitrate aerosols are not included.” is added to the aerosol description”**

“it is likely that the aerosol forcing is similar in both model versions” – can you actually make this statement given the different cloud tuning in the two configurations with potential implications for the marine stratocumulus for instance?

**The ERF has now been calculated so we can give the actual figures. We will add (see above) the aerosol ERF values in text (obtained from comparing two RFMIP simulations, i.e. piClim-aer and piClim-control) : we find an ERF of -1.36 ±0.05 W/m$^2$ in NorESM2-LM and -1.26 ±0.05 W/m2 in NorESM2-MM in 2014 (compared to 1850).**

Line 473 “sea level is lower” –> sea level anomaly is lower.

**Corrected**

Line 488: Fig 9 here I think should refer to Fig 20

**Referred to both (surface salinity and total precipitation)**

Line 489 Fig 10 –> Fig 11

**Corrected**

Line 550: North Atlantic –> North American continent (also biases prevalent in south America).

**We will modify this as suggested**

Line 554: “are mitigated”–> are reduced

**Done**

Figure 14: Why is the LM model so much warmer (at the surface) than MM? Is this a consequence of the different tunings? Its hard to tell from Figure 2 if the net TOA in the piControl is overall warmer in the LM model.

**The average 500 yr TOA imbalance in NorESM2-MM and LM is very similar (-0.057 W m$^{-2}$ for NorESM2-LM and -0.065 W m$^{-2}$ for NorESM2-MM). The small difference is a residual <u>at equilibration</u> (after SSTs have warmed) and is therefore not the *cause* for LM being warmer. A brief discussion is added**

**The stronger cool tropospheric and warm surface tropical bias of NorESM-LM compared with NorESM-MM is in line with the behaviour of both NorESM1 and CESM2. The systematic difference between the two atmosphere resolutions is also consistent between coupled and AMIP simulations, with CAM-Nor significantly cooler at two degree resolution than at one degree resolution for the same SSTs and the same physics parameters. At the same time, tropospheric specific humidity (and, a fortiori, relative humidity) is higher.**

**Both lead to higher corresponding RESTOM. The ultimate cause of this systematic dependence of the simulated climatology on the resolution of the atmosphere model is not known. There may be a sensitivity of the convection parameterization to the grid-scale variability of near-surface air parameters and to boundary-layer stability. Another possibility is a resolution dependence of cloud microphysics and the efficiency of stratiform precipitation. LWP and column precipitable water appear almost uniformly higher in CAM-Nor at two degree resolution than at one degree resolution.**

Line 625 “modelled cloud cover” – presumable the 70% here is referring to a global mean value. It would be helpful to refer the reader back to Table 2 here.

**Done**

Line 626: The reference to Fig 15 seems to be oddly placed in between Fig 19 and Fig 20. I suggest moving the location of the figure.

**Renumbered to figure 19. Figure 16-19 → 15-18**

Line 633: “reanalysis. along”? This whole sentence needs to be corrected as Figure 20 does not use GPCP data.

**Figure 20 does use GPCP data, but there was an error in the caption. Both the caption and the sentence (L633 in the submitted version) is corrected.**

Figure 20: Either the figure caption of the figure labelling is incorrect here in terms of what model is plotted in what panel, please correct.

**Corrected.**

Line 695: friquency –> frequency

**Corrected.**

Line 700: “NorESM-LM (Figure 25(b)) –> this is inconsistent with the Figure caption of Figure 25 which states panel b = NorESM-MM. Please double check all figure captions to make sure they are correct and consistent with the main text.

**Corrected (The figure is also moved to the supplement. )**

Figures 26 – 29 : It feels like there are a disproportionately a lot of figures for Section 5.9. Are all these needed in the main text, can some be moved to supplementary section?

**Figure 26 and 30 is moved to the supplement. As mentioned above, ENSO is a fundamental aspect of the coupled model climatology, and we do not foresee to publish another evaluation of NorESM2 in this respect. So we believe the text of the ENSO section should stay unchanged in this paper.**

Line 751: “medium-resolution version of the model” – you should clearly state here that the resolution differences here relate to the atmosphere.

**Included in the sentence atmospheric in the sentence. Now reads “ the atmospheric medium-resolution version of the model (NorESM2-MM) and a low-resolution version (NorESM2-LM).**

Figure 8: March and September lines should be clearly marked on the plots.

**Will be corrected before final publication.**

Figure 25: In the figure caption the final sentence is incomplete.

**This will be corrected.**

# Reply to anonymous Referee #2

Received and published: 15 May 2020

The paper is a high-level description of development, tuning, and key CMIP6 simulations of NorESM2. That includes a discussion of climate sensitivity and several aspects of the climatological state of the model.

The paper is half-way between an overview paper and an evaluation paper. It works well as an overview, covering the main development activities, simulations and results. I like the openness of the description of tuning strategies. Sharing components with CESM2 brings the interesting aspect of the impact of ocean/etc. on different on key metrics like sensitivity, which may provide interesting opportunities for new insights.

The paper does not work as an evaluation paper. The evaluation mostly looks at physical aspects (radiation, clouds, ocean state, sea ice, ENSO) and the more “Earth system” components are not evaluated at all. My suggestion is to focus on the overview, delegating the evaluation of specific components to companion papers. The present paper should then be re-titled and reframed, with minimal effort, as an overview paper only. This reframing would be a good opportunity to make section 5 more balanced in terms of text-to-figure ratio: many figures are only briefly mentioned in the text, so could go.

**The title and introduction have been modified to give more focus to the general overview of the model. Title reads now: “Overview of The Norwegian Earth System Model, NorESM2, and Key Climate Response of CMIP6 DECK, Historical, and Scenario Simulations”. Some of the more detailed analysis e.g. seasonal precipitation cycle and some of the ENSO aspects have been moved to a supplement. More details on which and how figures are moved are given in the answers to reviewer 1.**

**1 Specific comments**

Caption of Figure 1: The information given in parentheses could more efficiently be put in the boxes directly.

**I OVerlaying additional text to the plot panels results in less legibility; we therefore prefer to keep this Figure unchanged.**

Line 64: I'm curious to know how those modifications were chosen. In response to perceived deficiencies in CESM2? Different scientific priorities? Ad-hoc developments that happened to be ready?

**Our main initial thrust with regard to CAM dynamics and formulation was directed at improving the local and global conservation properties of the model, and, in a related way, to remove obvious model resolution dependencies, in the belief that this might also bring advantages to the fidelity of the coupled model to observations at both targeted resolutions. A posteriori, our results appear to support such belief. Scientific priorities of our own were to include the CAM Oslo aerosol scheme, allowing for coupling with marine biogeochemistry (DMS) not part of CESM along with associated further adjustments to emissions and fluxes.**

Line 99: That paragraph would be a good place to say what the time step of the different models is.

**We have added the following text:**

**"As in CAM6, a 30 minute physics timestep is used, with four-fold and eight-fold dynamics substepping for LM and MM, respectively."**

Line 99: That paragraph could be organised more efficiently. Related statements should be grouped together, for example all statements related to emissions; then chemistry; then volcanic forcing; then optical properties. Bullet points would work well here.

**The paragraph has been split up into changes in external forcings (given as bullet points) and other changes. The whole section is included here**

**The latest updates in the aerosol modules (that is, the changes between NorESM1.2 and NorESM2) are described by Olivié110et al. (in prep.). Very briefly these can be summarized as followsThe CMIP6 forcing input files now replace the corresponding CMIP5 files in NorESM2. These changes involve a large number of parameters: (i) Greenhouse gas concentrations of carbon dioxide ($CO_2$), methane ($CH_4$), nitrous oxide ($N_2O$),equivalent trichlorofluoromethane (CFC-11), and dichlorodifluoromethane (CFC-12) follow Meinshausen et al. (2017). (ii)Solar forcing is prescribed according to Matthes et al. (2017). (iii) Emissions of aerosols and aerosol precursors that are not calculated online by the model have been updated. Anthropogenic emissions of black carbon (BC), organic matter (OM),and sulfur dioxide ($SO_2$) are prescribed according to Hoesly et al. (2018), and biomass burning emission strengths follow van Marle et al. (2017) applying a vertical distribution according to Dentener et al. (2006). As in NorESM1, continuous tropospheric outgassing of $SO_2$by volcanoes is taken into account, but we have also added the tropospheric contribution of explosive volcanoes (Dentener et al., 2006). As in NorESM1, an OM/OC ratio of 1.4 is taken for fossil fuel emissions and 2.6 for biomass burning emissions, and sulphur emissions are assumed to be 97.5 % $SO_2$and 2.5 % $SO_4$. Nitrate aerosol is not included. (iv) The impact of stratospheric aerosol in NorESM1 was taken into account by prescribing volcanic aerosol mass concentrations. In NorESM2, prescribed optical properties from CMIP6 are instead used, and are integrated in the calculation of total optical parameters for use in the radiation module together with other**

aerosols. The monthly distributions of stratospheric sulfate aerosols follow now the CMIP6 recommendations (Thomason et al., 2018). (v) For oxidant concentrations (hydroxyl radical (OH), nitrate radical ($NO_3$), hydroperoxy radical ($HO_2$) and ozone ($O_3$)) needed for the description of secondary aerosol formation, we use the same fields as used in CESM2(CAM6) (Danabasoglu et al., 2020), which originate from pre-industrial control, historical, and scenario simulations of CESM2(WACCM6) (Gettelman et al., 2019b). The oxidant fields are 3-dimensional monthly varying fields, and are provided at a decadal frequency for the historical and scenario simulations(Danabasoglu et al., 2020). (vi) For ozone concentrations used in the radiative transfer calculations we also use fields from CESM2(WACCM6). They are zonally averaged 5-daily varying fields. (vii) Production rates of $H_2O$ from CH4 oxidation(mainly playing a role in the stratosphere) are also prescribed monthly climatologies based on CESM2(WACCM6) simulations,again with a decadal frequency.In NorESM2, oceanic dimethyl sulfide (DMS) emission is prognostically simulated by the ocean biogeochemistry compo-nent (Sect. 2.4), hence allowing for a direct biogeochemical climate feedback in coupled simulations. The DMS air-sea flux is simulated as a function of upper-ocean biological production following the formulation of Six and Maier-Reimer (1996) andwas first tested in the NorESM model framework by Schwinger et al. (2017). Currently the atmospheric deposition into the ocean is decoupled. The ocean biogeochemistry uses the monthly climatological aerial dust (iron) deposition of Mahowaldet al. (2005). The dust parametrisation has undergone two important changes with respect to NorESM1. First, dust emissions were effectively halved by reducing a scaling coefficient for the emission flux of prognostic dust. This brings CAM6-Nor140better in line with CAM6. Second, the assumed complex refractive index of mineral dust for wavelengths below 15μm has furthermore been changed according to more recent research (for details, see Olivié et al. (in prep.) and references therein),compared to the values applied in NorESM1.2.The aerosol nucleation formulation described by Kirkevåg et al. (2018) has been updated by allowing all pre-existing parti-cles to act as coagulation sinks for freshly nucleated particles (Sporre et al., 2019). This results in a more realistic rate of survival for these 2 nm nucleation particles into the smallest explicitly modeled mode/mixture of co-nucleated sulfate and secondary organic aerosols. In NorESM1 only the fine mode of co-nucleated sulfate and SOA (mixture no. 1) acted as a coagulation sink for the 2 nm particles. This reduces the number concentrations of fine-mode particles, while increasing their size, which in effect yields increased cloud condensation nuclei and cloud droplet concentrations. In NorESM1.2 the survival rates in the lower troposphere changed from typically 20 - 80% to 1 - 20% (zonally and annually averaged). Kuang et al. (2009) inferred survival probabilities from size distribution measurements and found that at least 80% of the nucleated particles measured at Atlanta, GA and Boulder, CO were lost by coagulation before the nucleation mode reached CCN sizes, even during days with high growth rates.The equation for sea-salt emissions has been modified by changing their dependence on 10-meter wind speed. NorESM2adopts the value recommended by Salter et al. (2015), 3.74, for the exponential factor, instead of 3.41 in NorESM1. This change was partly justified as an early tuning prior to the start of the spin-up simulations, in order to reduce the large positive top of the model radiative imbalance of the model before temperature equilibration. Even with the lower exponential factor however the model already produced excessive sea-salt aerosol optical depth (Glиß et al., 2020) and surface mass concentrations (Oliviéet al., in prep.) compared to in-situ observations. Thus the change results in an even larger overestimate. “The sea-salt emissions changes were tested in a predecessor model version, NorESM1.2 (Kirkevåg et

**al., 2018). Annual and globally averaged, this lead to increases from 99.5 to 228.3 ng$^{-2}$s$^{-1}$(129%) in sea-salt emissions, from 7.8 to 17.2 mg m$^{-2}$(121%) in sea-salt column burdens, with corresponding changes in total clear-sky AOD from 0.086 to 0.119 (38%), and cloud droplet numbers at top of the cloud (using the method of Kirkevåg et al., 2018) changed from 31.3 to 32.7 cm$^{-3}$(4.5%). Since the emission flux of oceanic primary organic aerosols is proportional to that of fine sea-salt aerosols (Kirkevåg et al., 2018), this specific change also has an impact on the natural oceanic organic matter emissions. CAM-Nor computes the effects of hygroscopic growth of aerosols on water uptake and optical properties by means of look-up tables that take relative humidity as an input. In NorESM1, the grid-point average relative humidity was used. In CAM6-Norwe instead use the mean cloud-free relative humidity, in line with CAM6 and a number of other atmospheric models (Textoret al., 2006; Kirkevåg et al., 2018; Gliß et al., 2020). The cloud-free relative humidity (RH) is calculated assuming 100 percentRH in the cloudy volume.**170**The other differences of CAM6-**

Lines 123-124: What do you mean? The model should not cool in such simulations... Do you mean improve the radiative balance of the model?

**We will rewrite this as follows:**

**This change was partly justified as an early tuning prior to the start of the spin-up simulations, in order to reduce the large positive top of the model radiative imbalance of the model before temperature equilibration.**

Line 128: Kirkevag 2018 is unclear as to what particles acted as coagulation sink in the previous version. It should be clarified here.

**In the previous version only the fine mode of co-nucleated sulfate and SOA (mixture no. 1) acted as a coagulation sink for the 2 nm particles. The text is changed accordingly.**

Line 129: “a more realistic rate” What was the previous value? How big is the change?

**Added text**

**“In NorESM1.2 the survival rates in the lower troposphere changed from typically 20 - 80% to 1 - 20% (zonally and annually averaged). Kuang et al. (2009) inferred survival probabilities from size distribution measurements and found that at least 80 % of the nucleated particles measured at Atlanta, GA and Boulder, CO were lost by coagulation before the nucleation mode reached CCN sizes, even during days with high growth rates.”**

Line 139: How is the mean cloud-free relative humidity calculated? Assuming 100% RH in the cloudy part?

**Added text**

**“The cloud-free relative humidity is calculated assuming 100 % relative humidity in the cloudy volume.”**

Line 254: Need to clarify your secondary tuning target. Was it absolute temperature of the preindustrial state, the present-day state, or present-day temperature anomalies? The latter two imply a tuning of the response.

> **We have given a fuller explanation in the text. Basically, at every coupled tuning step towards reducing TOA imbalance in the pre-industrial (PI) climate, we used parallel stand-alone atmosphere and ocean integration to validate against present-day, observed climate. In essence, we tuned TOA in coupled mode under PI forcings, and state in stand-alone mode und present-day (PD) forcings. So tuning to observations was performed on PD climate. There was no explicit tuning of the response, since we did not target a detailed PI state, but only the PI fluxes (to equilibrium under PI forcings) while trying to minimise coupled model drift; nor did we tune to PD (satellite-era) observed absolute TOA fluxes -- only, partially, the PD cloud radiative forcings -- when adjusting the AMIP/OMIP states towards observations.**

Paragraphs starting lines 270 and 275: Those two paragraphs are confusing. Which changes made it and which didn't?:

> **The paragraph starting at line 270 will be changed to.**
>
> **"Given the same gamma values and otherwise identical parameter values the RESTOM was higher in the low-resolution version of the model"**
>
> **We have also further elucidated the role of tuning gamma nd dcs in the two model version further down in this paragraph.**

Line 272: "the final parameter values" – might as well give those values here.

> **This will be added along with a reference to the tuning table in the supplementary material**

Line 281: That statement looks incomplete.

> **The whole paragraph is reformulated as follows:**
>
> **"A more effective tuning of low-cloud radiative effects was achieved by modifying air-sea fluxes of DMS. Compared to Schwinger et al. (2017), NorESM2 has doubled the diatom-mediated DMS production parameter in order to maintain the observed high DMS concentration at high latitudes. This tuning is necessary due to the lower biological production simulated in NorESM2 (relative to NorESM1), which is a better representation to the observations, during spring bloom in both hemispheres (Tjiputra et al., 2019).**

Lines 366-367: Is that drift related to the ocean temperature drift?

> **Most of the remaining drift in ocean biogeochemistry variables is likely not very dependent on the ocean temperature drift.**

Lines 379: Should cite the examples of long equilibrium studies.

> **Added Paynter et al. (2018) show results from simulations with GFDL-CM3 and GFDL-ESM2 run for more than 4000 years.**

Lines 389-391: It would be useful to show that 500-year simulation on Fig 3. Is there a change in warming rate at some point in time, or is it just a question of time to equilibrium?

**An extensive analysis of the low ECS, including time series of temperature, is given in Gjermundsen et al 2020 (submitted). There is no substantial change in the warming rate in NorESM2 (except for the first 20 years compared to the later), but the equilibrium time scale differs substantially from CESM2. The text was unclear on this point. We will change the text but leave further details to the paper Gjermundsen et al. (submitted).**

**The text has been modified**

**"An extensive analysis of the low ECS value in NorESM2 is given in Gjermundsen et al. (submitted, 2020). Note that the aerosol forcing is not very different between NorESM2 and CESM2 and can not explain the discrepancy in ECS values. Several sensitivity experiments have been conducted and are reported in Gjermundsen et al. in order to investigate the importance of different ice cloud schemes, CLUBB and interactive DMS. However, these NorESM2 experiments exhibit similar ECS values. The main reason for the low ECS in NorESM2 compared to CESM2 is, how the ocean models respond to GHG forcing. The behaviour of the BLOM ocean model (compared to the POP ocean model used in CESM2), contributes to a slower surface warming in NorESM2 compared to CESM2. Using the Gregory et al. (2004) method on the first 150 years leads to an ECS estimate which is considerably lower than for CESM2. However, if 500 years are included in the analysis, NorESM2 shows a sustained warming similar to CESM2. This suggests that the actual equilibrium temperature response to a large GHG forcing (the value one finds when the model is run for many hundred years) in NorESM2 and CESM2 is not very different, but that the Gregory et al. (2004) method based on the first 150 years does not give a good estimate of ECS for models."**

Line 396: I suppose that the slower warming in NorESM2 means that its TCR is lower than that of CESM2?

**Yes, that is correct.**

**"The TCR of both NorESM2-LM and NorESM2-MM are lower than the value of 2.0 K found for CESM2."**

Line 417: Is that really the explanation? Isn't it normally a good thing to have a low climate sensitivity when having a strong forcing?

**We agree that the mentioning of low climate sensitivity is not helpful here, and the reasons for the 50s cooling is currently under discussion: The sentence reads now: "The cooling over the period 1930–1970 in NorESM2 is probably caused by a relatively strong negative aerosol forcing.**

Line 418: Perhaps say that this is the effective radiative forcing

**Included**

Line 425: A good way to summarise the numbers in that paragraph is that the absolute temperature simulated by MM is almost 1 degree warmer than LM throughout the 1850-2100 period, but anomalies are similar.

**The sentence "Although the historical warming is slightly weaker in NorESM2-MM compared to NorESM2-LM, the warming at the end of the 21st century is rather similar in both versions of**

**NorESM2." is replaced with "The absolute temperature simulated by LM is almost 1 degree warmer than MM throughout the 1850-2100 period, but anomalies are similar."**

Line 468: Although I do not have specific comments on section 5, that section needs to focus on main results only, clearly summarising which model/model and model/obs differences are understood, which are not, and which differences affect the model response to forcing.

**We will strengthen the focus of the manuscript on the main results, and move some of the detailed analysis to the supplementary material.**

Paragraphs starting lines 436 and 444 and Figures 6-7: SSP126 looks like an outlier in a couple of these timeseries. Is that just variability among ensemble, or is there something more than that?

**Only one realisation of each scenario makes it uncertain if it is only internal variability or not.**

Figure 15 should be re-numbered, as it is used after Figure 19.

**done**

**2 Technical comments**

Line 15: Satisfactorily -> satisfactory

**done**

Line 47: Delete "Also"

**done**

Line 793: Typo "properties"

**done**

~~Graff~~

# ~~The~~ Overview of the Norwegian Earth System Model, NorESM2~~-~~ ~~Evaluation~~ , and Key Climate Response of ~~the~~ CMIP6 DECK, Historical, and ~~historical simulations~~Scenario Simulations

Øyvind Seland[1], Mats Bentsen[2], Dirk Olivié[1], Thomas Toniazzo[2,6], Ada Gjermundsen[1], Lise Seland Graff[1], Jens Boldingh Debernard[1], Alok Kumar Gupta[2], Yanchun He[3], Alf Kirkevåg[1], Jörg Schwinger[2], Jerry Tjiputra[2], Kjetil Schanke Aas[4], Ingo Bethke[5], Yuanchao Fan[2], Jan Griesfeller[1], Alf Grini[1], Chuncheng Guo[2], Mehmet Ilicak[7,2], Inger Helene Hafsahl Karset[4], Oskar Landgren[1], Johan Liakka[3], Kine Onsum Moseid[1], Aleksi Nummelin[2], Clemens Spensberger[5], Hui Tang[4], Zhongshi Zhang[2], Christoph Heinze[5], Trond Iversen[1,4], and Michael Schulz[1,4]

[1]Norwegian Meteorological Institute, Oslo, Norway
[2]NORCE Norwegian Research Centre AS and Bjerknes Centre for Climate Research, Bergen, Norway
[3]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[4]University of Oslo, Department of Geosciences, Oslo, Norway
[5]Geophysical Institute, University in Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[6]Stockholm University, Department of Meteorology, Stockholm, Sweden
[7]Eurasia Institute of Earth Sciences, Istanbul Technical University, Istanbul, Turkey

**Correspondence:** Øyvind Seland (oyvindse@met.no)

**Abstract.** The second version of the ~~fully~~ coupled Norwegian Earth System Model (NorESM2) is presented and evaluated. NorESM2 is based on the second version of the Community Earth System Model (CESM2), ~~but has~~ and shares with CESM2 the computer code infrastructure and many earth system model components. However, NorESM2 employs entirely different ocean and ocean biogeochemistry models~~; a new module for aerosols in the atmosphere model along with aerosol-radiation-cloud interactions and changes related to the moist energy formulation, deep convection scheme and~~ . The atmosphere component of NorESM2, CAM-Nor, includes a different module for aerosol physics and chemistry, including interactions with cloud and radiation; additionally, CAM-Nor includes improvements in the formulation of local dry and moist energy conservation, in local and global angular momentum conservation~~; modified albedo and~~ , and in the computations for deep convection and air-sea ~~turbulent flux calculations; and minor changes~~ fluxes. The surface components of NorESM2 have minor changes in the albedo calculations and to land and ~~sea ice models. We show results from low (∼2) and medium (∼1) atmosphere-land resolution versions of~~ sea-ice models.

We present results from simulations with NorESM2 that ~~have both been used to carry out simulations~~ were carried out for the sixth phase of the Coupled Model Intercomparison Project (CMIP6). Two versions of the model are used, one with lower (∼2°) atmosphere/land resolution, and one with medium (∼1°) atmosphere/land resolution. The stability of the pre-industrial climate and the sensitivity of the model to abrupt and gradual quadrupling of $CO_2$ is assessed, along with the ability of the model to simulate the historical climate under the CMIP6 forcings. ~~As compared~~ Compared to observations and reanalyses, NorESM2 represents an improvement over previous versions of NorESM in most aspects. NorESM2 ~~is~~ appears less sensitive to greenhouse gas forcing than its predecessors, with an estimated equilibrium climate sensitivity of 2.5 K in both resolutions on a 150 year ~~frame~~time frame; however, this estimate increases with the time window and the climate sensitivity at equilibration is much higher. We also consider the model response to future scenarios as defined by selected shared socioeconomic pathways (SSPs) from the Scenario Model Intercomparison Project defined under CMIP6. Under the four scenarios SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, the warming in the period 2090–2099 compared to 1850–1879 reaches 1.3, 2.2, 3.0, and 3.9 K in NorESM2-LM, and 1.3, 2.1, 3.1, and 3.9 K in NorESM–MM, robustly similar in both resolutions. NorESM2-LM shows a rather ~~satisfactorily~~ satisfactory evolution of recent ~~sea ice~~ sea-ice area. In NorESM2-LM an ~~ice free~~ ice-free Arctic Ocean is only avoided in the SSP1-2.6 scenario.

## 1 Introduction

The Norwegian Earth System Model version 2 (NorESM2) is the second generation of the coupled Earth System Model (ESM) developed by the Norwegian Climate Center~~(NCC)~~, and is the successor of NorESM1 (**????**) which ~~has been~~ was used in the 5th phase of the Coupled Model Intercomparison Project (CMIP5; **?**), and for ~~evaluation of the difference between a~~ the evaluation of potential climate impacts between the 1.5 and 2 °C ~~warmer world than pre-industrial~~ warming targets of "the 21st Conference of Parties" (COP21) (**?**). NorESM2 is based on the Community Earth System Model CESM2.1 ~~(**?**)~~(**?**). Although large parts of NorESM are similar to CESM, there are several important differences~~:~~. NorESM uses the ~~isopycnic coordinate~~ Bergen Layered Ocean Model (BLOM; Bentsen et al., in prep.) ~~,~~ coupled with the iHAMOCC model for ocean biogeochemistry (**?**). It also uses a different ~~aerosol module OsloAero6 (~~atmospheric aerosol module (OsloAero6; **?**, **?**; Olivié et al., in prep.)~~, contains~~. Additionally, NorESM2 features specific modifications and tunings of the physics and dynamics of the atmosphere component (~~**?**, **??**~~**?**, **?**; Toniazzo et al., in prep.)~~, and contains the iHAMOCC model to describe ocean biogeochemistry (**?**)~~.

Many changes have contributed to the development of NorESM1 into NorESM2. The model has benefited from the evolution of the parent model Community Climate System Model version 4 (CCSM4.0) into CESM2.1, comprising the change of the atmosphere component from CAM4 to CAM6 (**??**, see also the supplementary information in **?**, **?**), the land component from ~~CLM4~~ Community Land Model (CLM)4 to CLM5 (**?**), and the sea ice component from ~~CICE4~~ Community Ice CodE (CICE)4 to CICE5 (**?**). Also, ~~specific~~ NorESM-specific developments have been implemented in the description of aerosols and their coupling to clouds and radiation (**?**), in addition to harmonizing the implementation of the aerosol scheme with the standard aerosol schemes in CESM. To extend the capabilities of NorESM as an ~~Earth System Model~~ESM, a strong focus has been put on the interactive description of natural emissions of aerosols and their precursors, and tightening the coupling between the different ~~Earth System~~ earth system components. Finally, the ocean model (Bentsen et al., in prep.) and the ocean biogeochemistry module ~~(**??**)~~ (**??**) have been further developed.

This manuscript gives a description of NorESM2 ~~,~~ and a basic evaluation against observations of the simulation of the atmosphere, sea ice, and ocean in a small set of baseline long-duration experiments with the new model. It focuses on such aspects as the simulated climatology, ~~its~~ stability and internal variability, and also on ~~its~~ the response under historical and enhanced-greenhouse gas scenario forcings.

Currently, NorESM2 exists in three versions. The two versions presented here are NorESM2-LM and NorESM2-MM: they differ in the horizontal resolution of the atmosphere and land component (approximately 2° for LM and 1° in MM), but share the same horizontal resolution of 1° for the ocean and sea ice components. These versions are otherwise identical, except for a very limited number of parameter settings in the atmosphere component, and the parameterisation used to diagnose the fraction of ~~ice-clouds~~ice clouds. A third version of the model is the $CO_2$-emission driven NorESM2-LME (as opposed to concentration driven), which can be used for interactive carbon-cycle studies, but is identical to NorESM2-LM in all other aspects.

A range of climate models and model versions ~~participates~~ participate in the sixth phase of the Coupled Model Intercomparison Project (CMIP6; **?**). ~~Also~~ NorESM2 has been used to contribute to CMIP6, and all the data generated by the participating models, including NorESM2, can be downloaded from the CMIP6 multi-model data archive.

An overview of the model which highlights the differences since previous versions and from CESM2 is given in Sect. **??**, and a short summary of model initialization and tuning is presented in Sect. **??**. A short description of the CMIP6 experiments considered in this paper is provided in Sect. **??** along with results documenting model stability, climate sensitivity, and the time evolution of selected climate variables during the historical period and future scenarios. Section **??** documents the climatological mean state of the model and atmospheric circulation patterns, with emphasis on ocean temperatures, salinity, Sea Level Anomalies (SLA; Sect. **??**), sea ice (Sect. **??**), atmospheric temperature and zonal winds (Sect. **??**), extratropical storm tracks (Sect. **??**), precipitation and the fresh water cycle (Sect. **??**), Northern Hemisphere blocking (Sect. **??**), the Madden-Julian Oscillation (Sect. **??**), and the El Niño Southern Oscillation (ENSO; Sect. **??**). A summary and ~~discussion~~ conclusion is provided at the end in Sect. **??**.

## 2 From CESM2 and NorESM1 to NorESM2: description and updates

As described in the introduction, NorESM2 is built on the structure and many of the components of CESM2 ~~.1 (?)~~(**?**), but with several modifications. ~~The development work described in this section was based on a slightly older version, CESM2, but updated to CESM2.1 during the tuning phase of the model (Sect. ??).~~ An overview of the model components can be found in Fig. **??**.

Compared to the Community Atmosphere Model version 6 (CAM6; **?**) of CESM2, the atmospheric component of NorESM2, CAM6-Nor, incorporates a number of modifications. These involve the independently developed module for the life-cycle of particulate aerosols, and the representation of aerosol-radiation-cloud interactions (**??**); changes in the moist convection scheme and the local moist energy formulation (Toniazzo et al., in prep.); global conservation of rotational momentum ~~(?)~~(**?**); and an updated ~~parameterisation~~ parametrisation of the surface flux layer for the computation of air-sea fluxes. (The last two of these modifications have recently been included in the CESM CAM6 code repositories and are available as namelist options.) A summary of these changes is given in the atmospheric model section (Sect. **??**).

The ocean model BLOM is an updated version of the elaborated Miami Isopycnic Coordinate Ocean Model (MICOM) used in NorESM1 (?). BLOM is coupled to the isopycnic coordinate Hamburg Ocean Carbon Cycle Model (iHAMOCC; ?) (iHAMOCC; ?), an updated version of the carbon-cycle model found in NorESM1 (?). Brief descriptions of the ocean and ocean biochemistry models are given in Sect. **??** and **??**.

The sea ice model, version 5.1.2 of the Los Alamos Sea Ice Model (CICE5.1.2; ?), and the land-model, the Community Land Model version 5 (CLM5; ?), only differ from the versions used in CESM2.1 by minor changes which are summarised in Sect. **??** and **??**. The river model is the Model for Scale Adaptive River Transport (MOSART; ?) and is identical to the version found in CESM2.1 and hence is not described here. The coupler structure is retained as in CESM2.1 but with changes in flux and albedo calculations summarized below.

The interactive land-ice component (the Community Ice Sheet Model; CISM; ?) and ocean surface wave components included in CESM2 (the Community Ice Sheet Model; CISM; ?) is were not activated in NorESM2 at this timefor the CMIP6 model integrations. Our tests with an interactive ice-sheet model over Greenland show that the model does not maintain a realistic mass balance, indicating that further development is needed. For CESM, specific tuning was carried out in order to achieve a better Greenland ice-sheet mass balance. Although NorESM2 inherited such tunings, its warmer regional climate would have required additional, dedicated effort. Due to resource limitations we have postponed this until after CMIP6.

### 2.1 Model versions and the coupled model system

Due to the generally In view of the comparatively high computational cost of NorESM2the model, two different versions have been set up with different CPU time demands per simulated year. In these, the atmospheric of NorESM2 with different computational cost are presented. The two versions differ by the horizontal resolution of the atmosphere and land componentshave different horizontal resolutions, one with nominal 1 which is considered medium . The "medium-resolution" (M) resolution and another one with nominal 2version has a grid spacing of $1.25^\circ \times 0.9375^\circ$ which is considered low in these components, like CESM2 (?). The "low-resolution" (L) resolutionversion uses half that resolution in the atmosphere and land components. The ocean and sea ice components are run with medium "medium" (M) ($1^\circ$) resolution in both versions. To facilitate distinguishing between the different resolutions when discussing set-up and results, a two letter suffix is added to the NorESM2, "LM" for low-resolution atmosphere/land and medium resolution ocean/sea ice and "MM" for medium resolution of both atmosphere/land and ocean/sea ice. Both versions use the “low-top” version of CAM6, with 32 layers in the vertical and model top at 3.6 hPa (40 km). NorESM2-LM is used for most of the CMIP6 simulations, while NorESM2-MM is only used for a limited number of experiments.

### 2.2 Atmospheric Atmosphere model, CAM6-Nor

The atmospheric atmosphere model component of NorESM2 is built on the CAM6 version from CESM2.1, but with using the hydrostatic finite-volume dynamical core on a regular latitude-longitude grid at the two horizontal resolutions mentioned above. In the vertical, both versions use the same discretisation as CAM6, with 32 hybrid-pressure layers and a "rigid" lid at 3.6 hPa (40 km). As in CAM6, a 30 minute physics timestep is used, with four-fold and eight-fold dynamics substepping for

LM and MM, respectively. CAM6-Nor employs parametrisations for particulate aerosols and the for aerosol-radiation-cloud interaction parameterisation interactions from NorESM1 and NorESM1.2 as described by **??**. NorESM2-specific changes to model physics and dynamics which are not aerosol related, are described by ? ? and Toniazzo et al. (in prep.).

The latest updates in the aerosol modules (that is, the changes between NorESM1.2 and NorESM2) are described by Olivié et al. (in prep.). Very briefly these can be summarized as follows :

The CMIP6 forcing input files now replace the corresponding CMIP5 files in NorESM2. These changes involve a large number of parameters: (i) Greenhouse gas concentrations of carbon dioxide ($CO_2$), methane ($CH_4$), nitrous oxide ($N_2O$), equivalent trichlorofluoromethane (CFC-11), and dichlorodifluoromethane (CFC-12) follow **?**, and solar . (ii) Solar forcing is prescribed according to **?**. Emissions (iii) Emissions of aerosols and aerosol precursors that are not calculated online by the model have been updated. Anthropogenic emissions of black carbon (BC), organic matter (OM), and sulfur dioxide ($SO_2$) are prescribed according to **?**, and biomass burning emission strengths follow **?** applying a vertical distribution according to **?**. As in NorESM1, continuous tropospheric outgassing of $SO_2$ by volcanoes is taken into account, but we have also added the tropospheric contribution of explosive volcanoes (**?**). As in NorESM1, an OM/OC ratio of 1.4 is taken for fossil fuel emissions and 2.6 for biomass burning emissions, and sulphur emissions are assumed to be 97.5 % $SO_2$ and 2.5 % $SO_4$. For oxidant concentrationsused to describe the formation of secondary aerosols , for ozone (ONitrate aerosol is not included. (iv) The impact of stratospheric aerosol in NorESM1 was taken into account by prescribing volcanic aerosol mass concentrations. In NorESM2, prescribed optical properties from CMIP6 are instead used, and are integrated in the calculation of total optical parameters for use in the radiation module together with other aerosols. The monthly distributions of stratospheric sulfate aerosols follow now the CMIP6 recommendations (**?**). (v) For oxidant concentrations (hydroxyl radical (OH), nitrate radical ($NO_3$)concentrations used in the radiative transfer calculations, and for H, hydroperoxy radical ($HO_2$O production rates due to $CH_4$ oxidation in the stratosphere) and ozone ($O_3$)) needed for the description of secondary aerosol formation, we use the same fields as used in CESM2(CAM6) (Danabasoglu et al., 2019)(**?**), which originate from a pre-industrial controland three historical, historical, and scenario simulations of CESM2(WACCM6) (**?**). The oxidant fields (hydroxyl radical (OH), nitrate radical ($NO_3$), hydroperoxy radical ($HO_2$) and $O_3$) and $H_2O$ emission rates are 3-dimensional monthly varying fields, and the $O_3$ fields are provided at a decadal frequency for the historical and scenario simulations (**?**). (vi) For ozone concentrations used in the radiation radiative transfer calculations we also use fields from CESM2(WACCM6). They are zonally averaged 5-daily varying fields; in the historical simulations these fields are provided at . (vii) Production rates of $H_2O$ from $CH_4$ oxidation (mainly playing a role in the stratosphere) are also prescribed monthly climatologies based on CESM2(WACCM6) simulations, again with a decadal frequency(Danabasoglu et al. , submitted). The impact of stratospheric aerosol in NorESM1 was taken into account by prescribing volcanic aerosol mass concentrations.

In NorESM2, prescribed optical properties from CMIP6 are instead used, and are integrated in the calculation of total optical parameters for use in the radiation module together with other aerosols. The NorESM2, oceanic dimethyl sulfide (DMS) emission is prognostically simulated by the ocean biogeochemistry component (Sect. **??**), hence allowing for a direct biogeochemical climate feedback in coupled simulations. The DMS air-sea flux is simulated as a function of upper-ocean biological production following the formulation of **?** and was first tested in the NorESM model framework by **?**. Currently the atmospheric deposition into the ocean is decoupled. The ocean biogeochemistry uses the monthly climatological aerial dust (iron) deposition of **?**. The dust parametrisation has undergone two important changes with respect to NorESM1. First, dust emissions were effectively halved by reducing a scaling coefficient for the emission flux of prognostic dust. This brings CAM6-Nor better in line with CAM6. Second, the assumed complex refractive index of mineral dust for wavelengths below 15 $\mu$m has furthermore been changed according to more recent research (for details, see Olivié et al. (in prep.) and references therein), compared to the values applied in NorESM1.2.

There have also been some changes in parameterisations and aerosol-specific tunings. Due to exaggerated extinction by mineral dust in dust-dominated regions in the previous model version, a scaling coefficient in the emission flux for interactive dust emissions has been reset to the original CAM6 value, in effect halving the emissions. Sea-salt emissions have been tuned up by changing the wind speed (at 10 m height, U10) dependency to the recommended value by **?**, now being proportional to $U10^{3.74}$ instead of $U10^{3.41}$. This has been done as a measure to help cool the model sufficiently in the spin-up and control simulation in spite of already exaggerated sea-salt aerosol optical depth (Gliss et al., in prep.) and surface mass concentrations (Olivié et al., in prep.) compared to in-situ measurements. Since the emission flux of oceanic primary organic aerosols is proportional to that of fine sea-salt aerosols (**?**), this specific change also has an impact on the natural oceanic organic matter emissions.

The aerosol nucleation formulation described by **?** has been updated by allowing all pre-existing particles to act as coagulation sinks for freshly nucleated particles (**?**)to give . This results in a more realistic rate of survival for these 2 nm nucleation particles into the smallest explicitly modeled mode/mixture of co-nucleated sulfate and secondary organic aerosols. In NorESM1 only the fine mode of co-nucleated sulfate and SOA (mixture no. 1) acted as a coagulation sink for the 2 nm particles. This reduces the number concentrations of fine-mode particles, while increasing their size, which in effect yields increased cloud condensation nuclei and cloud droplet concentrations.

In In NorESM1.2 the survival rates in the lower troposphere changed from typically 20 - 80% to 1 - 20% (zonally and annually averaged). **?** inferred survival probabilities from size distribution measurements and found that at least 80% of the nucleated particles measured at Atlanta, GA and Boulder, CO were lost by coagulation before the nucleation mode reached CCN sizes, even during days with high growth rates.

The equation for sea-salt emissions has been modified by changing their dependence on 10-meter wind speed. NorESM2 , oceanic dimethyl sulfide (DMS)emission is prognostically simulated by the ocean biogeochemistry component (Sect. **??**), hence allowing for a direct biogeochemical climate feedback in coupled simulations. The DMS air-sea flux is simulated as a function of upper-ocean biological production following the formulation of **?** and was first tested in adopts the value recommended by **?**, 3.74, for the exponential factor, instead of 3.41 in NorESM1. This change was partly justified as an early tuning prior to the start of the NorESM model framework by **?**.spin-up simulations, in order to reduce the large positive top of the model radiative imbalance of the model before temperature equilibration. Even with the lower exponential factor however the model already produced excessive sea-salt aerosol optical depth (**?**) and surface mass concentrations (Olivié et al., in prep.) compared to in-situ observations. Thus the change results in an even larger overestimate. “The sea-salt emission changes were tested in a predecessor model version, NorESM1.2 (Kirkevåg et al., 2018). Annual and globally averaged, this lead to increases from 99.5 to 228.3 ng $^{-2}$ s$^{-1}$(129%) in sea-salt emissions, from 7.8 to 17.2 mg m $^{-2}$(121%) in sea-salt column burdens, with corresponding changes in total clear-sky AOD from 0.086 to 0.119 (38%), and cloud droplet numbers at top of the cloud (using the method of Kirkevåg et al., 2018) changed from 31.3 to 32.7 cm$^{-3}$ (4.5%). Since the emission flux of oceanic primary organic aerosols is proportional to that of fine sea-salt aerosols (?), this specific change also has an impact on the natural oceanic organic matter emissions.

~~While hygroscopic swelling of aerosols in earlier versions always used the grid averaged relative humidity as input to look-up tables which take into account~~ CAM-Nor computes the effects of hygroscopic growth of aerosols on water uptake and optical properties ~~, in~~ by means of look-up tables that take relative humidity as an input. In NorESM1, the grid-point average relative humidity was used. In CAM6-Nor we instead use the mean cloud-free relative humidity, ~~as in the host model~~ in line with CAM6 and a number of other atmospheric models (**?**, **?**; **?**, **?**; ~~Gliss et al., in prep.~~ )**?**, **?**). The cloud-free relative humidity (RH) is calculated assuming 100 percent RH in the cloudy volume.

The other differences of CAM6-Nor relative to CAM6 are summarised as follows. A correction to the zonal wind increments due to the **?** dynamical core is introduced in order to achieve global conservation of atmospheric angular momentum along the Earth's axis of rotation, as described and discussed in ~~(?)~~**?**. The local energy update of the model is also modified by including a missing term (the hydrostatic pressure work) related with changes in atmospheric water vapour and thus ~~achieve~~ achieves better local energy conservation. Finally, a set of modifications to the deep convection scheme is introduced which eliminate most of the resolution dependence of the scheme, and mitigate the cold tropospheric bias of CAM6. The energy and convection changes (which are not available in the CAM6 code repository) are described in Toniazzo et al. (in prep.).

### 2.3 Ocean model

The ocean component BLOM is based on the version of MICOM used in NorESM1 and shares the use of near-isopycnic interior layers and variable density layers in the surface well-mixed boundary layer. The dynamical core is also very similar but with notable differences in physical parameterisations and coupling. For vertical shear-induced mixing a second-order turbulence closure (**??**) using a one equation closure within the family of $k-\varepsilon$ models has replaced a parameterisation using the local gradient Richardson number according to **?**. Parameterised eddy-induced transport is modified to more closely follow the **?** parameterisation with the main impact of increased upper ocean stratification and reduced mixed layer depths. As for NorESM1-MICOM, the estimation of diffusivity for eddy-induced transport and isopycnic eddy diffusion of tracers is based on the **?** implementation of **?** with their diagnostic equation for the eddy length scale, but modified to give a spatially smoother and generally reduced diffusivity. Hourly exchange of state and flux variables with other components is now used compared to daily ocean coupling in NorESM1. The sub-diurnal coupling allows for the parameterisation of additional upper ocean mixing processes. Representation of mixed layer processes is modified to work well with the higher frequency coupling and in general to mitigate a deep mixed layer bias found in NorESM1 simulations. The penetration profile of shortwave radiation is modified, leading to a shallower absorption in NorESM2 compared to NorESM1. With respect to coupling to the sea ice model, BLOM and CICE now use a consistent salinity dependent seawater freezing temperature (**?**). Selective damping of external inertia–gravity waves in shallow regions is enabled to mitigate an issue with unphysical oceanic variability in high latitude shelf regions, causing excessive sea ice formation due to breakup and ridging in CMIP5 versions of NorESM1.

For the CMIP6 contribution, BLOM uses identical parameters and configuration in coupled ocean-sea ice OMIP (Ocean Model Intercomparison Project; **?**) experiments and fully coupled NorESM2-LM and NorESM2-MM experiments, except for sea surface salinity restoring in OMIP experiments. As for NorESM1, 53 model layers are used with two non-isopycnic surface layers and the same layer reference potential densities for the layers below. A tripolar grid is used instead of the bipolar grid in CMIP5 versions of NorESM1, allowing for approximately a doubling of the model time step. At the equator the grid resolution is $1^\circ$ zonally and $1/4^\circ$ meridionally, gradually approaching more isotropic grid cells at higher latitudes. The model bathymetry is found by averaging the S2004 (**?**) data points contained in each model grid cell with additional editing of sills and passages to their actual depths. The metric scale factors are edited to the realistic width of the Strait of Gibraltar so that strong velocity shears can be formed, enabling realistic mixing of Mediterranean water entering the Atlantic Ocean.

OMIP provides protocols for two different forcing datasets, OMIP1 (**?**) and OMIP2 (**?**). **?** is a model intercomparison evaluating OMIP1 and OMIP2 experiments, including BLOM/CICE of NorESM2. Further details on the BLOM model and its performance in OMIP coupled ocean-sea ice simulations can be found in Bentsen et al. (in prep.).

### 2.4 Ocean biogeochemistry

The ocean biogeochemistry component iHAMOCC (isopycnic coordinate HAMburg Ocean Carbon Cycle model) is an updated version of the ocean biogeochemistry module used in NorESM1. The model includes prognostic inorganic carbon chemistry following **?**. A Nutrient Phytoplankton Zooplankton Detritus (NPZD) type ecosystem model (**?**) represents the lower trophic biological productivity in the upper ocean. The updated version includes riverine inputs of biogeochemical constituents to the coastal ocean. Atmospheric nitrogen deposition is prescribed according to the data provided by CMIP6. The parameterisations of the particulate organic carbon sinking scheme, dissolved iron sources and sinks, nitrogen fixation, and other nutrient cycling have been updated as well. NorESM2 also simulates preformed and natural inorganic carbon tracers, which can be used to facilitate a more detailed diagnostic of interior ocean biogeochemical dynamics. Due to the identical ocean component between NorESM2-LM and NorESM2-MM, the performance in ocean biogeochemistry is very similar in both model versions. Compared to NorESM1, the climatological interior concentrations of oxygen, nutrients, and dissolved inorganic carbon have improved considerably in NorESM2. This is mainly due to the improvement in the particulate organic carbon sinking scheme, allowing more efficient transport and remineralization of organic materials in the deep ocean. The seasonal cycle of air-sea gas exchange and biological production at extratropical regions was improved through tuning of the ecosystem parameterizations. The simulated long-term mean of sea-air $CO_2$ fluxes under the pre-industrial condition in NorESM2-LM is -0.126 ± 0.067 Pg C $yr^{-1}$. Under the transient historical simulation, the ocean carbon sink increases to 1.80 and 2.04 Pg C $yr^{-1}$ in the 1980s and 1990s, which is well within the present day estimates. Details on the updates and improvements of the ocean biogeochemical component of NorESM2 are provided in ~~?~~**?**.

## 2.5 Sea ice

The sea ice model component is based upon version 5.1.2 of the CICE sea ice model of **?**. A NorESM2-specific change is including feature however is to include the effect of wind drift of snow into ocean following **?**, as described in Bentsen et al. (in prep).

The CICE model uses a prognostic ice thickness distribution (ITD) with five thickness categories. The standard CICE elastic–viscous–plastic (EVP) rheology is used for ice dynamics (**?**). The model uses mushy-layer thermodynamics with prognostic sea ice salinity from **?**. Radiation is calculated using the Delta-Eddington scheme of **?**, with melt ponds modeled on level, undeformed ice, as in **?**.

CICE uses is discretised on the same horizontal grid as the ocean model (Sect. **??**), and is configured with 8 layers of ice and 3 layers of snow.

## 2.6 Land

The NorESM2 land model is CLM5 (**?**) with one minor modification described below. A general description of the model will therefore not be presented here. It should however be noted that CLM5 has a new treatment of nitrogen-carbon limitation, which is very important for the carbon cycle in NorESM2 and has increased the land carbon uptake substantially relative to NorESM1 (**?**). An overview of gross primary productivity (GPP) and soil and vegetation carbon pools are provided in Table **??**, showing a substantially better agreement with observations for both resolutions of NorESM2 than NorESM1. There is consistency between observations and model simulations at different resolutions for GPP and vegetation carbon, whereas both NorESM2 versions produce a negative bias in soil carbon. These results broadly agree with results from offline (land only) simulations with CLM described by **?**, who also describe the individual model updates from CLM4 (used in NorESM1) to CLM5.

In NorESM2, one specific modification was made to the surface water treatment in CLM. The surface water pool is a new feature replacing the wetland land unit in earlier versions of CLM (introduced in CLM4.5). This water pool does not have a frozen state, but is added to the snow-pack when frozen. To avoid water being looped between surface water and snow during alternating cold and warm periods, we remove infiltration excess water as runoff if the temperature of the surface water pool is below freezing. This was done to mitigate a positive snow bias and an artificial snow depth increase found in some Arctic locations during melting conditions.

## 2.7 Coupler

The state and flux exchanges between model components and software infrastructure for configuring, building and execution of model experiments is handled by the CESM2 coupler Common Infrastructure for Modeling the Earth (CIME; **?**)(CIME; **?**). The coupler computes the turbulent air-sea fluxes of heat and momentum and in NorESM2 this is implemented as a version of the COARE-3 (**?**) scheme, replacing the calculation based on **?** in CESM2. State and flux exchanges via the coupler between atmosphere, land and sea ice components occur half-hourly, aligned with the atmosphere time step, while the ocean exchanges with the coupler every hour. CIME also provides common utility functions and among these are estimation of solar zenith angle. In NorESM2, this utility function is modified with associated changes in atmosphere, land and sea ice components, ensuring that all albedo calculations use zenith angle averaged over the components time-step instead of instantaneous angles.

## 3 NorESM2 initialisation and tuning

Most of the general development of the model as described in Sect. **??** was tested in ~~stand-alone versions of the different model components,~~ configurations with reduced number of interactive components. CAM6-Nor ~~in present-day AMIP-mode under year 2000 conditions and~~ was tuned in Atmospheric Model Intercomparison Project (AMIP) configuration with mean climatological radiative forcings and boundary conditions (sea-surface temperatures – hereafter "SST"s – and sea-ice) derived from observations over the period 1990-2010. Similarly, BLOM and iHAMOCC ~~forced by a data-atmosphere. The main targets~~ were primarily tuned with prescribed atmosphere and runoff forcing of the OMIP1 protocol. The scope of these separate experiments ~~were~~ was to test improved representations of the physical processes in the simulations, ~~to mitigate~~ with the twin aims of mitigating model systematic biases when compared to the observed climate, and to ~~reduce the residual radiative~~ achieve a net radiative flux imbalance at the top of the model atmosphere (hereafter RESTOM~~) given prescribed SSTsfrom observations~~; defined as positive inward, i.e. warming the climate) more in line with satellite-based estimates, given the observed SSTs.

The first coupled version of NorESM2 included all changes described in Sect. **??**. This version was heavily tested in a pre-industrial setting (as defined in Sect. 4).

This initial version of the coupled model was initialized using a hybrid of observational estimates and earlier model simulations. The ocean model was initialised with zero velocities and temperature and salinity fields from the Polar science center Hydrographic Climatology (PHC) 3.0 (updated from **?**). Following the OMIP protocol (**?**), the nutrients (phosphate, nitrate, and silicate) and oxygen fields in NorESM2 were initialized with the gridded climatological fields of the World Ocean Atlas database (**??**). For dissolved inorganic carbon and total alkalinity, we used the pre-industrial and climatological values from the Global Ocean Data Analysis Project (GLODAPv2) database (**?**). Other biogeochemical tracers are initialized using values close to zero. CAM and CLM were initialized using the files included in the CESM2 release. Aerosols and aerosol precursors were initialised to near zero values. As there were no low-resolution pre-industrial initial files for the land model available this was replaced by an interpolation of the 1° initial file from CESM2. At a later stage in the coupled spin-up, the land surface fields were re-initialised from a long (approximately 1400 years) stand-alone CLM spin-up simulation driven by ~~a repeat 50-yrs climatology fields of~~ repeating 50 years of coupling exchange fields obtained from the earlier coupled run.

~~While preparing the coupled model for the spin-up, it was found that the sensitivities of important climatological variables, including RESTOM, to changes in parameterisations were often different in the coupled configuration compared to stand-alone simulations with the individual components using prescribed boundary conditions. The coupled response could be both amplified or damped with respect to single-component simulations. As a result, tuning test simulations had to be performed in coupled mode and the model had to be restarted from the initial state several times.~~ Similar to CESM, NorESM2 adjusted towards its own coupled climatology with an initial phase of strong cooling in the high latitudes of the northern hemisphere, after which an intensification of ocean heat advection stabilised the simulation. After that point, the climatology tended to settle to a steady drifttowards a steady-state. During major tuning steps, the coupled model had to be restarted from the initial state several times. In order to save computer resources, minor tuning, especially toward balanced towards reducing RESTOM, was performed during this second stage of the spin-up phase of the modelon the best-candidate simulation after this initial, large adjustment. Alongside the final tuning, the CESM components were updated to the versions found in CESM2.1. The changes from CESM2.0 to CESM2.1 are mostly technical but also include minor bug fixes and updated forcing fields (?). The update was done after an initial adjustment, but early in both spin-ups, approximately 1000 model years before the start of the control, at both resolutions. The impact on the global fields are quite small as can be seen in Fig. S1 and Fig. S2 in the supplement. In this second phase of coupled spin-up, it was found that the sensitivity of some aspects of the simulated coupled climatology to small changes in parameters or parametrisations could be different than that found in stand-alone simulations of the individual components with prescribed boundary conditions. The coupled response could be both amplified or damped with respect to single-component simulations. As a result, some of the final parameter tuning of the model had to be performed in coupled mode. No tuning was performed during the pre-industrial control simulation as described in Sect. **??**

The main goal of the coupled tuning process was to create a reasonably stable an energy balanced pre-industrial control simulation with a reasonably stable, adjusted equilibrium state. The simulation can produce a steady climatology only if the time-average radiative imbalance on the top of the model (RESTOM) vanishes. In practice, a commonly used target is for RESTOM to be to bring RESTOM to within $\pm 0.1\ \mathrm{W\,m^{-2}}$ . Secondary tuning targets are to obtain and maintain while maintaining values of mean atmospheric and ocean temperatures close to observations. As the ocean heat again reflects To achieve this, each change in the coupled model was tested in parallel in atmosphere-only (AMIP) and ocean-only (OMIP) mode. As ocean heat gain and tropospheric air temperature, humidity and cloudiness are strongly associated with the fluxes at the top of the atmosphereimbalance, the two requirements are strongly connected . One additional constraint was that the tuning , improving the state in the coupled simulation, and reducing RESTOM and drift in AMIP and OMIP simulations, are closely connected goals. On the other hand, fine tuning of the coupled state should not significantly degrade other important climatological variables such as temperature, precipitation, cloud, and clouds, or the main mode of coupled variability, i.e. the El-Niño Southern Oscillation (ENSO). Our parallel testing procedure ensured that the model simulation maintained a degree of consistency both with the present-day, observed climatology, and with a steady pre-industrial climate. Where available, notably in SST and sea-ice, observational estimates of the state of Earth's pre-industrial climate were also considered against the coupled integrations. Each tuning step was performed in isolation, and an effort was made to ensure the greatest possible similarities in the two model configurations LM and MM. No tuning was performed that attempted to target other modes of variability beside ENSO, or a particular climate response to external forcings, e.g. from changes in greenhouse gas concentration, anthropogenic aerosol emissions, or volcanic or solar forcing.

As found in Similar to CESM2 (?), also (?), NorESM2 had development of tended to develop excessive sea ice cover in the Labrador Sea (LS) region, although the temporal development in NorESM2 differed from CESM2. For any tested combination of parameter choices, NorESM2 developed excessive LS sea ice cover starting around year 60 after model initialisation. This was however only a temporary model state and in all experiments the sea ice returned close to observed state in the LS region after additional 60–80 model years of simulation.

One of the most common methods to tune RESTOM is to change the amount and thickness of low clouds. The main parameter used for tuning the low clouds in the CLUBB Cloud Layers Unified By Binormals (CLUBB) scheme is the "gamma" parameter, which controls the skewness of the assumed Gaussian PDF for subgrid probability density function for subgrid vertical velocities. A low gamma implies weaker entrainment at the top of the clouds, in particular for marine stratocumulus. This increases the amount of low clouds and results in a higher short-wave cloud forcing.

Given the same gamma values, the RESTOM was higher in the low resolution version of the model. In addition the sensitivity to the change of the gamma parameter was different in the two model resolutions, so a different choice of gamma was needed for the two resolutions. The final parameter values are well within the gamma range of 0.1–0.5 tested by **?**. The resulting bias in , although smaller than the values used in CESM2 at the same resolution. A small gamma pushes up short-wave cloud radiative forcing (SWCF), which led to a high bias in SWCF in NorESM2-LM. This bias was somewhat off-set by regulating the parameter dcs (autoconversion size threshold for cloud ice to snow)in NorESM2-LM but this had only , with a small impact on the tropospheric temperature bias. Changing dcs in NorESM2-MM did not improve the overall skill of this model version compared to the initial value so was not used for this versions

While the amount of change in SWCF could be estimated by running the atmosphere and land model in a stand-alone configuration, the change in RESTOM in coupled set-up was small compared to the change in cloud forcing. Further attempts at reducing positive RESTOM by tuning the boundary layer stability were neutralised by SST adjustment, while worsening the tropospheric cold bias. A more effective tuning of low cloud radiative effects was achieved by modifying air-sea fluxes of sea salt and DMSdetail here. As described in Sect. **??** the disadvantage of increasing the sea-salt flux, however, is that this resulted in too dominant marine aerosols with respect to optical thickness and surface mass concentrations. DMS. Compared to **?**, the parameter controlling DMS production by diatoms was doubled in NorESM2, which allowed to maintain high DMS concentration at high latitudes during spring and summer seasons in both hemispheres, as in observations (**?**). This tuning compensates for the reduced primary production simulated in NorESM2 compared to that in NorESM1 (**?**).

RESTOM was decisively reduced, both in stand-alone (AMIP) and in coupled simulations (before SST adjustment), by increasing outgoing long-wave radiation. This was achieved in three ways. First, alterations were made to the **?** convection scheme, as described in Toniazzo et al. (in prep.), aimed at increasing mid- and high-altitude latent heating of the atmosphere . Second, higher sea-surface temperatures were achieved by reverting to the NorESM1 level of ocean background vertical mixing after having used up to 50% higher diffusivity for the purpose of reducing upper ocean biases. Thirdfor a given amount of precipitation. Second, positive cloud radiative forcing in the terrestrial radiation spectrum was reduced by intervening on the parameterisation of ice cloud fraction. Finally, higher sea-surface temperatures in coupled simulations were achieved by reducing the value of the parameter controlling background vertical mixing in the ocean back to that used in NorESM1. Initial optimisation in stand-alone configurations had led to increase the value of this parameter by about 50%.

Several versions A remarkable sensitivity of the model climatology on the parametrisation of the ice cloud fraction parameterisation are provided (as namelist options) was found. This purely empirical part of the cloud parametrisation of CESM2 is rather ad-hoc and poorly constrained by observations. Several namelist-controlled options for ice-cloud fraction are provided in CESM. Initial tuning of the parameters of the CESM2 default option appeared promising, but coupled adjustment again tended to neutralise the effect on model radiative imbalance. ~~An~~ In NorESM2-LM, an effective reduction in the high- and mid-level cloud cover could only be achieved by switching ~~parameterisation in NorESM2-LM, such that~~ to a different parameterisation option, in which there is no direct functional dependence of ice cloud fraction on environmental relative humidity (this is option number 4 in CESM). By contrast, in NorESM2-MM the CESM default scheme (option number 5, with explicit RH dependence) could be ~~tuned sufficiently in NorESM2-MM, by including a minor modification that narrows~~ modified by allowing a continuous narrowing of the range of cloud sensitivity to environmental RH~~(and thus provides~~. This modification thus constitutes a continuous switch between the two ~~parameterisations). This purely empirical part of the cloud parametrisation of CESM2 is very poorly constrained by observations, and its~~ parameterisation options. A target for future development might be to represent ice clouds in a way better rooted in physical processes.

~~Compared to **?**, NorESM2 has doubled the diatom-mediated DMS production parameter in order to maintain the observed high DMS concentration at high latitudes. This tuning is necessary due to the lower biological production simulated in~~ We give a concise summary of the parameters that were used for tuning NorESM2~~(relative to NorESM1), which is a better representation to the observations, during spring bloom in both hemispheres (?)~~, with their final value and a comparison with CESM2, in Table **??**.

## 4 Control simulations and model response to forcing

This section presents a basic description of the climatology simulated in CMIP6 experiments with the two versions of the model, NorESM2-LM and NorESM2-MM (Sect. **??**). We consider the time evolution of temperature in historical and enhanced greenhouse gas future climate scenarios, along with aspects of the ocean circulation and sea ice. We validate the historical coupled simulations against observational estimates and reanalyses, and compare them with results from simulations with previous versions of NorESM (Sect. **??**): NorESM1-M (**???**) used in CMIP5 and NorESM1-Happi (**?**) used for HAPPI (Half a degree Additional warming Prognosis and Projected Impacts; **?**) and a set of CMIP5 experiments carried out for model evaluation purposes (**?**). NorESM1-Happi is an upgraded version of NorESM1-M with differences including doubled horizontal resolution in the atmosphere and land components (1°in NorESM1-Happi and 2°in NorESM1-M) and improved treatment of sea ice. The motivation for including NorESM1-Happi in the present paper is to present results from a low-resolution (-M) and medium-resolution version (-Happi) of NorESM1 alongside the results from the low-resolution (-LM) and medium-resolution versions (-MM) of NorESM2.

We consider three sets of experiments that are important for documentation and application of CMIP6 models: the DECK experiments (**?**), the CMIP6 Historical experiment (**?**), and the Tier 1 experiments of the ScenarioMIP (**?**). A brief description of the set-up of these experiments is given in Sect. **??**.

The analysis is divided into three parts. Section **??** focuses on the stability of the pre-industrial control simulation. In Sect. **??**, we consider the simulated climate sensitivity to abrupt and gradual quadrupling of $CO_2$. A brief analysis of the warming, sea ice, AMOC, and the transport through the Drake Passage in the historical simulations and the scenarios is given in Sect. **??**.

### 4.1 Experiment set-up

As described by **?**, a set of common experiments known as DECK (Diagnostic, Evaluation, and Characterization of Klima) has been defined to better coordinate different model intercomparisons and provide continuity for model development and model progress studies. The DECK consists of the following four baseline experiments: (1) the Historical Atmospheric Model Intercomparison Project (AMIP) experiment; (2) the pre-industrial control (piControl) experiment defined by estimated forcings from 1850, started from initial conditions obtained from a spin-up with the same, constant forcings during which the coupled model climatology stabilises towards a stationary statistics; (3) the experiment corresponding to the piControl, but where an experiment otherwise identical to piControl, except that the $CO_2$ concentrations are instantaneously quadrupled at the start of the run (abrupt-4x$CO_2$set to four times the piControl concentrations, from piControl initial conditions (abrupt-4xCO2); (4) the experiment corresponding to the an experiment otherwise identical to piControl, but where the $CO_2$ concentrations are gradually increased by 1% per year (1pctCO$_2$starting from piControl concentrations and initial conditions (1pctCO2). Both abrupt-4xCO2 and 1pctCO2 were started from year 1 of the control.

The DECK was run produced with both versions of the model (NorESM2-LM and NorESM2-MM) and we here here we consider results from the pre-industrial control and the abrupt-4x$CO_2$ and 1pctCO$_2$ abrupt-4xCO2 and 1pctCO2 (Sect. **??**–**??**). As this paper focuses on the coupled aspect of NorESM2, the AMIP runs are not included here, but are described in Olivié et al. (in prep.) and Toniazzo et al. (in prep.).

Another qualifying experiment required for entry in CMIP6, and important for model evaluation with respect to observations, is the historical experimentwhich is run with forcings from . In this experiment, time-dependent forcings are specified to reflect observational estimates valid for the so-called historical period, defined as viz. 1850–2014. For the low-resolution version of the model (NorESM2-LM), we have Following CMIP6 guidelines, for this experiment we carried out a small ensemble of integrations, consisting of 3 members. The first ensemble member was initialised using initial conditionsfrom the first year of the control experiment, while members number 2 and 3 are initialised from years 32 and 62 respectively. For NorESM2-MM, only a single ensemble member had been carried out when this paper was written. Consistent with historical member 1 from NorESM2-LM, the NorESM2-MM historical experiment was started from identical initial conditions to the NorESM2-MM control simulationThis helps isolate the forced signal from internal climate variability. The three model integrations of the ensemble differ only in their initial conditions, which were obtained from model states late in the spin-up at intervals of 30 model years apart. This is analogous to the historical ensemble of NorESM1 produced for CMIP5.

One Beyond the DECK, one of the most important applications for Earth system models ESMs is to provide estimates for future climate developmentof future climate change. This is typically done using scenarios where critical input for climate models through description and quantification of both which specify future anthropogenic forcing of the climate that include changes in land-use change and amount of climate forcing agents in the atmosphereis provided(such as deforestation) and the addition of greenhouse gases and other pollutants to the atmosphere. The latter can be ~~described either as changes in emissions or concentrations~~prescribed either directly as atmospheric concentrations (as a function of time), or as time-evolving in emissions into the atmosphere (which then interact with ocean and land biogeochemical processes before yielding atmospheric concentrations). The design of scenarios are based on a combination of socioeconomic and technological development, named the Shared Socioeconomic Pathways (SSPs), with future climate radiative forcing (RF) pathways, Representative Concentration Pathways. (RCPs) in a scenario matrix architecture (**?**).

The simulations included ~~here~~ in this paper are the Tier 1 experiments of the ScenarioMIP (**?**): SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. The forcing fields for all the experiments are generally the same as used in CESM2.1. This includes solar forcing, prescribed oxidants used for describing secondary aerosol formation, greenhouse gas concentrations, stratospheric $H_2O$ production from $CH_4$ oxidation, ozone used in the radiative transfer calculations, and land-use. While the experiments in this paper use prescribed greenhouse gas concentrations, NorESM2 can also be run with $CO_2$ emissions as described by ~~?~~**?**.

~~Files for~~ NorESM2 lacks a physical representation of the stratosphere; instead, appropriate upper-boundary conditions need to be specified. Accordingly, stratospheric aerosols and emissions of aerosols and aerosol precursors were ~~created~~ prescribed based on the ~~input found at~~ data provided by the input4mips website: https://esgf-node.llnl.gov/projects/input4mips/. In addition, sulphur from tropospheric volcanoes was included similarly to **?**, see Sect **??**.

### 4.2 ~~Stability of the~~ Simulated control ~~climate~~climatology and residual drift

After ~~the tuning period and the~~ tuning and an initial spin-up, both NorESM2-LM and NorESM2-MM were integrated for 500 years with steady pre-industrial forcings to produce the piControl experiments. Below, we present a basic analysis of the general state and drift of important parameters ~~in the system~~of the simulated climatology.

During the control integration the forcings as well as the parameter choices were kept constant. ~~Given a sufficiently long spin-up there~~ There should be no long-term drift in the model state variables or ~~fluxes. In practice any residual trends in~~ their partial tendencies (hence, a fortiori, in radiative fluxes). More precisely, any residual drift of the simulated control climatology should be negligibly small compared with the signal resulting from the response to changes in climate forcings as prescribed in the historical~~and~~ , enhanced-greenhouse gas~~experiments. A~~ , and scenario experiments. In practice, a reasonable target is to maintain ~~RESTOM~~ the RESTOM of piControl within $\pm 0.1\,\mathrm{W\,m^{-2}}$ in the time mean. Any small imbalance in RESTOM is typically reflected in a small trend in ocean temperature. A time-series of the Atlantic Meridional Overturning Circulation (AMOC) can give an indication of the stability of the general ocean circulation.

Figure **??** shows time-series of related global-means in the piControl simulations from NorESM2-LM and NorESM2-MM. As can be seen in the figure the drift is generally small and comparable for the two model versions. The top-of-the-atmosphere radiative imbalance is -0.057 $\mathrm{W\,m^{-2}}$ for NorESM2-LM and -0.065 $\mathrm{W\,m^{-2}}$ for NorESM2-MM. The ocean volume temperature change of 0.03 K over 500 years is much smaller than the rate of warming observed during the last 50 years. Similarly, there are positive trends in global mean ocean salinity of $2.6 \times 10^{-5}$ g kg$^{-1}$ and $4.7 \times 10^{-5}$ g kg$^{-1}$over 500 years for NorESM2-LM and NorESM2-MM, respectively, that we consider small since for NorESM2-MM this is equivalent to an average surface freshwater loss of $2.9 \times 10^{-5}$ mm day$^{-1}$. The remaining trends are not significantly different from 0 ~~on a 5% level t-test~~at the

95% confidence level, as estimated from a $t$-test. We found however a slight decrease in DMS sea-to-air flux of 2% over the 500 year control period, reflecting a residual drift in ocean bio-geochemistrybiogeochemistry. AMOC variations are reasonably small and show no significant trend.

### 4.3 Equilibrium climate sensitivity and transient response

The two enhanced greenhouse gas experiments of the CMIP-DECK aim to facilitate a comparison of climate change in response to a standardized specified forcing across different models. The corresponding NorESM2 simulations were started at the same nominal model year and with the same initial conditions as piControl (and consequently assigned the same notional model year). They are referred to as abrupt4×$CO_2$ and 1pctCO$_2$abrupt-4xCO2 and 1pctCO2.

Figure ?? shows the time evolution of near-surface temperature for abrupt4×$CO_2$, 1pctCO$_2$ abrupt-4xCO2, 1pctCO2 and piControl for both model configurations. Three commonly used metrics for the response to $CO_2$ forcing, based on the evolution of the simulated global-mean temperature, are the Equilibrium Climate Sensitivity (ECS), the Transient Climate Response (TCR), and the Transient Climate Response to cumulative $CO_2$ Emissions (TCRE). Their values are given in tableTable ?? for the NorESM2 experiments, and compared to those for NorESM1. The ECS is defined as the change in global near-surface temperature when a new climate equilibrium is obtained with an atmospheric $CO_2$ concentration that is doubled compared to the pre-industrial amount. In order to reach a new equilibrium, a model simulation of several thousand years is required (Boer and Yu, 2003)(?). There are some examples in the literature of models for which this has been done , but in general ECS is more commonly e.g. ? show results from simulations with GFDL-CM3 and GFDL-ESM2 run for more than 4000 years. Given certain assumptions, ECS may be estimated from the relationship between surface temperature and RESTOM from the abrupt4×$CO_2$ abrupt-4xCO2 experiment using the so-called Gregory method (?). The numbers in tableThis estimate has become a standard in CMIP6. The figures reported in Table ?? are calculated using years 1–150 from the simulations shown in Fig. ??, and are divided by 2 to get the number for $CO_2$ doubling instead of quadrupling. The ECS is 2.54 K for NorESM2-LM, which is slightly lower than the equivalent value for NorESM1 of 2.8 K. Both are significantly lower than the CMIP5 mean value of 3.2 K but well inside the bounds of the likely range of 1.5–4.5 K (?). On the other hand, the ECS in NorESM2 is markedly smaller than the ECS found in CESM2 of 5.3 K by ??, despite sharing many of the same component models. The ECS An extensive analysis of the low ECS value in NorESM2 is discussed in more detail in Gjermundsen et al. (in prep.). There are indications that the different given in ?. Note that the aerosol forcing is not very different between NorESM2 and CESM2 and can not explain the discrepancy in ECS values. Several sensitivity experiments have been conducted and are reported in ? in order to investigate the importance of different ice cloud schemes, CLUBB and interactive DMS. However, these NorESM2 experiments exhibit similar ECS values. The main reason for the low ECS in NorESM2 compared to CESM2 is how the ocean models respond to GHG forcing. The behaviour of the BLOM ocean model (compared to the POP ocean model used in CESM2), contributes to a delayed warming in the first 150 years of abrupt-4xCO2 in slower surface warming in NorESM2 compared to CESM2. Using the Gregory et al. (2004) method on that period ? method on the first 150 years leads to an ECS estimate which is considerably lower than for CESM2. However, after the initial slow warming in the abrupt-4xCO2 experimentif 500 years are included in the analysis, NorESM2 shows a sustained warming similar to CESM2, when the

~~abrupt-4xCO2 experiment is extended to 500 years or longer~~. This suggests that the actual ~~ECS~~ equilibrium temperature response to a large GHG forcing (the value one finds when the model is run for ~~thousands of yearsuntil equilibrium~~many hundred years) in NorESM2 and CESM2 is not very different, but that the **?** method based on the first 150 years ~~only~~ does not give a good estimate of ECS for models.

The ~~TCR~~ transient climate sensivity (TCR) is defined as the global-mean surface temperature change at the time of $CO_2$ doubling, and accordingly it was calculated from the temperature difference between the ~~1pct$CO_2$~~ 1pctCO2 experiment averaged over years 60–80 after initialisation and piControl. The TCR is 1.48 K and 1.33 K for NorESM2-LM and NorESM2-MM, respectively. As for ECS, these values fall in the lower part of the distribution obtained from the CMIP5 ensemble (**?**), similar to those obtained for NorESM1. The TCR of both NorESM2-LM and NorESM2-MM are lower than the value of 2.0 K found for CESM2 (**?**). A recent observational estimate for the 90 % likelihood range of TCR is 1.2–2.4 K (**?**).

We also give an estimate of the transient climate response to cumulative carbon emissions (TCRE) calculated from TCR and the corresponding diagnosed carbon emissions. Following **?**, TCRE is defined as the ratio of TCR to accumulated $CO_2$ emissions in units of K $EgC^{-1}$. As $CO_2$ fluxes were not calculated in NorESM1-M and NorESM1-Happi, the NorESM1 values are obtained from the carbon cycle version of NorESM1 (NorESM1-ME; **?**). TCRE is reduced from 1.93 K $EgC^{-1}$ in NorESM1-ME to 1.36 K $EgC^{-1}$ and 1.21 K $EgC^{-1}$ in NorESM2-LM and MM, respectively. Since TCR is comparable, the main difference is due to changes in carbon uptake. NorESM1, with CLM4 as the land component, had a very strong nitrogen limitation on land carbon uptake. This limitation is weaker in CLM5 (**?**) used in NorESM2.

### 4.4 Climate evolution in historical and scenario experiments

In this section we provide a very brief analysis of the response of the model to historical forcings in the three historical members carried out ~~with~~ in both NorESM2-LM and ~~the one realisation carried out with~~ NorESM2-MM. We also consider the model response for the Tier 1 experiments from ScenarioMIP (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5). The focus here will be on the response in global-mean near-surface temperature, the Atlantic Meridional Overturning Circulation (AMOC), the volume transport through the Drake Passage, and on sea ice area.

Figure **??** shows the time evolution of the surface atmospheric temperature in the historical simulations from NorESM2-LM and NorESM2-MM along with observations. Both versions of NorESM2 follow the observations rather closely for the first 80 years. After 1930 the model displays somewhat weaker warming than the observations until around 1970. After that the rate of the warming in the models are similar to that seen in the observations. The cooling over the period 1930–1970 in NorESM2 is probably caused by the combination of a low climate sensitivity (see Sect. **??**) and a strong negative aerosol forcing. Atmosphere-only simulations with NorESM2-LM (see Olivié et al., in prep.) show that the aerosol ~~forcing~~ effective radiative forcing (ERF) strengthens from around -0.3 $W\,m^{-2}$ around 1930 to -1.5 $W\,m^{-2}$ in the period 1970–1980, becoming slightly weaker again in 2014 with a value of ~~-1.2~~ -1.36 $W\,m^{-2}$. On a global scale anthropogenic $SO_2$ emissions have risen strongly in the period 1950–1980, and these are assumed to contribute most to the anthropogenic aerosol forcing. ~~Although no such experiments have been done with~~ The ERF are quite similar in both model versions. We find an ERF of -1.36 ±0.05 $W\,m^{-2}$ in NorESM2-LM and -1.26 ±0.05 $W\,m^{-2}$ in NorESM2-MM ~~yet, it is likely that the aerosol forcing is similar in~~

both model versionsfor the year 2014 (compared to 1850). Figure S3b shows the time evolution of ERF for the first ensemble member of NorESM2-LM. Given that the ERF is not an observable quantity, we have also included time series of aerosol optical depth which can be related to measurements (Fig. S3a) along with a comparison of aerosol optical depth with observations (Fig. S4). Detailed analysis of the aerosol properties is done in Olivié et al. (in prep.). Note also that our choice of the reference period for temperature anomaly computation (1850-1880) enhances the NorESM2 negative bias with respect to observations in the last half of the 20th century.

Figure **??** shows again the evolution of the surface air temperature in the historical simulations (only the first ensemble member for NorESM2-LM), followed by the temperature evolution under the four SSP scenarios for NorESM2-LM and NorESM2-MM. Compared to the 1850–1879 period, the model shows a warming in 2005–2014 of 0.72 and 0.54 K for NorESM2-LM and NorESM2-MM, respectively. Under the four scenarios SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, the warming in the period 2090–2099 compared to 1850–1879 reaches 1.30, 2.15, 2.95, and 3.94 K in NorESM2-LM, and 1.33, 2.08, 3.06, and 3.89 K in NorESM-MM. Although the historical warming is slightly weaker in NorESM2-MM compared to NorESM2-LM, the warming at the end of the 21st century is rather similarin both versions of NorESM2The absolute temperature simulated by LM is almost 1 degree warmer than MM throughout the 1850-2100 period, but anomalies are similar. For SSP1-2.6, the temperature stabilizes in the second half of the 21st century. In NorESM1, under the scenarios RCP2.6, RCP4.5, and RCP8.5, the surface air temperature in the period 2071–2100 was 0.94, 1.65, and 3.07 K higher than in 1976–2005 (**?**). For the same periods and looking at SSP1-2.6, SSP2-4.5, and SSP5-8.5, we find rather similar (but slightly stronger) warmings of 1.06, 1.81, and 3.22 K in NorESM2-LM, and 1.11, 1.83, and 3.26 K in NorESM2-MM.

The simulated Atlantic Meridional Overturning Circulation (AMOC) at 26.5° N shows a multi-centennial variability that is 15% of the mean in the control simulation (Fig. **??**). In the historical simulations the AMOC peaks for both MM and LM in the 1990's at around 24 Sv before starting a rapid decline at around year 2000 (Fig. **??**). In both versions the AMOC reaches a quasi-equilibrium by the end of the century at around 15-10 Sv depending on the scenario. Since we only have a few ensemble members, it remains unclear how fast the AMOC declines in response to the greenhouse gas forcing and which part of e.g. the initial decline is due to the multi-decadal variability. In any case, it is noteworthy that the initial AMOC decline begins already during the historical period in both versions, which is also consistent with the NorESM2 and multimodel mean response to the OMIP2 forcing (1958-2018, **?**).

In addition to the AMOC, also the Antarctic Circumpolar Current (ACC) strength, as measured in the Drake Passage, shows multi-centennial variability that is about 3% of the mean (Fig. **??**). Similar variability in the ACC has been linked to convection within the Weddell and Ross seas in the CMIP5 ensemble (**?**). Also in our simulations the Weddel Sea convection has similar long term variability as the ACC. Unlike the AMOC, there is no clear trend emerging from the scenario simulations, but rather the multidecadal variability continues throughout the 21st century. Again, a larger number of ensemble members could help identify the forced signal.

The time evolution of Northern Hemisphere sea ice area (March and September) through the historical and scenario periods is shown in Fig. **??**. Both model versions are compared with the sea ice area from OSISAF (OSI-V2.0) reprocessed climate data record (**?**) for the years 1979–2019. The total sea ice area from NorESM2-LM compares rather well with the observations, while NorESM2-MM has too much ice, especially during summer. The trend in sea ice area found in the observations during summer is also rather well captured by NorESM2-LM, while this trend is too small in NorESM2-MM. Both models have a reasonable March sea ice area compared to observations. However, the negative trends in winter sea ice area are small compared to observed trends.

During the scenario period both models show a strong reduction in summer sea ice area. The Arctic Ocean is often considered ice free when the total sea ice area drops below 1 million square km. This threshold is denoted by dotted gray lines in Fig. **??**. NorESM2-LM loses summer ice shortly after year 2050. This occurs first in the SSP5-8.5 scenario, but also the SSP2-4.5 ensemble shows values close to this threshold even before 2050. SSP3-7.0 scenarios become ice free at around 2070. Any prediction of which year the Arctic Ocean first becomes ice free must therefore be considered rather uncertain due to forcing evolution uncertainty and internal variability. This is consistent with the overall assessment of sea ice evolution in CMIP6 assessed by the ~~SIMIP Community (?)~~**?**. In NorESM2-LM an ice free Arctic Ocean is only avoided in the SSP1-2.6 scenario. NorESM2-MM loses ice slower and shows the first ice free summer around 2070. In that model, also the SSP2-4.5 scenario keeps the ice area above 1 mill. square km all years before 2100. However, the SSP1-2.6 scenario stabilizes at a sea ice area comparable with present day observations, even with SSP1-2.6 warming levels present. Therefore, the sea ice area simulated by NorESM2-MM for the future Arctic seems to be unrealistically high.

## 5 Climatological mean state and circulation patterns compared to observations and NorESM1

### 5.1 Ocean state

In the surface ocean, the large-scale climatological biases are similar in the two NorESM2 versions (Fig. **??**), but overall the MM version is closer to the observations (smaller global-mean root-mean-square error; RMSE; $\sqrt{A^2}$ in Fig. **??**). In general the Southern Ocean is too warm (Fig. **??**b-c), the Atlantic (and the Arctic) are too saline, but the Pacific is too fresh (Fig. **??**e-f). The sea level anomaly is lower than observed in the Atlantic basin, but higher in the Indo-Pacific basin and thus the gradient between the two basins is larger than in the observations (Fig. **??**h-i). If we remove the global-mean biases, the two versions produce even more similar mean errors, suggesting that some of the regional biases are largely independent of the atmosphere and land resolution.

Indeed, the regional patterns are common to many other models with coarse resolution ocean components (**?**). Both NorESM2 versions are too warm and (relatively) saline over the western boundary currents (the Gulf Stream and the Kuroshio in the Northern Hemisphere and the Brazil current and the Agulhas current in the Southern Hemisphere) and over the major eastern boundary upwelling systems (Canary, Benguela, Humboldt, and California). The biases over the western boundary currents are due to the errors in the location of the currents, which are linked to the ocean model resolution (**???**). The ocean-model resolution also explains two well known biases in the North Atlantic also seen in NorESM2: the southern bias in the Gulf Stream/North Atlantic current path causes the cold (and fresh) bias in the subpolar North Atlantic (**???**), while the lack of the Labrador Current waters on the east coast of North America causes a large warm and saline bias there (**?**).

While the above mentioned biases are mostly linked to the ocean model, in the Pacific there are biases that are not present in the ocean-only simulations (not shown). Specifically, a fresh bias over the Southern Pacific subtropical gyre and cold biases over the northern Pacific subtropical gyre and the equatorial Pacific.

The fresh bias in the Southern Pacific (Fig. **??**) is linked to the co-located positive net precipitation bias ~~(Fig. **??**)~~ as shown in Fig. **??** and extends throughout the surface mixed layer (Fig.~~**??**~~ **??**). The salinity bias also causes a negative density bias (not shown) as it is not fully compensated by temperature, supporting an atmospheric origin. A comparison with the OMIP1 and OMIP2 simulations shows that the net precipitation bias in the LM simulation, 250 mm $\text{year}^{-1}$ in the mean over the region where the salinity bias is larger than 1 g $\text{kg}^{-1}$, would be large enough to cause the simulated salinity bias (assuming mixed layer depth of 100 m and a residence time of 10 years). Therefore, we suggest that the net precipitation bias leads to accumulation of excess freshwater that is spread throughout the subtropical gyre by the ocean circulation.

Most of the large-scale surface biases are also visible in the subsurface (Figs. **??**–**??**). The upper ocean is too warm and fresh, while the deep ocean is too cold and saline. The biases are again larger in the LM version. The cold deep ocean is due to the cold bias in the Antarctic bottom water, while the warm bias in the mid-depth Atlantic (between 500–3500 m) is due to the Antarctic Intermediate Water and the North Atlantic deep water being too warm. There are also subsurface biases without a large surface signature. The Mediterranean outflow and the Red Sea outflow form too warm and saline cores visible at around 20° N and 1000 m depth in the Atlantic and Indian oceans (respectively, Figs. **??**–**??**). These biases are stronger in the LM version and not visible or much less pronounced in the OMIP simulations (not shown), which suggests that they are due to biases in the surface heat and freshwater budgets in these semi-enclosed basins. In addition, there is a strong cold and fresh (warm and saline) bias in the Pacific (Atlantic) centered around 15° S and 200–400 m depth. These anomalies are likely linked to the biases in the tropical upwelling and the resulting thermocline depth that is too shallow (deep) in the Pacific (Atlantic).

Overall, many of such sub-surface ocean biases are similar in the ocean-only simulations and may be linked to coarse ocean resolution and shortcomings in parameterised processes. In some regions, air-sea coupling tends to act to reinforce biases that may be generated in either atmosphere or ocean model components separately. The biases over the upwelling systems for example have generally a complex cause rooted in both local (including mesoscale) and remote (including equatorial) biases in both atmosphere and ocean model components (**???**). For NorESM2 the biases in the coupled simulations have a similar pattern as, but approximately twice the magnitude of the biases in the OMIP simulations (not shown). The cold bias in the northern subtropical Pacific has a contribution from weak oceanic mixing as there is a large warm bias just below the surface (Fig. **??**), but may be amplified by increased atmospheric stability and correspondingly enhanced boundary-layer clouds. Excessively negative short-wave cloud forcing is seen in that region, in contrast to AMIP simulations which show no such regional bias. In the central and eastern equatorial Pacific NorESM2 displays a characteristic "cold tongue" bias with cold SSTs and easterly wind stress bias. An equatorial easterly bias is present in the NorESM2 AMIP simulations. **?** show that off-equatorial net precipitation biases alone can initiate a feedback leading to an equatorial Pacific cold tongue in coupled simulations, and CAM6-Nor tends to develop such a bias. Finally, the near-surface ocean temperature bias pattern in OMIP1 simulations is cold along the equator, and warm on each side, which may further enhance off-equatorial precipitation. It should be noted that OMIP2 simulations with BLOM/CICE have a warm bias along the equator (**?**). The cold equatorial bias can affect ENSO variability and teleconnections. These are discussed further below.

### 5.2 Sea ice

The geographic distribution of sea ice in March and September, compared with observations are shown in Fig. **??** for NorESM2-LM (**??**e-h), and NorESM2-MM (**??**i-l). In common for both models for the Northern Hemisphere (Fig. **??**e,f,i,j) are too large sea ice extents in the Barents Sea and Greenland Sea and a too small extent in the Labrador Sea, Bering Sea, and Sea of Okhotsk during winter. The total areas are quite close to the observations as shown in Fig. **??**. These regional biases are most likely due to persistent biases in the oceanic and atmospheric circulation.

During summer, the distribution of sea ice in NorESM2-LM (Fig. **??**f) seems to be more variable. Apart from the persistent, positive bias in the East Greenland Current, the regional biases within the Arctic Ocean are more likely due to inter-annual variability, and the effect that the observations show a larger downward trend than the model.

NorESM2-MM (Fig. **??**j) shows too much sea ice in the central Arctic in September. In general, the model is colder in the Arctic than NorESM2-LM (Fig. **??**), and it has thicker sea ice in the Arctic Ocean. The Northern Hemisphere sea ice volume in NorESM2-MM is ~~19–23% (38–60~~21% (36%) larger in March (September) compared with the NorESM2-LM (not shown). The smaller seasonal cycle in ice area (Fig. **??**) and volume is consistent with a thicker sea ice cover in NorESM2-MM, both due to less winter growth because of increased insulation, and less summer melt due to higher albedo. The situation encountered in NorESM2-MM is similar to the results from NorESM1-M (**?**) and NorESM1-Happi (**?**). These models simulate ice cover that is too thick, with the reduction in the Northern Hemisphere summer ice area being too slow.

The winter sea ice area and extent is too low in the Southern Ocean in NorESM2 as seen in Fig. **??** and Fig. **??**(g-h,k-l). Winter area in September is around 4 million square km too small. The largest bias is found in the Atlantic-Indian sector. This bias seems to be associated with the warm bias in the ocean model, and the too warm ~~intermediate Antarctic water~~ Antarctic Intermediate Water (AAIW). The exact reason for this problem is not known, but the warm bias in AAIW is also evident in the OMIP simulations (not shown). However, these uncoupled simulations have a reasonable representation of the upper ocean temperature and the winter sea ice extent that are most likely due to the inherent relaxation towards observed atmospheric temperatures in those experiments. With the interactive atmosphere these problems increase.

### 5.3 Atmospheric temperature and winds

In terms of mean surface temperatures, NorESM2 is a warmer model than its preceding versions. The global-mean near-surface temperature (Fig. **??**) in NorESM1-M and NorESM1-Happi is generally too low with global-mean biases of ~~-0.62~~-0.76 K and ~~-0.94~~-1.08 K (see legends above panels in Fig. **??**). NorESM2-MM is closer to the reanalysis with a global-mean bias of ~~-0.05~~-0.19 K. Regionally, cold biases are mostly found in the polar regions and over the sub-tropical oceans. Warm biases are found over the Southern Ocean, North ~~Atlantic~~ American continent and in central Eurasia. NorESM2-LM (panel a) is warmer still, and overestimates the near-surface temperatures in the Arctic and in the global-mean, with a bias of ~~0.58~~0.43 K. NorESM2-

MM has the best overall performance also in terms of the global-mean RMSE, with ~~1.33~~1.35 K compared to ~~1.76~~1.62 K for NorESM2-LM, and ~~1.71~~1.83 K for NorESM1-Happi, and ~~1.79~~1.86 K for NorESM1-M (cf Fig. **??**).

Temperature biases are ~~mitigated~~ reduced in NorESM2 compared to NorESM1, not only near the surface, but also and especially in the mid and upper troposphere (Fig. **??**). ~~In particular~~ Tropospheric air temperatures tend to be systematically cold in all versions of both CESM and NorESM. NorESM2 has a reduced cold bias compared to NorESM1 particularly in the tropics and sub-tropics. This is mostly a consequence of the changes made to the cumulus convection scheme (Toniazzo et al., in prep.). The higher tropical SSTs in NorESM2-LM ~~being generally warmer in the tropics than~~ compared to NorESM2-MM ~~, its cold biases there are smaller~~lead to a reduced cold tropospheric tropical bias; however persistent cold mid- and high-latitude biases imply an excessive meridional temperature gradient. By contrast, NorESM2-MM shows improvements at all latitudes.

The stronger cool tropospheric and warm surface tropical bias of NorESM2-LM compared with NorESM2-MM is in line with the behaviour of both NorESM1 and CESM2. The systematic difference between the two atmosphere resolutions is also consistent between coupled and AMIP simulations, with CAM-Nor significantly cooler at two degree resolution than at one degree resolution for the same SSTs and the same physics parameters. At the same time, tropospheric specific humidity (and, a fortiori, relative humidity) is higher. Both lead to higher corresponding RESTOM. The ultimate cause of this systematic dependence of the simulated climatology on the resolution of the atmosphere model is not known. There may be a sensitivity of the convection parametrisation to the grid-scale variability of near-surface air parameters and to boundary-layer stability. Another possibility is a resolution dependence of cloud microphysics and the efficiency of stratiform precipitation. LWP and column precipitable water appear almost uniformly higher in CAM-Nor at two degree resolution than at one degree resolution.

All four of NorESM2-MM, NorESM2-LM, NorESM1-Happi and NorESM1-M tend to produce tropospheric westerly biases in zonal-mean zonal winds (Fig. **??**). At tropical and sub-tropical latitudes, these are more widespread in NorESM2 than NorESM1-M and NorESM1-Happi~~, and at the same time the easterlysurface biases are mitigated~~. Surface wind biases, which by contrast tend to be easterly, are reduced. At higher latitudes, all models tend to have westerly biases on the poleward side of the sub-polar surface jet (between 50° and 60°) in both hemispheres. The overestimation on the poleward flank is generally more pronounced in NorESM2 than in NorESM1. Comparing NorESM1-M to NorESM1-Happi and NorESM2-LM to NorESM2-MM, the biases in the zonal wind tend to be ameliorated with increased resolution. The differences in the tropics between NorESM2 and its predecessors ~~is~~ are in part attributable to the enforcement of conservation of atmospheric global angular or rotational momentum in NorESM2 ~~(?)~~(**?**). In all versions, in common with CAM6/CESM2, there is accumulation of westerly momentum near the model lid, where it is insufficiently damped.

### 5.4 Extratropical storm tracks

Extratropical storm tracks can be defined as regions of storminess associated with cyclogenesis, cyclone development, and cyclolysis which take place in the baroclinic zones between the sub-tropics and polar regions. They are important features at mid- and high latitudes as they are responsible for eddy transport of heat and momentum between low and high latitudes, and associated with potentially high-impact weather such as heavy precipitation and strong winds. Here, we diagnose storm-track activity by applying a bandpass filter to retain fluctuations in the geopotential height field at 500 hPa with periodicity corresponding to that of baroclinic waves, that is, between 2.5 and 6 days (**??**). The variability of the bandpass-filtered field is dominated by propagating low-pressure and high-pressure systems, and the storm tracks can be defined as geographically localized maxima in bandpass-filtered variability (**????**).

The climatological winter storm tracks are shown as the solid black contours in Fig. **??**. There are two maxima in the Northern Hemisphere, one over the North Atlantic and one over the North Pacific. The colors show the bias with respect to ERA-Interim (**?**). In NorESM1-M, storm-track activity is underestimated in both storm-track regions. In particular, the North-Atlantic storm track is overly zonal with too little activity on the equatorward side of the climatological maximum as well as over the Norwegian and Barents Sea (**??**). The magnitude of the bias is reduced in NorESM1-Happi compared to NorESM1-M in both storm-track regions. This is likely associated with the increased resolution in the atmosphere and land components (1° in NorESM1-Happi versus 2° in NorESM1-M).

Similar improvements are seen when comparing NorESM2-LM and NorESM2-MM. Both versions of NorESM2 are, furthermore, better able to simulate the North-Atlantic storm track with the size of the negative bias on its equatorward side being reduced. Overall, NorESM2-MM displays the smallest biases in Northern Hemisphere storm-track activity out of the four models. There remains, however, too little activity over the Norwegian Sea with extension into the Barents Sea.

In the Southern Hemisphere, the climatological winter storm track surrounds Antarctica with the largest variability occurring over the Indian Ocean (Fig. **??**). Storm-track activity is generally too weak on the equatorward side, with the largest biases being located over the Indian Ocean, close to the storm-track maximum. As in the Northern Hemisphere, the largest biases are found in NorESM1-M and the smallest biases in NorESM2-MM.

While the bandpass-filter approach yields a measure of storm-track activity, it cannot be used to isolate the individual cyclone centers. To further assess the robustness of the improvements between NorESM2-LM and NorESM2-MM, we therefore also consider results from the cyclone detection algorithm described in **?**. The method detects cyclones as minima in the sea-level pressure fields and sets the perimeter as the outermost closed sea-level pressure contour. The storm tracks are then seen as maxima in the local frequency of occurrence of surface cyclones, i.e. the fraction of time when cyclones are present in a given point (Fig. **??**a–b).

As for the bandpass-filter approach, the cyclone detection shows a clear reduction in the bias between NorESM2-LM and NorESM2-MM, which is likely to be associated with the higher horizontal resolution in the atmosphere and land components. The cyclone occurrence is underestimated on the equator-ward side of the North Pacific and Southern Hemisphere storm tracks and overestimated on the poleward side. Over the North Atlantic, the cyclone occurrence is underestimated on the equator-ward side of the storm track and over the Norwegian Sea extending into the Barents Sea, and overestimated between The British Isles and Greenland. The magnitude of the bias is clearly reduced in all regions in NorESM2-MM, with the improvement being particularly evident in the regions where the cyclone occurrence is overestimated.

Note that both the climatology and the biases should be expected to differ somewhat between the two approaches considered here because they capture different aspects of the storm tracks. The bandpass-filter approach does not distinguish between cyclones and anti-cyclones, and is dominated by growing and propagating baroclinic waves (**?**). The cyclone occurrence reflects the regions where cyclone centers are identified most frequently, and is for instance more sensitive to systems that are ~~slow~~ slowly moving or too long lived.

### 5.5 Clouds and forcing

Table **??** gives an overview of major forcing fluxes in NorESM2 compared to NorESM1 and observational estimates. Despite the large differences in physics and tuning, the overall numbers for top of the atmosphere fluxes and forcings are very similar to the numbers found in NorESM1-Happi and are generally within the observational range. There is however a slightly stronger negative bias in clear-sky LW flux and long wave cloud forcing. The latter is an unfortunate consequence of the tuning of high clouds in the model implemented in order to increase the outgoing long wave radiation. As seen from the upward LW flux estimate the outgoing long wave radiation is still within the estimate from satellite retrievals. SWCF values are very similar to the values of NorESM1-Happi and within the observational range. This number hides, however, a major weakness in NorESM1 stratiform cloud parameterisation which underestimated the cloud cover and compensated this by overestimating the cloud liquid water.

The major updates in cloud physics from CAM4 to CAM6 (**?**) improved the cloud cover, and the cloud liquid water path is now quite close to the observational estimate. The global cloud cover is still slightly lower than observed (Table **??**). This is partly connected to the tuning in NorESM2. Prior to the tuning the modelled cloud cover was higher than 70%. As seen from Fig. ~~**??**~~S5 in the supplement, the cloud cover underestimate is most pronounced in the tropics and subtropics in both hemispheres, while there is good agreement around the extra-tropical stormtrack regions and an overestimate in the high Arctic. Before the tuning (not shown) there was no bias at the low latitudes.

The modelled liquid water path seems to have a systematic bias towards low values at low latitudes and high values in the extra-tropics. Possible connections between cloud cover biases and the hydrological cycle are discussed in the next section.

### 5.6 Precipitation and hydrological cycle

The bias in the annual-mean total precipitation rate is shown in Fig. **??** for the two versions of NorESM1 and NorESM2 relative to ~~the ERA-Interim re-analysis. along with climatology~~ data from the Global Precipitation Climatology Project (GPCP; **?**). While the bias of the global-mean average is not systematically reduced between NorESM1 and NorESM2, there is a reduced RMSE, indicating that there is less cancellation between positive and negative biases in the global mean.

The reduction of the RMSE is also seen when considering the four seasons separately in Fig. ~~**??**~~ S6 in the supplement along with climatology from the GPCP. The evaluation of the mean bias, RMSE, and correlation included in the bottom left corner of each panel shows that RMSE and correlation have improved in NorESM2 compared to NorESM1 for all seasons. While the overall wet bias has increased slightly, mostly due to strong biases over the Pacific ocean, there are regions with a large reduction in mean bias. This is especially pronounced over Africa and equatorial Atlantic ocean. The largest improvement compared to NorESM1-M is seen for NorESM2-MM during northern hemisphere winter, when all three metrics (bias, RMSE, and correlation) consistently indicate higher skill.

As a measure of interannual variability, the standard deviation of monthly means for each season was calculated. The differences compared to GPCP are presented in Fig. ~~??~~S7 in the supplement. While NorESM1 slightly underestimates the precipitation variability, it is somewhat too high in NorESM2, with the magnitude of the bias being larger in all seasons~~except DJF and SON in NorESM2-MM~~. ~~As~~ NorESM2-MM improves RMSE of precipitation variability in all seasons except northern hemisphere autumn. As also seen for the mean climatology in Fig. ~~??~~S6 in the supplement, the correlation has improved for all seasons in both NorESM2-LM and MM.

The hydrological cycle (or cycling of fresh water) is of major importance for the climate system. Global means of precipitation and evaporation can serve as integrated measures of the properties of many processes in an ~~earth system model~~ESM. Results presented in Table **??** indicate that the intensity of hydrological cycle, as measured by evaporation, in NorESM2 is about ~~1.1~~1.4% larger globally (~~4.1~~4.9% over oceans) than in NorESM1-M. This is also manifested in the positive precipitation biases in ~~Figs.?? and ??~~Fig. **??**. While the values in Table **??** for NorESM2 are higher than for GPCP, they are closer to results from ERA-Interim calculated by ~~?~~**?**. Although NorESM1-Happi has the highest precipitation globally, NorESM2 has the highest precipitation over ocean, suggesting a larger re-cycling of oceanic water vapor and a lower fraction transported from oceans to continents (measured by E-P over oceans). The overestimated evaporation over oceans is likely linked to the underestimated cloudiness in the tropics and subtropics (~~see discussion above about~~ Fig. ~~??~~ S5 in the supplement). Solar radiation over subtropical ocean regions is an important driver of evaporation. The net moisture transport from oceans to continents is nevertheless smaller in NorESM2 than in NorESM1, consistent with more clouds in the extra-tropics and more marine precipitation in NorESM2. This analysis is only preliminary, however, and needs more in-depth studies which is out of scope of the present paper.

In ~~the~~ NorESM2 ~~earth system model~~ a closed hydrological cycle is present, with the difference between evaporation and precipitation being close to zero in the long-term average at equilibrium. In NorESM2-MM the discrepancy is ~~only 0.001~~ slightly improved to 0.023 $km^3$/year, whereas it is 0.027 $km^3$/year in NorESM1-M and ~~0.016~~ 0.031 $km^3$/year in NorESM2-LM (all values are means from members 1–3).

### 5.7 Northern Hemisphere blocking

While storm tracks are closely tied to precipitation, atmospheric blocking is associated with persistent anti-cyclones that inhibit precipitation for time scales up to several weeks. To diagnose blocking, we apply the variational Tibaldi and Molteni (vTM) index (**????**). Blocks are identified when there is persistent reversal of the 500 hPa geopotential height field around a central latitude that last for at least five days and cover at least 7.5 consecutive longitudes. The central longitude varies with the position of the maximum in the Northern Hemisphere climatological storm track.

The seasonal blocking frequency is mostly underestimated over the North Atlantic and in Europe in the four versions of NorESM (Fig. **??**), particularly during winter (DJF). During spring (MAM), NorESM2-MM is closest to the reanalysis, while during summer (JJA) and autumn (SON), NorESM1-Happi performs best in these regions. While NorESM1 tends to overestimate the blocking frequency over the Pacific, NorESM2 generally lies closer to the reanalysis in that sector. Consider, for instance, the region between 120° E and 180° E during summer, or the region between 130° W and 90° W during winter. In summary, although the use of 30 years from ERA-Interim for verification may not be fully representative for blocking climatology, the representation of NH blocking continues to be a challenge in NorESM, and in particular over the Atlantic-European sector in winter.

### 5.8 Madden-Julian Oscillation

In the tropical atmosphere, the Madden-Julian Oscillation (MJO) is the dominant mode of variability on timescales between 30 and 90 days (**??**). The MJO is characterized by ~~large-scale regions of enhanced and suppressed convection that slowly propagate eastwards~~ a meridional dipole of convective precipitation anomalies along the equator, that slowly propagates eastwards and interacts with a number of other circulation features such as El Niño events (**?**), the Indian summer monsoon (**?**), tropical cyclones (**?**), and even the North Atlantic Oscillation and extratropical variability (**?**).

~~We diagnose the MJO in two ways. One is in terms of temporal correlations between subseasonally filtered anomalies of precipitation and winds along the equatorial Indian ocean. The second is in terms of~~ The MJO is characterised by a specific feature in wavenumber-frequency spectrum ~~for~~ of equatorial 850 hPa zonal wind (U850) and ~~for~~ outgoing longwave radiation (~~Fig. **??**). These diagnostics have been proposed and described in detail in **?**.~~

~~Positive wavenumbers and frequencies indicate eastward propagation, while negative frequencies (or wavenumbers) indicate westward propagation. The energy in the spectrum for U850 from ERA-Interim (Fig. **??**a)shows that the energy in the reanalysis is~~ OLR; **?**), associated with wavenumbers 1–3~~with a maximum for~~ , a maximum at wavenumber 1, and ~~with the energy being more or less contained within timescales of~~ periods between 30 to 80 days. NorESM2-LM and NorESM2-MM ~~also show maximum energy for the same wavenumbers, with the maximum occurring for wavenumber 1, as in ERA-Interim. The maximum is, however, somewhat too strong in both models and the energy is spread out over a wider range of timescales. Both NorESM2-MM and NorESM2-LM peak at longer timescales (lower frequency) than ERA-Interim, and NorESM2-LM has an additional peak at shorter timescales (higher friquency). Similar results are found when comparing the wave-number frequency spectra for outgoing longwave radiation from NorESM2-LM and~~ possess this mode in U850 spectra (Fig. S8 in the supplement), but its spread in wavenumber is too narrow and its spread in frequency too wide. Furthermore, OLR variability is too weak, and the mode appears preferentially as a stationary oscillation in the Indian-ocean sector, with too little zonal and meridional propagation (Fig. S9 in the supplement). The relationship between zonal winds and precipitation anomalies, with the former in quadrature with respect to the latter, are similar in the simulations and in observations. In NorESM2-MM ~~with that from NOAA, however here the peak energy is underestimated.~~

~~Lead-lag correlations with respect to precipitation anomalies during extended winter in the equatorial Indian Ocean around 90E are characterised in observations (Figure **??**(a))by a marked, slow eastward propagation and some poleward propagation. There is a strong relationship with zonal winds in quadrature such that westerly wind anomaly maxima precede the precipitation maxima. These characteristics are simulated fairly clearly in NorESM-LM (Figure **??**(b)), although there is little propagation into the Indian Ocean from the West, or poleward propagation. The MJO in NorESM-MM (Figure **??**~~(Fig. S9(c) ~~) has in general similar properties, but it is too weakin amplitude~~in the supplement) however the anomalies appear to be generally too weak. Composite plots of MJO events (not shown) indicate a tendency in both model versions to generate westward-propagating convective anomalies, which may weaken activity in the MJO region of the spectrum. The ability of the simulated MJO mode to propagate eastwards as observed appears to be sensitive to the distribution of tropical SSTs in both CESM2 and NorESM2 (Richard Neale, personal communication; Toniazzo et al., in prep.).

### 5.9 El Niño Southern Oscillation (ENSO)

The coupled model internally generates a self-sustained ENSO mode with spatial and temporal characteristics similar to observations. (The timeseries of NINO3.4 SST anomalies are shown in Figure ??Fig. S10 in the supplement, alongside the observed one). The ENSO in LM and MM model versions are very similar in magnitude (Figures ??, ??, Fig. ??–??), spatial pattern (Figure **??**), and spectral power distribution in frequency space (Figure Fig. **??**). ENSO SST anomalies are very large compared to observations (with a NINO3.4 anomaly greater than 2.5° C in the average El-Niño event, compared with 1.5° C in observations), and they tend to peak early in the season, i.e. between November and December instead of between December and January as observed (Figure Fig. **??**a). The early peak and termination may be partly attributable to weak zonal wind-stress anomalies over the equatorial region, which also peak early, notwithstanding a robust response in equatorial precipitation (Figure Fig. **??**b,c). Such weak surface wind response may be caused by the general displacement, with respect to observations, of the maximum of climatological precipitation north of the Equator along the Pacific ITCZ. Especially in MM, precipitation anomalies also have their maximum north of the equator (Figure **??**b), which tends to result in weaker equatorial anomalous westerlies. A second origin of the early simulated El-Niño SST peak may however also be found in the early rapid demise of positive thermocline depth anomalies in the NINO3 region during El-Niño events, which is seen also in OMIP1 and OMIP2 simulations forced with prescribed wind-stress (Figure Fig. **??**d). Given the weak coupled wind-stress and thermocline activity, the large SST anomalies may be partly the result of insufficient surface damping by the action of anomalous surface heat fluxes.

Correlation analysis shows that indeed over the eastern equatorial Pacific the model tends to generate positive downward net short-wave radiative flux anomalies when SST anomalies are positive, in contrast to observations. This might also explain the growth of positive SST anomalies in the NINO3.4 region early during El-Niño events even before positive 20° C isotherm depth anomalies have fully reached the area; and the long persistence of both SST and precipitation anomalies in the later stages of El-Niño events. The model climatological bias of a pronounced double ITCZ, with strong ITCZ precipitation away from the equator and a dry, cold equatorial region dominated by marine stratocumulus, rather than trade-cumulus cloud in the eastern Pacific probably contributes to this behaviour. Toniazzo et al. (in prep.) shows that changes in the convection scheme that were made in order to mitigate the tropospheric cold bias and the positive TOA net residual have contributed to this error. Off-equatorial precipitation tends to couple ~~with westward-propagating equatorial modes and can lead to a tendency for westward propagation of convective activity~~ less effectively with eastward-propagating equatorial modes (cf. ~~Figure ??~~also previous section and Fig. S9). Westward propagation is also evident in the model's ENSO during the phase change from El-Niño to La-Niña. In spite of such shortcomings, ENSO-related variability in NorESM2 is generally similar to the observed one. In particular, NINO3.4 spectra in the two model versions and in observations are formally indistinguishable (Figure Fig. **??**). The simulated composite El-Niño SST, precipitation, and geopotential height anomalies have good global pattern correlations with the observed composite El-Niño anomalies (Figure ??(Fig. S11 in the supplement), indicating that the simulated ENSO adds important and useful features to the climatology simulated by the model. Particularly prominent and fairly realistic are teleconnections into both hemispheres during and after ENSO peaks, with a PNA pattern that extends into the storm-track entry region of the western Atlantic, as observed. In this respect NorESM2-MM validates better than NorESM2-LM, in spite of its equivalent of slightly worse equatorial ENSO biases, probably due to a better overall sub-tropical and high-latitude atmospheric circulation.

## 6 Summary and conclusions

This paper presents and evaluates NorESM2 (the second version of the Norwegian Earth System Model) used for conducting experiments for CMIP6. NorESM2 is based on CESM2 (the second version of the Community Earth System Model), but with several important differences. While the land and sea ice components are largely the same as in CESM2, NorESM2 has entirely different models for the ocean and ocean biogeochemistry, namely BLOM and iHAMOCC. There are also several differences in the atmosphere model (CAM6-Nor), including a different module for aerosol life-cycle, aerosol-radiation-cloud interactions and with changes related to the moist energy formulation, deep convection scheme and angular momentum conservation. Finally, the turbulent air-sea flux calculations are modified and proper time-averaging of solar zenith angle in albedo estimation implemented.

We report results from the CMIP6 DECK experiments, including the pre-industrial control, the abrupt quadrupling of $CO_2$ concentration levels, and 1% increase per year of $CO_2$ concentrations until quadrupling, along with the CMIP6 historical experiments, and the ScenarioMIP Tier 1 experiments (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5). The experiments were all carried out with both a the atmospheric medium-resolution version of the model (NorESM2-MM) and a low-resolution version (NorESM2-LM).

The drift over the 500 year long pre-industrial control experiment is generally very small for both versions of the model. NorESM2 in both model resolutions is slightly less sensitive than its predecessors and at the lower end of the CMIP5 and CMIP6 multi-model mean (preliminary calculations, Gjermunden et al, in prep)for both model resolutions (**?**), with the equilibrium climate sensitivity of 2.5 K estimated using the Gregory method (**?**).

The historical reconstruction of surface temperatures is similar in both model versions. A significant temperature increase due to enhanced climate forcing is setting in found late in the historical period. Both model versions reach present day warming levels to within 0.2° C of observed temperatures in 2015. Aerosol forcing may be responsible for the delayed warming in the late 20th centuraycentury. Aerosol effective radiative forcing reaches levels of -1.5 W $m^{-2}$ in the period 1970–1980, becoming slightly weaker again in 2014 with a value of -1.2 -1.36 W $m^{-2}$ in NorESM2-LM and -1.26 W $m^{-2}$ in NorESM2-MM.

Under the four scenarios SSP126, SSP245, SSP370, and SSP585SSP-1.26, SSP-2.45, SSP-3.70, and SSP-5.85, the warming in the period 2090–2099 compared to 1850–1879 reaches 1.3, 2.2, 3.0, and 3.9 K in NorESM2-LM, and 1.3, 2.1, 3.1, and 3.9 K in NorESM-MMNorESM2-MM, robustly similar in both resolutions.

In particular NorESM2-LM shows a ~~satisfactorily~~ satisfactory evolution of recent ~~sea ice~~ sea-ice area. In NorESM2-LM~~an ice free~~ , an ice-free Arctic Ocean is only avoided in the ~~SSP1-26~~ SSP1-2.6 scenario. NorESM2-MM simulates ~~higher sea ice~~ larger sea-ice area both at present and in future scenarios.

The pattern of some biases seen in the fully coupled simulations considered here are similar in coupled ocean-sea ice simulations carried out for OMIP and can thus be linked to the ocean model having too coarse resolution and shortcomings in parameterised processes. NorESM2-LM and MM largely share the same biases in the surface ocean, although the MM version is somewhat closer to the observations. Most of the large-scale biases in the surface ocean are also seen in the subsurface.

Like CESM2, NorESM2 is generally a "cold" model, with an initial deficit in atmospheric long-wave cooling that causes a positive RESTOM and leads to heat gain by the ocean and positive SST biases particularly in the tropics. NorESM2 represents an improvement in this respect compared to NorESM1. This is particularly evident in the tropical and sub-tropical troposphere (Fig. **??**). In addition, the medium-resolution version of the model has more realistic ~~upper tropospheric~~ upper-tropospheric meridional temperature gradients, and reduced near-surface temperature biases.

The extratropical storm tracks are generally better simulated in NorESM2 than in NorESM1, particularly over the North Atlantic. The storm tracks additionally improve with higher resolution, both in the Northern and Southern Hemisphere.

Several aspects of the modeled cycling of fresh water are improved in NorESM2 compared to NorESM1, including the RMSE and spatial correlation of the bias in the total precipitation rate. The intensity of the hydrological cycle as compared to the ~~observationally based findings of **?**~~ observationally-based findings of **?** is slightly exaggerated in NorESM2, as it was in NorESM1, consistent with the underestimated cloudiness and thus overestimated solar radiation in the tropics and sub-tropics. The transport of oceanic water vapor over the continents is smaller in NorESM2 than NorESM1, indicating a slightly too efficient re-cycling of oceanic water vapor associated with over-estimated oceanic precipitation and higher cloudiness in the extratropics.

The seasonal blocking frequency in the Northern Hemisphere is ~~in particular~~ especially underestimated over the Atlantic - European sector during winter (DJF) by NorESM2. During spring (MAM), NorESM2-MM is closest to the reanalysis, while during summer (JJA) and autumn (SON), NorESM1-Happi performs best in these regions. While NorESM1 tends to overestimate the blocking frequency over the Pacific, NorESM2 generally lies closer to the reanalysis in that sector. Although the use of 30 years from ERA-Interim for verification may not be fully representative for blocking climatology, the simulation of ~~NH~~ Northern Hemisphere blocking continues to be a challenge for NorESM.

The coupled model internally generates a self-sustained ENSO mode with spatial and temporal characteristics similar to observations. ENSO SST anomalies are very large compared to observations (with a NINO3.4 anomaly greater than 2.5° C in the average El-Niño event, compared with 1.5° C in observations), and they tend to peak early in the season, i.e. between November and December instead of between December and January as observed. Nevertheless many ~~proprties~~ properties of the ENSO are similar to those observed, and El-Niño teleconnections are quite realistic both in the tropics and at mid- and high latitudes. Less satisfactory is the performance of the coupled model in terms of the Madden-Julian oscillation. Here the ~~low resolution version~~ model version with low resolution in the atmosphere appears to produce more intense and more realistic sub-seasonal tropical variability than the medium-resolution version.

*Code and data availability.* The CMIP6 and CMIP5 data considered here is available through data archives operated and maintained by the Earth System Grid Federation: https://esgf-node.llnl.gov/search/cmip6/ and https://esgf-node.llnl.gov/projects/cmip5/. The model code is available on github, is available on request from the corresponding author and will be fully publicly available by September 2020.

*Author contributions.* ØS coordinated the writing of this article. ØS, DO, AK and TT wrote the CAM6-Nor description section. MB, AN, JS, JT wrote the ocean component description section. AN performed the evaluation of the model climatology for the oceanic variables and wrote the corresponding sections. JBD performed the evaluation and wrote the sea ice section. TT performed the evaluation and wrote the ENSO section. LSG performed the analysis of the climatological mean state of near-surface temperature, zonally averaged temperature and zonal winds, extratropical storm tracks, atmospheric blocking, and the MJO and wrote the corresponding sections with contributions from TT. CS contributed to the storm-track analysis. OL and TI performed the analysis and wrote the section on precipitation and the hydrological cycle. DO and Ada G conducted the ECS and feedback analysis. TL performed the analysis and wrote the section on future scenarios. The development of the NorESM2 model was co-ordinated by CH and TI during the first years (EVA-project), and is co-ordinated by MB and MS since fall 2017. All co-authors contributed to the model development, the modelling infrastructure, the setting up of the CMIP6 experiments, data processing and/or data distribution and gave feedback to the manuscript.

*Competing interests.* There are no competing interests.

*Acknowledgements.* We thank all the scientists, software engineers, and administrators who contributed to the development of CESM2, on which NorESM2 is based. We are particularly grateful to Cecile Hannay, M. Vertenstein, A. Gettelmann, J.F. Lamarque and others for invaluable advice on numerous scientific and technical issues when using CESM code, and the support by the CESM program directors during the NorESM2 development period. We acknowledge support from the Research Council of Norway funded projects EVA (229771), HappiEVA (261821), INES (270061), and KeyClim (295046), Horizon 2020 projects CRESCENDO (Coordinated Research in Earth Systems and Climate: Experiments, Knowledge, Dissemination and Outreach, no. 641816,), APPLICATE no. 727862 and IS-ENES3 , no. 824084, NS Nordic project eSTICC (57001). High performance computing and storage resources were provided by the Norwegian infrastructure for computational science (through projects NN2345K, NN9560K, NN9252K, NS2345K, NS9082K, NS9560K, NS9252K, and NS9034K) and the Norwegian Meteorological Institute. The observational SLA dataset used in this work was obtained from the obs4MIPs project hosted on the Earth System Grid Federation and the original altimeter products were produced by Ssalto/Duacs and originally distributed by Aviso+, with support from CNES (https://www.aviso.altimetry.fr). Monthly distributions of stratospheric sulfate aerosols were adjusted for use in NorESM by Luo Beiping at ETHZ.