Framework for an ocean-connected supermodel of the Earth System

Francois Counillon[1], Keenlyside Noel S[2], Shuo Wang[2], Marion Devilliers[3], Alok Kumar Gupta[4], Koseki Shunya[5], and Mao-Lin Shen[2]

[1]University of Bergen
[2]Geophysical Institute, University of Bergen
[3]DMI
[4]NORCE
[5]Geophysical Institute, University of Bergen, Bjerknes Centre for Climate Reseacrh

February 26, 2024

Abstract

A supermodel connects different models interactively so that their systematic errors compensate and achieve a model with superior performance. It differs from the standard non-interactive multi-model ensembles (NI), which combines model outputs a-posteriori. We formulate the first supermodel framework for Earth System Models (ESMs) and use data assimilation to synchronise models. The ocean of three ESMs is synchronised every month by assimilating pseudo sea surface temperature (SST) observations generated from them. Discrepancies in grid and resolution are handled by constructing the synthetic pseudo-observations on a common grid. We compare the performance of two supermodel approaches to that of the NI for 1980—2006. In the first (EW), the models are connected to the equal-weight multi-model mean, while in the second (SINGLE), they are connected to a single model. Both versions achieve synchronisation in locations where the ocean drives the climate variability. The time variability of the supermodel multi-model mean SST is reduced compared to the observed variability; most where synchronisation is not achieved and is bounded by NI. The damping is larger in EW than in SINGLE because EW yields additional damping of the variability in the individual models. Hence, under partial synchronisation, the part of variability that is not synchronised gets damped in the multi-model average pseudo-observations, causing a deflation during the assimilation. The SST bias in individual models of EW is reduced compared to that of NI, and so is its multi-model mean in the synchronised regions. The performance of a trained supermodel remains to be tested.

Framework for an ocean-connected supermodel of the Earth System

François Counillon[1,2], Noel Keenlyside[1,2], Shuo Wang[1], Marion Devilliers[3], Alok Gupta[4], Shunya Koseki[1], Mao-Lin Shen[1]

[1]Geophysical Institute, University of Bergen, Bjerknes Centre for Climate Research, Bergen, Norway
[2]Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, Bergen, Norway
[3]Danish Meteorological Institute, Copenhagen, Denmark
[4]NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

Key Points:

- Data assimilation can be used to synchronise different Earth System models.
- Partial synchronisation is achieved and bias is reduced in key regions by exchanging monthly sea surface temperature.
- Synchronising the models towards a weighted-mean causes a deflation of variability.

Corresponding author: François Counillon, `francois.counillon@nersc.no`

Abstract

A supermodel connects different models interactively so that their systematic errors compensate and achieve a model with superior performance. It differs from the standard non-interactive multi-model ensembles (NI), which combines model outputs a-posteriori. We formulate the first supermodel framework for Earth System Models (ESMs) and use data assimilation to synchronise models. The ocean of three ESMs is synchronised every month by assimilating pseudo sea surface temperature (SST) observations generated from them. Discrepancies in grid and resolution are handled by constructing the synthetic pseudo-observations on a common grid. We compare the performance of two supermodel approaches to that of the NI for 1980—2006. In the first (EW), the models are connected to the equal-weight multi-model mean, while in the second (SINGLE), they are connected to a single model. Both versions achieve synchronisation in locations where the ocean drives the climate variability. The time variability of the supermodel multi-model mean SST is reduced compared to the observed variability; most where synchronisation is not achieved and is bounded by NI. The damping is larger in EW than in SINGLE because EW yields additional damping of the variability in the individual models. Hence, under partial synchronisation, the part of variability that is not synchronised gets damped in the multi-model average pseudo-observations, causing a deflation during the assimilation. The SST bias in individual models of EW is reduced compared to that of NI, and so is its multi-model mean in the synchronised regions. The performance of a trained supermodel remains to be tested.

Plain Language Summary

Supermodelling is a novel approach in which different models are run interactively and are connected to achieve a model with superior performance. The method exploits model diversity and reduces model error by training the connection terms using observations. Structural differences between the models and the amount of data exchange among models caused technical challenges and have limited the applicability of supermodelling to Earth system models. Here, we show that data assimilation can handle these discrepancies. A demonstration is done with three Earth system models having different resolution and structural differences and using a limited amount of data exchange (monthly ocean surface temperature). Synchronisation is achieved in several key regions. There, the ocean surface temperature error is smaller than by taking the a-posteriori average of the non-interactive multi-model ensemble, an approach commonly used for handling Coupled Model Intercomparison Project simulations. Connecting the model via their weighted mean (one of the two approaches tested) causes a spurious deflation of variability. This study opens the application and training of supermodelling to Earth system models.

1 Introduction

Climate models have been key tools for understanding fundamental questions about our climate systems. However, large uncertainties exist with many key processes being parameterised and biases in several of the Earth system components (e.g., ocean, atmosphere, sea ice and land) being larger than the projected changes in climate and larger than the variability being predicted (Palmer & Stevens, 2019). While models have improved through the successive generations of coordinate model intercomparison project (CMIP) — version 6 being the latest (Eyring et al., 2016) — many of the large biases persist (e.g., the double-Intertropical Convergence Zone (ITCZ) problem (Tian & Dong, 2020); the tropical Atlantic bias (Richter et al., 2014) and the signal-to-noise paradox (Scaife & Smith, 2018). Although some simulations are being tested at a breakthrough resolution (Zhongming et al., 2020), we are still decades away from being able to perform operational climate simulations with models that can explicitly resolve the most important physical processes. In the meantime, one can test alternate methods to understand and mitigate these biases using the current generation of models.

The classical approach to mitigate model error is to take the multi-model average of independent model simulations (a-posteriori) so that errors in the different models cancel. This approach is standard for climate projection but can also improve predictions (Branicki & Majda, 2015). One can refine the post-processing by taking a weighted mean of the different runs — an approach referred to as a super ensemble (Krishnamurti et al., 2016). However, this approach has limitations as most models share the same deficiencies (e.g., double ITCZ, warm bias in the tropical Atlantic) and because linear post-processing does not necessarily correct non-linear responses, such as climate sensitivity.

Supermodel builds on the interactive ensemble (Kirtman & Shukla, 2002) where multiple realizations of the same atmospheric general circulation model are simultaneously coupled to a single ocean general circulation model through averaging their air-sea fluxes. Supermodels couple simultaneously different models and take advantage of model diversity to compensate for their errors (Duane et al., 2018). Models are connected as they run via their state variables or their tendencies. Models can either be connected to each other (e.g. Mirchev et al., 2012; Smith, 2001) or towards their weighted mean (Wiegerinck et al., 2013; Schevenhoven & Carrassi, 2021). During a training phase, the connection terms are optimised to formulate a new synchronised dynamical system that achieves enhanced performance. Supermodels rely on two important properties: first, it is possible to synchronise different models through a few variables — an approach referred to as chaos synchronisation of non-linear dynamical systems (Pecora et al., 1997; Duane & Tribbia, 2001) — and second, model diversity can encompass the true behaviour of the dynamical system.

Supermodels have been tested with various models and experimental designs of different complexity. Idealised framework experiments (or observing system simulation experiments, Halem & Dlouhy, 1984) are convenient because they allow controlling challenges faced with the real framework and because the truth (constructed from a model) is known. One can introduce model error (e.g., by perturbing parameter values or using a different model) and disclose only part of the true model state as observations (perfect or not). Supermodelling was successfully demonstrated for parametric model error forming a convex envelope around the truth and that with low dimensional dynamical systems (e.g., Lorenz 63, Lorentz 96, Rossler systems; see Mirchev et al., 2012; Van den Berge et al., 2011; Du & Smith, 2017), quasi-geostrophic atmospheric models (Schevenhoven & Selten, 2017; Wiegerinck & Selten, 2017), and the global atmosphere–ocean–land model of intermediate complexity SPEEDO (Schevenhoven & Selten, 2017; Selten et al., 2017). However, when the model error does not cancel out (parameters do not form a convex envelop around the truth), one can use negative weights (Schevenhoven et al., 2019; Schevenhoven & Carrassi, 2021), which raises new challenges. Furthermore, the supermodel can degrade performance at a different time scale than it was trained for when the imperfect models do not fully resolve the processes of the truth (e.g., in Wiegerinck & Selten, 2017, using models at a coarser resolution). Nevertheless, the first demonstration of supermodelling with real data successfully mitigated the double ITCZ bias and improved the representation of the dynamics in the equatorial Pacific (Shen et al., 2016, 2017). These results were achieved with two versions of the ECHAM5 Atmospheric General Circulation Models (AGCM) — each of them using a different convection scheme — providing the weighted average fluxes to a single Oceanic General Circulation Model (OGCM), MPIOM (Shen et al., 2016). One may expect to enhance the performance of supermodelling by broadening model diversity that expands the convex envelop so that it may enclose the true Earth system – to the extent possible, recognising that the real climate system is far more complicated than any numerical model.

Building a supermodel with fundamentally different ESMs raises new challenges. Models do not share the same model-state space, nor do they have comparable resolution and representativity (Janjić et al., 2018). For this purpose, Du and Smith (2017) suggested formulating pseudo-observations of the different models and assimilating them back into the different models. The approach was demonstrated with low-dimensional systems and in an idealised framework. Another challenge is practical limitations with feasible data volume exchanged among models and frequency of the synchronisation steps. Here, we aim to assess whether a supermodel can achieve synchronisation (a requirement of supermodelling) in a configuration where data exchange is sparse, synchronisation steps infrequent, and models have a radically different structure. We will apply part of the formalism proposed by Du and Smith (2017)— connecting models via assimilation

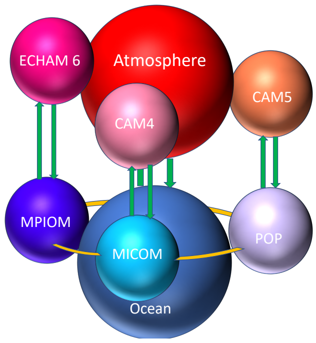

Figure 1. Schematics of the supermodel. The green arrows denote dynamical coupling in the individual models, and the yellow band denotes synchronisation done in the different ocean models.

of synthetic pseudo-observations generated from the multi-model ensemble. We will examine the impact of generating the pseudo-observations from a single model or constructing it from an equal-weighted mean. In future work, we will implement training and analyse the resulting supermodel's performance.

This paper is organised as follows. Section 2 presents the practical implementation of the supermodelling framework: the description of the individual ESMs, the synchronisation methodology and the data assimilation method. Section 3 introduces the validation data sets and metrics, and Section 4 presents the result of two prototype supermodelling approaches, first globally with a focus on the damping of internal variability and impact on the SST bias and second with a focus on the ENSO region.

2 Supermodelling framework

This section describes the practical implementation of the supermodel framework for Earth system models using ocean connection. It combines the Norwegian Earth System Model (NorESM), the Community Earth System Model (CESM) and the Max Planck Institute Earth System Model (MPI-ESM) (see Figure 1). The three ESMs are connected via their SST every month. We implement an individual assimilation system based on the Ensemble Optimal Interpolation. The data assimilation method updates the whole water column based on the synthetic SST pseudo-observations constructed from the multi-model ensemble mean or from a single model.

2.1 CESM

The Community Earth System Model is a global, fully coupled model for climate simulations. We used the CESM Large Ensemble Project (LENS) version based on CESM1.1.2 (Kay et al., 2015), with all components at approximately 1° horizontal resolution. External forcing complies with CMIP5's historical experiment. The atmospheric component is the Community Atmosphere Model version 5, (CAM5, Hurrell et al., 2013) with 30 vertical levels and a finite-volume grid (f09, i.e., approximately 1°). The Community Land Model, version 4 (CLM4), is on the same grid as the atmosphere. The Parallel Ocean Program, version 2 (POP), is run with 60 vertical levels. The horizontal resolution of the ocean is approximately 1°, but it is enhanced in the meridional direction around the equator and both in zonal and meridional directions at high latitudes (g16 grid). The sea ice [Los Alamos Sea Ice Model (CICE)version 4] component model is on the same grid as the ocean model. We use historical forcing from 1920 to 2005 (Lamarque et al., 2010) and retrieve the initial conditions from NCAR repository (b.e11.B20TRC5CNBDRD.f09_g16.001) in 1950. As POP2 only allows modification for one time level of the leapfrog scheme, we follow the approach of DART (Anderson et al., 2009). It uses a forward Euler scheme for the first time step but reverts to the leapfrog scheme afterwards. The barotropic velocities and surface pressure gradients are adjusted to preserve the ocean volume. The flag POPDART can activate this option in POP2.

2.2 NorESM

We use the medium-resolution NorESM1-ME (Tjiputra et al., 2013) that contributed to the Coupled Model Intercomparison Project Phase 5 (CMIP5). It is based on the Community Earth System Model version 1.0.3 (CESM1, Vertenstein et al., 2012). However, the atmospheric component (CAM4-OSLO) features an advanced aerosol chemistry scheme (Kirkevåg et al., 2013), and the ocean is an updated version of the isopycnal coordinates ocean model MICOM (Bentsen et al., 2012). The atmosphere and land components are configured on a finite-volume grid with a latitude and longitude resolution of $1.9 \times 2.5^{\circ}$. The atmosphere component uses 26 hybrid sigma-pressure levels with a model top at approximately three hPa. The horizontal resolution of the ocean and sea-ice model is approximately 1°. The ocean uses 51 isopycnal layers and two layers representing the bulk mixed layer with time-evolving thicknesses and densities. The initial condition is taken from a historical simulation in 1980 that started from a stable pre-industrial simulation in 1850.

2.3 MPI-ESM

We use the MPI-ESM1-LR (Block & Mauritsen, 2013; Dunstone et al., 2018; Giorgetta et al., 2013) model that contributed to the CMIP5. The AGCM of MPI-ESM1-LR is the 6^{th} generation European Centre Hamburg general circulation model (ECHAM6 Stevens et al., 2013), and the OGCM is the Max Planck Institute Ocean Model (MPIOM) (Marsland et al., 2003; Jungclaus et al., 2013). The land model (JSBACH, Reick et al., 2013; Schneck et al., 2013), which includes vegetation, and the marine biogeochemistry model (HAMOCC5, Ilyina et al., 2013) are considered as subsystems of ECHAM6 and MPIOM, respectively. ECHAM6 employs T63 spectral resolution (approximately 1.9° horizontal resolution) and 47 vertical levels, and MPIOM employs a rotated curvilinear grid with an approximate 1.5° horizontal resolution and 40 z-levels. The poles of MPIOM are moved to Greenland and the Weddell Sea.

Table 1. We summarise the key characteristics of the different ESMs used. The first column details the model versions. The second column reports the name of the ocean models, their resolution and coordinate system in the vertical — i.e., isopycnal (σ-coordinate) or geopotential depth (z-coordinate). The third column is the name of the atmospheric models and their discretisation scheme. The last column provides a reference to the model version.

Model version	Ocean	atmosphere	reference
NorESM1-ME	MICOM(σ; 1°)	CAM4 (finite-volume, 2°)	Tjiputra et al. (2013)
CESM1.1.2	POP2 (z, 1°)	CAM5 (finite-volume, 1°)	Kay et al. (2015)
MPI-ESM1-LR	MPIOM (z,1.5°)	ECHAM6 (spectral, 2°)	Block and Mauritsen (2013)

2.4 Synchronisation methodology

The three ESMs use different models, grids, coordinates, and resolutions (Table 1). NorESM and CESM may be more similar, but they use a fundamentally different ocean model (in geopotential depth for CESM and isopycnal coordinates for NorESM), and the atmospheric model in CESM is a more advanced version (CAM5 vs CAM4) and has a higher resolution (1° vs 2°). Consequently, one cannot simply use state replacement or nudging, as they will generate imbalances and may lead the models to crash. Data assimilation (DA) can estimate the best possible (most likely) state based on observations, a dynamical model, and their uncertainties. It is designed to preserve the dynamical consistency of the individual models and optimally handle observation uncertainty (Carrassi et al., 2018).

A limitation of DA methods for this application is that they are (with few exceptions, e.g., S. Zhang et al., 2007; Nerger et al., 2020) working offline - meaning that the model is stopped, the state written on disk, data assimilation applied on the files and the model restarted. With such large systems as the ESMs, the time required for initialising the model and writing the input/output is burdensome (see, e.g., Karspeck et al., 2018), limiting the feasible frequency of the synchronisation. Similarly, one needs to limit the number of variables that will be synchronised to keep the cost of the DA-step low.

As a first attempt before developing a more advanced connected supermodel of ESMs, we try to synchronise the three models through their SST at a monthly frequency. SST is sufficient to constrain the variability in many regions of the earth system, particularly in the tropics (Shukla, 1998; Zhu et al., 2017; Wang et al., 2019). It is observed over a long period with a good level of accuracy, enabling the possibility to effectively train our supermodel and validate it for an independent period. With monthly synchronisation, the additional computational cost of the monthly assimilation remains small.

We test two supermodelling approaches that differ in their formulation of the pseudo-observations. The first scheme belongs to the category of state-constrained weighted supermodel (Schevenhoven & Carrassi, 2021). The three models are integrated forward for one month, and their SSTs are interpolated to a common 1° grid. The pseudo-observations are the equal-weighted mean of these outputs (referred to as **EW** in the following). Weights should be trained using observations to optimise model performance, but this optimisation is out of the scope of the paper. The pseudo-observations are then assimilated into the individual models using the EnOI method (see Section 2.5), and the models are then restarted for the next cycle. The first synchronisation step in the two supermodel framework started on the first of February 1980 for practical reasons. We do not synchronise models under a the union of all three models sea ice mask.

In the second scheme, the workflow is similar, but the pseudo-observations are formulated from a single model (hereafter referred to as **SINGLE**). This approach is, for instance, used in the cross-pollination in time method (Smith, 2001; Du & Smith, 2017). Another objective of this experiment is that it should not be affected by variability damping and can serve as a benchmark for the EW. It can serve to assess the potential synchronisation that can be achieved with SST and monthly synchronisation steps. We have selected arguably CESM that has higher resolution and provides overall the best performance. We interpolate the CESM pseudo-observations onto the common grid to have a comparable interpolation error between the two schemes. Formulating the pseudo-observations in the native CESM grid enhances the performance slightly but does not change any conclusion unless reported.

The performances of the two supermodel approaches are compared to a non-interactive multi-model ensemble (hereafter referred to as **NI**) which start from the same initial condition than the two supermodels, but do not stop for the synchronisation steps. All models start from a historical simulation in 1980 using CMIP5 historical forcing, and RCP8.5 is used for 2006 (Taylor et al., 2012).

The system runs on 11 nodes (1408 CPU) on the Norwegian high performance computer Betzy (a BullSequana XH2000) and can achieve about 10 model-year per day. CESM runs on eight nodes and runs one month in approximately seven minutes, NorESM runs on two nodes and performs the 1-month simulation in approximately seven minutes, MPI-ESM uses one node and performs a 1-month simulation in approximately four and half minutes. The model integration accounts for approximately seven minutes, while the DA step accounts for approximately two minutes. The assimilation is currently performed sequentially for each model, and this step could have been reduced to 40 seconds if parallelised.

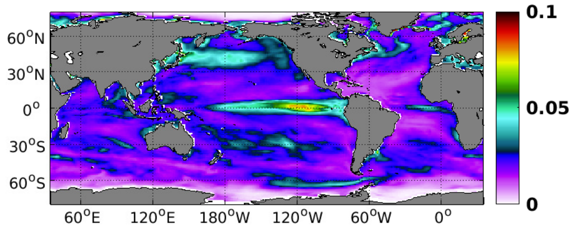

Figure 2. Pseudo-observation error standard deviation (in °C) used for assimilation.

2.5 Ensemble Optimal Interpolation

The Ensemble Optimal Interpolation (EnOI, Oke et al., 2002; Evensen, 2003) is a computationally cheap sequential data assimilation method derived from the Ensemble Kalman Filter (EnKF, Evensen, 2003).

The EnOI provides multivariate updates based on the model's historical covariance. Formulating the covariance from the same model ensures the preservation of linear quantities (such as geostrophic balance) and limits initialisation shock (Counillon & Bertino, 2009). The covariance matrix is constructed based on an ensemble of N model snapshots (each of dimension n, where n is the state dimension):

$$\mathbf{X}^s = [\mathbf{x}_1, ..., \mathbf{x}_N] \in \mathcal{R}^{n\times N}. \tag{1}$$

The static ensemble anomaly $\mathbf{A}^s$ is calculated so that $\mathbf{A}^s = \mathbf{X}^s - \overline{\mathbf{X}^s}\,\mathbb{1}^{\mathrm{T}}$, with $\overline{\mathbf{X}^s}$ being the static ensemble mean and $\mathbb{1}_m = [1, 1, ..., 1] \in \mathcal{R}^{1\times N}$.

To correct a forecast $\mathbf{x}^f$, using the observation vector $\mathbf{d}$, one can estimate a new analysis state $\mathbf{x}^a$ as follows:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f). \tag{2}$$

where $\mathbf{H}$ is the observation operator that relates the prognostic model state variables to the measurements. The Kalman Gain $\mathbf{K}$, is computed as follows:

$$\mathbf{K} = \mathbf{A}^s\mathbf{A}^{s\mathrm{T}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{A}^s\mathbf{A}^{s\mathrm{T}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}. \tag{3}$$

For each model, the static ensemble is composed of the monthly snapshot outputs from a stable pre-industrial control run. Sampling the model states from a pre-industrial condition was preferred over sampling it from a historical run because composing the static ensemble from a model with transient forcing can introduce spurious correlation. The monthly static ensemble comprises 72 members for MPI-ESM, 80 for NorESM, and 80 for CESM. We use a local framework analysis (Evensen, 2003) with a radius of 235 km without tapering — this radius ensures at least one observation per grid cell. We limit the local observations to one (retaining only the nearest). We update all prognostic state variables in the vertical. The update is done in the model's native coordinates — i.e., in geopotential depth for MPI-ESM and CESM and isopycnal coordinates for NorESM. We use the k-factor formulation (Sakov et al., 2012) which artificially inflates the observation error if the assimilation pushes the update beyond twice the ensemble spread. We set the pseudo-observation error for both supermodel approaches as the pointwise de-seasoned (with the mean seasonal cycle removed) time standard deviation of the three models divided by 30 (see Figure 2). We have played with the scaling factor (from 3 to 100), but the results were not very sensitive to this choice (not shown) due to the k-factor formulation.

3 Validation data sets and metrics

For validating the simulations, we focus on SST and use the NOAA OI-SST V2 (Reynolds et al., 2002) analysis data set. We use the monthly averaged product available on a 1° grid, which extends back to 1982. We assess performance by comparing the climatological difference between the models and the observation and calculating grid cell area-weighted root mean square error (RMSE).

Internal variability can be suppressed when combining models, as for example through the averaging of multi-instance fluxes in the interactive ensemble framework (Kirtman & Shukla, 2002; Kirtman et al., 2004; W. Zhang & Kirtman, 2019). Hence, we introduce two metrics to investigate how internal variability is affected by the connection. The parameter δ is the ratio between the de-seasoned time standard deviation of the multi-model mean and the time average of the inter-model standard deviation. This metric is computed for every grid cell (see Equation 4). It can be demonstrated that this quantity should be equal to $\sqrt{\frac{1}{N_s-1}}$. With three models, this is about 0.7. If δ is larger than 0.7, some synchronisation is achieved. For example, a value of three indicates that the time standard deviation of the multi-model mean is three times larger than the inter-model standard deviation. Values lower than 0.7 can occur when the bias of the individual models is larger than the time variability of the multi-model mean (i.e., models are strongly attracted to their bias and have little variability).

$$\delta = \frac{\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(\overline{\mathbf{x}_i}^s - \overline{\overline{\mathbf{x}}^s}^t)^2}}{\frac{1}{N_t}\sum_{i=1}^{N_t}\sqrt{\frac{1}{N_s}\sum_{j=1}^{N_s}(\mathbf{x}_i^j - \overline{\mathbf{x}_i}^s)^2}} \tag{4}$$

With $\mathbf{x}$ being the model SST. The superscript j refers to the model indices (N_s=3; between 1–3), the superscript i is the time indices (N_t=312; i.e., 26*12). The overbar with superscript "s" denotes the multi-model mean, and the overbar with superscript "t" denotes the time average.

A second metric, called λ, quantifies the ratio between the multi-model mean time variability (denominator) and the mean of the variability in the unconnected individual models (numerator; Eq. 5). In a perfect synchronisation regime, there is no damping (Duane & Tribbia, 2001). Assuming that the variability among the individual models is comparable, an equal-weight multi-model mean would have comparable variability to the mean variability of the individual models (λ should be one). If only partial synchronisation is achieved, the value gets larger than one. For example, λ equal to 2 means that the standard deviation of the multi-model mean is half that of the unconnected model.

$$\lambda = \frac{\overline{\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(\mathbf{x}_i^j - \overline{\mathbf{x}_i}^t)^2}}^S}{\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(\overline{\mathbf{x}_i}^s - \overline{\overline{\mathbf{x}}^s}^t)^2}} \tag{5}$$

4 Supermodel results

We compare the performance of EW, SINGLE and NI (see Section 2.4) over the period 1982–2006. In Figure 3, we show the de-seasoned time standard deviation of the multi-model mean of NI, EW and SINGLE and the OISST observations. The time variability in the NI multi-model mean is substantially lower than in the observations. In EW, it is also reduced—albeit less than in the NI—, while SINGLE has nearly comparable amplitude to the observations.

In Figure 4, we are analysing the properties of each system in achieving synchronisation and causing damping. We present individual 2-dimensional maps of δ and λ (metrics introduced in Section 3) and the 2-dimensional probability density function (PDF) of δ versus λ. An ideal supermodel will have the highest probability centred on the brown line and on the right-hand side of the red line (i.e., $\delta > .7$).

We can see that δ for NI is lower or equal to 0.7 in most places, and values of λ are above one and mostly close to two. There is barely any synchronisation, and taking the multi-model

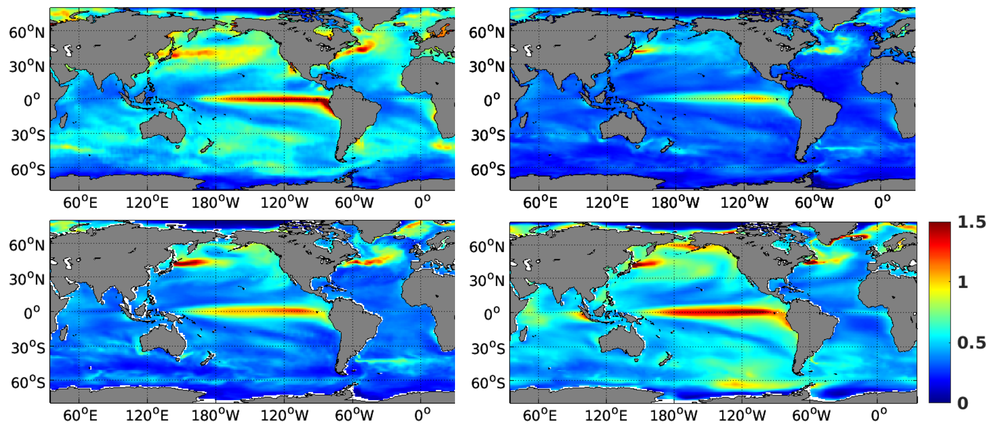

Figure 3. De-seasoned time standard deviation of SST in the NOAA OISST2 observations (upper left panel), the multi-model average of NI (upper right), EW (lower left) and SINGLE (lower right).

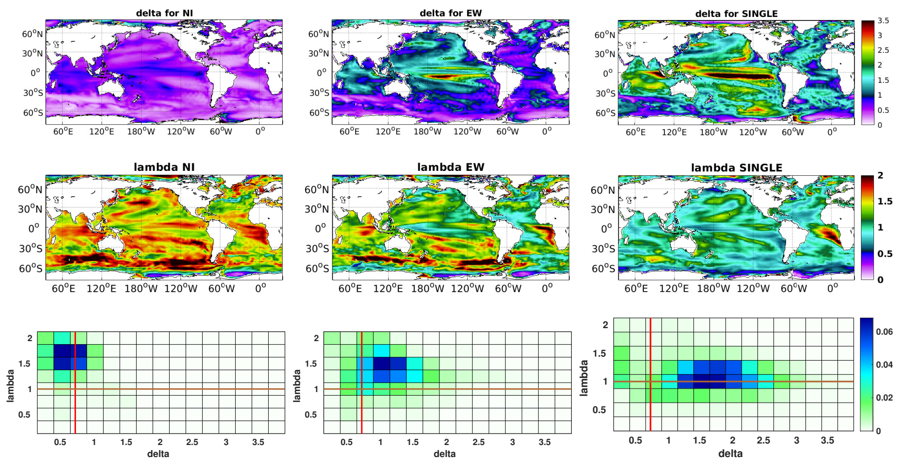

Figure 4. The first line is the synchronisation metric δ with NI mean, EW and SINGLE. The second line is the same for the synchronisation metric λ, and the last line is the 2-dimensional PDF of δ vs λ in the different systems. The red and brown lines highlight the threshold of 1.

mean causes damping nearly identical to a random process (i.e., $\sqrt{3}$ with three models). Hence, in NI, models are only connected via their historical CMIP5 forcing, while the internal variability overpowers the de-seasoned time variability compared to the climate change signal from 1980–2005.

In the EW supermodel, there is some synchronisation ($\delta > 0.7$) and the damping λ is mostly between [1.1 1.5]. The synchronisation and the damping are improved compared to NI, and the maximum likelihood moves towards the optimal. Synchronisation is maximum in the equatorial Pacific, reaching a value of 3.5 and is above 1 in the tropical Pacific, part of the North Pacific and part of the North Atlantic (entrance of the Nordic Seas). In those regions, λ lowers towards 1 (albeit remaining above). However, δ remains below or close to 0.7 in several regions: e.g., in the equatorial Atlantic and the eastern boundary upwelling system and most of the Southern oceans. No synchronisation is achieved there, and internal variability of the multi-model mean is damped and of comparable amplitude to NI. The shape of δ in the equatorial Pacific shows maxima on either side of the equator but achieves only moderate synchronisation at the equator. Hence, the fast Kelvin waves are driven by wind bursts in the equatorial Pacific, while atmospheric variability is poorly constrained in our ocean-constrained system. Slower Rossby waves control the thermocline variability off the equator, and the effect of winds is weaker and slower. Ocean data assimilation with monthly synchronisation steps can control such a process better.

The pattern of the value of δ with SINGLE resembles that of EW, but the values are considerably larger. Synchronisation δ is greater than 1.5, and the damping λ is reduced (approaching one and rarely exceeding 1.5). Actually, the CESM model (to which all models are connected) has a lower time standard deviation than MPI-ESM and NorESM, while the denominator in λ is the average of the time standard deviation of the three models. It partly explains the slight damping in λ. Synchronisation is now achieved in most parts of the southern oceans, the Atlantic Ocean, and the Indian Ocean. In the equatorial Atlantic and the eastern boundary upwelling system, synchronisation is still not achieved, and the damping is substantial. This region is notorious for being very challenging for models (Richter et al., 2014) and a considerable fraction of the bias relates to atmospheric origin (e.g., wind bias Koseki et al., 2018).

The above analysis focused on the time standard deviation of the multi-model mean. We analyse now the de-seasoned time standard deviation of the individual models. In Figure 5 we show the quantile-quantile plot of CESM-NI versus CESM-EW and CESM-SINGLE. Focusing on CESM allows us discarding discrepancies of variability between models, e.g., CESM has weaker temporal variability than the two other models. In EW, the time variability of CESM is reduced. The other models show comparable damping (not shown). Contrarily, SINGLE does not cause deflation, and the regression line yields a slight overestimation. It may relate to the fact that assimilating pseudo-observations in a common grid adds energy to the system (because of the imbalance). A perfect fit is obtained if pseudo-observations are formulated in the CESM native grid (not shown).

In Figure 6, we analyse the SST bias of the individual models compared to OISSTV2 observations. In NI, CESM has a lower bias than MPI-ESM and NorESM. NorESM has a pronounced cold bias. In the EW supermodel, the bias of all models is reduced compared to that of NI (albeit identical in MPI-ESM). Achieving such a result is highly promising, considering that weights have not been trained. In SINGLE, the bias structure in all models resembles that of CESM in NI. It yields a bias reduction for MPI-ESM and NorESM but increases the bias in CESM compared to NI. The interpolation to the common grid causes the degradation (not shown).

The bias of the multi-model mean of EW and NI (see Figure 7) are comparable in pattern and amplitude (NI has a slightly lower area-weighted RMSE). Still, some discrepancies exist. Figure 8 shows the absolute difference between the two. The bias of EW reduces in the tropics, except in the equatorial Atlantic and the eastern boundary upwelling system. Improvement tend to coincides with regions where synchronisations is achieved (i.e., $\delta > 1$ and $\lambda < 1.5$, see Figure 4). In the equatorial Pacific (5°S–5°N and 150°W–90°W), the error reduces from 0.91 to 0.67°. A bias reduction compared to the a-posteriori average results from the non-linear response of the

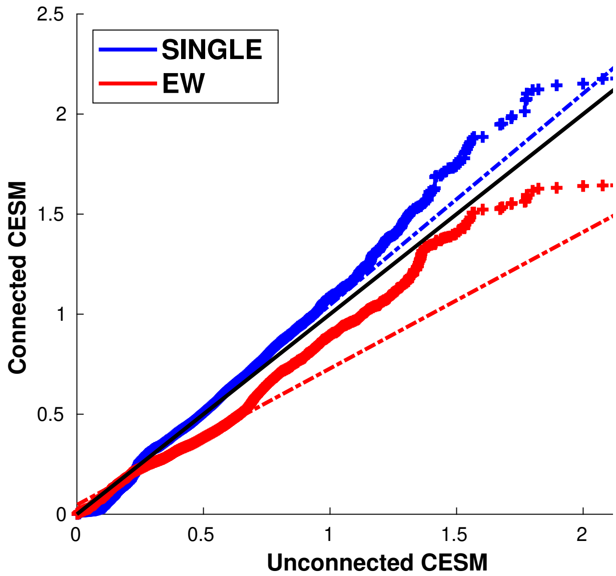

Figure 5. Quantile-quantile plot of the pointwise de-seasoned time standard deviation of CESM in NI versus that of EW (red) and SINGLE (blue). The dashed colour line show the regression line and the solid black line shows the perfect regression line.

climate system to model bias. Contrarily, performance degrades where synchronisation is poor (Antarctic Circumpolar current, Equatorial Atlantic and eastern boundary upwelling system). There, ocean synchronisation is quickly lost, and noise adds energy to the system. In the north Pacific, synchronisation is achieved, but the bias is degraded. The reason is unclear, but we propose several hypotheses. First, variability there during winter is driven by storm systems mixing the ocean. The EW experiment without atmospheric synchronisation effectively damps the impact of fluxes in the ocean, which can cause a bias. Second, the Pacific Decadal Oscillation (PDO) over 1982-2006 is almost solely positive, while the interannual variability of our models is not synchronised with the observations. The bias may relate to a modulation of the PDO. Third, we did not apply synchronisation under sea ice, which may cause some artefacts near the ice edge.

The bias in the SINGLE multi-model mean has a comparable spatial pattern than in the CESM model of NI, but is larger because of the interpolation error in the pseudo-observation grid. It does not outperform the multi-model mean of NI.

We further analyse the results in the ENSO region, which stands out as one of the regions where our ocean-connected supermodel reaches good synchronisation and reduced bias. The time series of the Niño 3.4 index (SST anomaly in the region, 5°S–5°N and 120°W–170°W) is presented in Figure 9. In the unconnected model (Figure 9), the individual models produce large ENSO variability that is not synchronised. The ensemble mean shows one prominent peak when a strong El Niño event occurs (by chance) in phase in 2003 in all models. The probability of such a coincidental occurrence scale with the power of the number of models. In our example (with three models and El Niño occurring typically every four years), such an event will happen every 64 years. Consequently, the NI methods will underestimate extreme events, particularly with more models (as in CMIP). Contrarily, in EW and SINGLE, all models are evolving in phases.

To analyse the representation of extreme events of ENSO, we compare the PDF of the Niño 3.4 index for 1982-2006 in Figure 10. The tropical Pacific variability is notoriously asymmetric, with the magnitude of SST anomalies over the eastern equatorial Pacific being more prominent during the warm phase than during the cold phase (e.g., T. Zhang et al., 2017). By taking the mean of unconnected and independent models, one expects the PDF to show less variabil-

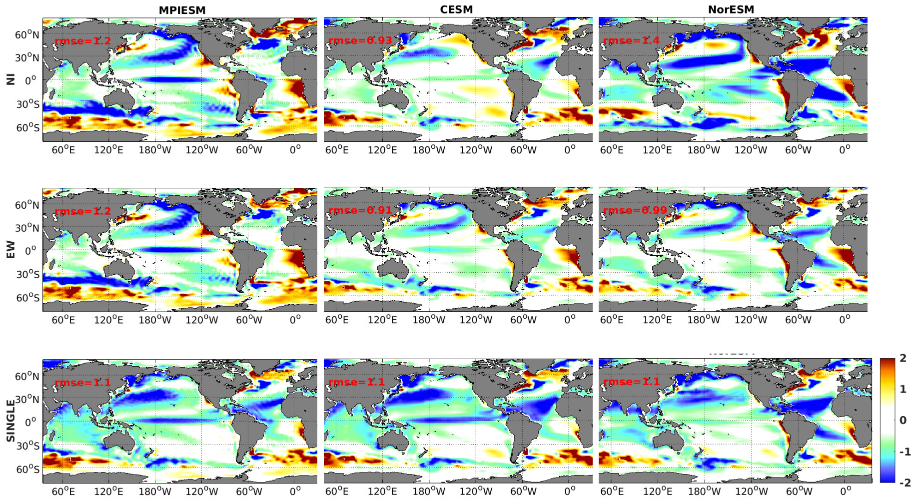

Figure 6. SST climatological bias computed over 1982-2006 for MPI-ESM, CESM and NorESM with the NI ensemble (first row) in the EW (second row) and the SINGLE (last row). The quantity in red reports the global spatial RMSE normalised by grid cell area.

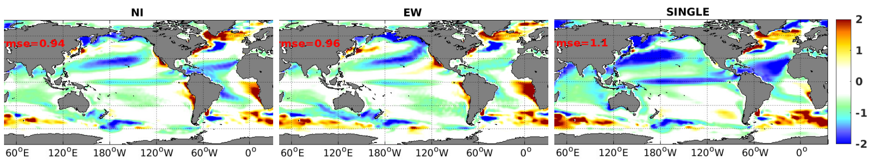

Figure 7. Climatological SST bias computed over 1982–2006 for the multi-model mean of NI ensemble (left), EW (middle) and SINGLE (right). The quantity in red reports the global spatial RMSE normalised by grid cell area.

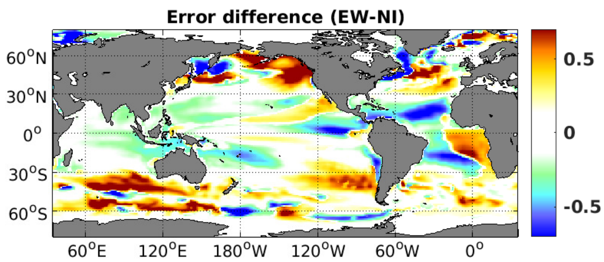

Figure 8. Difference of the absolute (EW-NI) of the climatological SST bias error of the multi-model mean over 1982–2006. Negative values indicate that the error in EW is smaller than in NI.

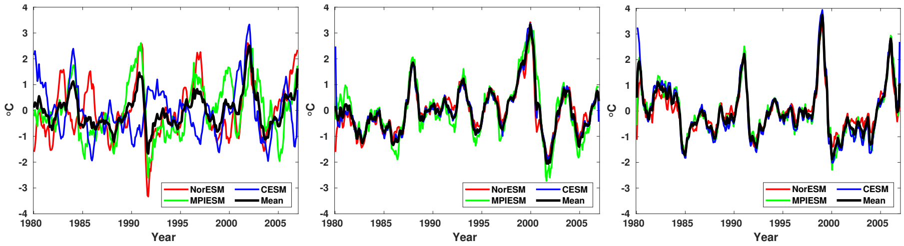

Figure 9. Niño 3.4 time series of the three models in the NI ensemble (left), the EW (middle) and the SINGLE (right). The black line shows the multi-model ensemble mean.

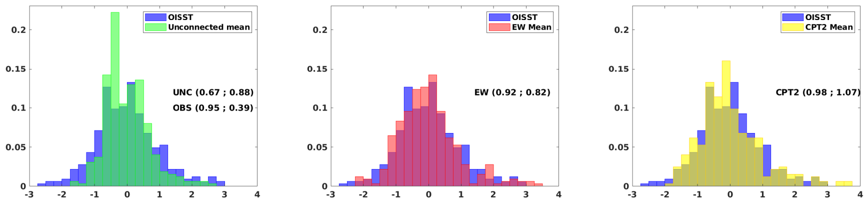

Figure 10. Histogram of Niño 3.4 over 1982–2006. The blue bars are for OISSTV2; the green bars are for the multi-model mean of NI (left), the red for EW (middle) and the yellow for SINGLE. The standard deviation and skewness of the distribution are reported in parenthesis for observations and the multi-model mean.

ity (being steeper around 0) and becoming more Gaussian (getting less skewed). The latter is because the average of skewed distributions will converge towards a Gaussian distribution (central limit theorem). As expected, there is damping of variability in the NI model mean, while in EW and SINGLE, the standard deviation reasonably matches the observed PDF. All multi-model means show a more skewed PDF than in the observations. However, the period 1982-2006 was quite anomalous in the observations, as no significant ENSO events occurred beyond 1998. In comparison, the standard deviation and skewness computed for 1950-2005 were respectively 0.87, and 0.89 (T. Zhang et al., 2017). We do not see the reduction of skewness expected in NI, but we think it is artificially high here because of the coincidental El Niño event referred to above. Obtaining more statistically robust results requires much longer runs and thus we limit our analysis to these basic metrics (Wittenberg, 2009).

5 Conclusions and future perspectives

This paper investigates critical characteristics for developing the first supermodel of ESMs with ocean synchronisation. Synchronisation with models having different grids and structures is handled by assimilating synthetic pseudo-observations of SST constructed from the multi-model every month. We show that such a framework can achieve partial synchronisation in distinct regions where the ocean drives the climate variability — where the impact of atmospheric fluxes is limited in relation to the oceanic timescales.

We compared two methodologies for constructing the pseudo-observations, either from a single model or an equal-weight mean. The latter tends to reduce model bias in the synchronised region compared to the unconnected version. The variability of the multi-model mean in the two connected models is smaller than in the observations. This damping is reduced in synchronised regions and converges to the non-interactive multi-model mean in the unsynchronised regions. The damping is more pronounced in the equal-weight version than in the version connected to a single model, as in the former, the variability of the individual models is also reduced. The multi-model averaging reduces the unsynchronised variability (e.g., driven by chaotic atmospheric variability) in the assimilated pseudo-observations causing a deflation when updating the model snapshots.

We will investigate in a future study how model error can reduce with the training of the weights. A first supermodel with monthly ocean connection via SST and with trained weights (varying spatially and monthly) will be presented in Schevenhoven et al. (in prep). The system constrains the SST bias and the double ITCZ problem in the equatorial Pacific. However, the system faces similar challenges regarding variability damping to those presented in this paper.

The system presented only uses a minimal amount of data exchange (monthly SST). Increasing the frequency of the synchronisation steps and assimilating more pseudo-observations will enhance synchronisation and may help to reduce the damping in the weighted mean supermodel framework. Ongoing works include a complete ocean pseudo-observations network (sea surface elevation, 3D hydrography), increasing the frequency of the ocean connection (to weekly synchronisation step) and complementing the system with synchronisation of other components of the Earth System Model. A supermodel with complementary synchronisation of the atmospheric component is also in development.

It is indeed unclear to speculate which of the two framework presented here will, at term, achieve the best performance. Weighted-mean supermodels have drawbacks in a partial synchronisation regime (variability damping) not faced with a connection to a single model. However, a weighted supermodel can provide locally a more accurate fit to the observation than a version where models are connected to a single model because — in the latter, skill is bounded by the best model. However, the single model to which we connect can vary spatially and for different variables (Smith, 2001; Du & Smith, 2017; Schevenhoven & Carrassi, 2021). Furthermore, if one can afford an ensemble of supermodels (with several members for each model), models could be synchronised from a randomly drawn single member/model every time, so that the frequency of the optimal weight is satisfied. Such a scheme resembles the cross-pollination in time, envisioned in Du and Smith (2017).

We will explore other alternatives to reduce the spurious damping in the weighted mean supermodel version by 1) adding back the reduced atmospheric driven variability term in the pseudo-observations or 2) isolating the part of variability that can be synchronised from the model snapshots (observation operator). In particular, machine learning techniques have emerged as powerful tools to carry process-identification (e.g., Sonnewald et al., 2019).

6 Open Research

Data presented in this article has been organised and made available. It contains a separate for each of the experiments presented (SINGLE, EW and NI). Each folder contains the multi-model mean and the individual SST outputs from the individual models provided on a common grid and in netcdf format. The full simulations will be made available on https://archive.sigma2.no with a specific doi upon acceptance of the manuscript. To retrieve the simulations, one can use the following link (457 MB):

wget -c http://ns9039k.web.sigma2.no/Synchronisation_supermodel.tar.gz

Acknowledgments
We acknowledge Cecile Hannay for providing the CESM initial condition and Gokhan Danabasoglu for sharing the Code needed only to update one time level in POP2. This study was partly funded by ERC-STERCP (#648982) and from the Trond Mohn Foundation, under project number: BFS2018TMT01. This work has also received a grant for computer time from the Norwegian Program for supercomputing (NOTUR2, project number nn9039k) and a storage grant (NORSTORE, NS9039k).

References

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., ... Kristjánsson, J. E. (2012). The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation. *Geosci. Model Dev. Discuss.*, *5*, 2843–2931. doi: 10.5194/gmdd-5-2843-2012

Block, K., & Mauritsen, T. (2013). Forcing and feedback in the mpi-esm-lr coupled model under abruptly quadrupled co2. *Journal of Advances in Modeling Earth Systems*, *5*(4), 676–691.

Branicki, M., & Majda, A.-J. (2015). An information-theoretic framework for improving imperfect dynamical predictions via multi-model ensemble forecasts. *Journal of Nonlinear Science*, *25*(3), 489–538.

Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, *9*(5), e535.

Counillon, F., & Bertino, L. (2009). Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico. *Tellus A*, *61*(2), 296–308.

Du, H., & Smith, L.-A. (2017). Multi-model cross-pollination in time. *Physica D: Nonlinear Phenomena*, *353*, 31–38.

Duane, G.-S., & Tribbia, J.-J. (2001). Synchronized chaos in geophysical fluid dynamics. *Physical review letters*, *86*(19), 4298.

Duane, G.-S., Wiegerinck, W., Selten, F., Shen, M.-L., & Keenlyside, N. (2018). Supermodeling: synchronization of alternative dynamical models of a single objective process. In *Advances in nonlinear geosciences* (pp. 101–121). Springer.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Fereday, D., O'Reilly, C., ... others (2018). Skilful seasonal predictions of summer european rainfall. *Geophysical Research Letters*, *45*(7), 3246–3254.

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, *53*(4), 343–367.

Eyring, V., Bony, S., Meehl, G.-A., Senior, C.-A., Stevens, B., Stouffer, R.-J., & Taylor, K.-E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Giorgetta, M.-A., Jungclaus, J., Reick, C.-H., Legutke, S., Bader, J., Böttinger, M., ... others (2013). Climate and carbon cycle changes from 1850 to 2100 in mpi-esm simulations for the coupled model intercomparison project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597.

Halem, M., & Dlouhy, R. (1984). Observing system simulation experiments related to space-borne lidar wind profiling. part 1: forecast impacts of highly idealized observing systems. *Res. Rev., 1983*.

Hurrell, J., Holland, M., Gent, P., Ghan, S., Kay, J., Kushner, P., ... others (2013). S. marshall (2013). the community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9).

Ilyina, T., Six, K.-D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model hamocc: Model architecture and performance as component of the mpi-earth system model in different cmip5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, *5*(2), 287–315.

Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S., Dance, S.-L., ... Weston, P. (2018). On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *144*(713), 1257–1278.

Jungclaus, J., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., ... Von Storch, J. (2013). Characteristics of the ocean simulations in the max planck institute ocean model (mpiom) the ocean component of the mpi-earth system model. *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446.

Karspeck, A.-R., Danabasoglu, G., Anderson, J., Karol, S., Collins, N., Vertenstein, M., … others (2018). A global coupled ensemble data assimilation system using the community earth system model and the data assimilation research testbed. *Quarterly Journal of the Royal Meteorological Society*, *144*(717), 2404–2430.

Kay, J.-E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., … others (2015). The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349.

Kirkevåg, A., Iversen, T., Seland, Ø., Hoose, C., Kristjánsson, J., Struthers, H., … others (2013). Aerosol–climate interactions in the Norwegian Earth System Model–NorESM1-M. *Geoscientific Model Development*, *6*(1), 207–244.

Kirtman, B.-P., Pegion, K., & Kinter, S.-M. (2004). *Internal atmospheric dynamics and tropical indo-pacific climate variability* (Vol. 165). Center for Ocean-Land-Atmosphere Studies.

Kirtman, B.-P., & Shukla, J. (2002). Interactive coupled ensemble: A new coupling strategy for cgcms. *Geophysical research letters*, *29*(10), 5–1.

Koseki, S., Keenlyside, N., Demissie, T., Toniazzo, T., Counillon, F., Bethke, I., … Shen, M.-L. (2018). Causes of the large warm bias in the angola–benguela frontal zone in the norwegian earth system model. *Climate Dynamics*, *50*(11), 4651–4670.

Krishnamurti, T., Kumar, V., Simon, A., Bhardwaj, A., Ghosh, T., & Ross, R. (2016). A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. *Reviews of Geophysics*, *54*(2), 336–377.

Lamarque, J.-F., Bond, T.-C., Eyring, V., Granier, C., Heil, A., Klimont, Z., … others (2010). Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmospheric Chemistry and Physics*, *10*(15), 7017–7039.

Marsland, S. J., Haak, H., Jungclaus, J.-H., Latif, M., & Röske, F. (2003). The max-planck-institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean modelling*, *5*(2), 91–127.

Mirchev, M., Duane, G.-S., Tang, W. K., & Kocarev, L. (2012). Improved modeling by coupling imperfect models. *Communications in Nonlinear Science and Numerical Simulation*, *17*(7), 2741–2751.

Nerger, L., Tang, Q., & Mu, L. (2020). Efficient ensemble data assimilation for coupled models with the parallel data assimilation framework: example of awi-cm (awi-cm-pdaf 1.0). *Geoscientific Model Development*, *13*(9), 4305–4321.

Oke, P.-R., Allen, J.-S., Miller, R.-N., Egbert, G.-D., & Kosro, P.-M. (2002). Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research: Oceans*, *107*(C9), 5–1.

Palmer, T., & Stevens, B. (2019). The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences*, *116*(49), 24390–24395.

Pecora, L.-M., Carroll, T.-L., Johnson, G.-A., Mar, D.-J., & Heagy, J.-F. (1997). Fundamentals of synchronization in chaotic systems, concepts, and applications. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *7*(4), 520–543.

Reick, C., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in mpi-esm. *Journal of Advances in Modeling Earth Systems*, *5*(3), 459–482.

Reynolds, R.-W., Rayner, N.-A., Smith, T.-M., Stokes, D.-C., & Wang, W. (2002). An improved in situ and satellite sst analysis for climate. *Journal of climate*, *15*(13), 1609–1625.

Richter, I., Behera, S.-K., Doi, T., Taguchi, B., Masumoto, Y., & Xie, S.-P. (2014). What controls equatorial atlantic winds in boreal spring? *Climate dynamics*, *43*(11), 3091–3104.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., & Korablev, A. (2012). TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, *8*, 633–656.

Scaife, A.-A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1), 1–8.

Schevenhoven, F., & Carrassi, A. (2021). Training a supermodel with noisy and sparse observations: a case study with cpt and the synch rule on speedo-v. 1. *Geoscientific Model Development Discussions*, 1–23.

Schevenhoven, F., Keenlyside, N., Carrassi, A., Counillon, F., Devilliers, M., Koseki, S., ... Duane, G. (in prep). Supermodeling: improving predictions with an ensemble of interacting models. *BAMS*.

Schevenhoven, F., & Selten, F. (2017). An efficient training scheme for supermodels. *Earth System Dynamics*, *8*(2), 429–438.

Schevenhoven, F., Selten, F., Carrassi, A., & Keenlyside, N. (2019). Improving weather and climate predictions by training of supermodels. *Earth System Dynamics*, *10*(4), 789–807.

Schneck, R., Reick, C.-H., & Raddatz, T. (2013). Land contribution to natural co2 variability on time scales of centuries. *Journal of Advances in Modeling Earth Systems*, *5*(2), 354–365.

Selten, F.-M., Schevenhoven, F.-J., & Duane, G.-S. (2017). Simulating climate with a synchronization-based supermodel. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126903.

Shen, M.-L., Keenlyside, N., Bhatt, B.-C., & Duane, G.-S. (2017). Role of atmosphere-ocean interactions in supermodeling the tropical pacific climate. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126704.

Shen, M.-L., Keenlyside, N., Selten, F., Wiegerinck, W., & Duane, G. S. (2016). Dynamically combining climate models to ?supermodel? the tropical pacific. *Geophysical Research Letters*, *43*(1), 359–366.

Shukla, J. (1998). Predictability in the midst of chaos: A scientific basis for climate forecasting. *science*, *282*(5389), 728–731.

Smith, L.-A. (2001). Disentangling uncertainty and error: On the predictability of nonlinear systems. In *Nonlinear dynamics and statistics* (pp. 31–64). Springer.

Sonnewald, M., Wunsch, C., & Heimbach, P. (2019). Unsupervised learning reveals geography of global ocean dynamical regions. *Earth and Space Science*, *6*(5), 784–794.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., ... others (2013). Atmospheric component of the mpi-m earth system model: Echam6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485.

Tian, B., & Dong, X. (2020). The double-itcz bias in cmip3, cmip5, and cmip6 models based on annual mean precipitation. *Geophysical Research Letters*, *47*(8), e2020GL087232.

Tjiputra, J., Roelandt, C., Bentsen, M., Lawrence, D., Lorentzen, T., Schwinger, J., ... Heinze, C. (2013). Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geoscientific Model Development*, *6*(2), 301–325.

Van den Berge, L.-A., Selten, F.-M., Wiegerinck, W., & Duane, G.-S. (2011). A multi-model ensemble method that combines imperfect models through learning. *Earth System Dynamics*, *2*(1), 161–177.

Vertenstein, M., Craig, T., Middleton, A., Feddema, D., & Fischer, C. (2012, 12). CESM1.0.3 User Guide. *UCAR Doc.* (Available at: http:www.cesm.ucar.edumodelscesm1.0cesmcesm_doc_1_0_3ug.pdf)

Wang, Y., Counillon, F., Keenlyside, N., Svendsen, L., Gleixner, S., Kimmritz, M., ... Gao, Y. (2019). Seasonal predictions initialised by assimilating sea surface temperature observations with the enkf. *Climate Dynamics*, *53*(9-10), 5777–5797.

Wiegerinck, W., Burgers, W., & Selten, F. (2013). On the limit of large couplings and weighted averaged dynamics. In *Consensus and synchronization in complex networks* (pp. 257–275). Springer.

Wiegerinck, W., & Selten, F. (2017). Attractor learning in synchronized chaotic systems in the presence of unresolved scales. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126901.

Wittenberg, A. T. (2009). Are historical records sufficient to constrain enso simulations? *Geophysical Research Letters*, *36*(12).

Zhang, S., Harrison, M., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, *135*(10), 3541–3564.

Zhang, T., Shao, X., & Li, S. (2017). Impacts of atmospheric processes on enso asymmetry: A comparison between cesm1 and ccsm4. *Journal of Climate*, *30*(23), 9743–9762.

Zhang, W., & Kirtman, B.-P. (2019). Estimates of decadal climate predictability from an interactive ensemble model. *Geophysical Research Letters*, *46*(6), 3387–3397.

Zhongming, Z., Linong, L., Wangqiang, Z., Wei, L., et al. (2020). Europe is building a 'digital twin' of earth to revolutionize climate forecasts. *Science*.

Zhu, J., Kumar, A., Lee, H.-C., & Wang, H. (2017). Seasonal predictions using a simple ocean initialization scheme. *Climate Dynamics*, *49*(11), 3989–4007.