Training of supermodels - in the context of weather and climate forecasting

Francine Schevenhoven

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

Training of supermodels - in the context of weather and climate forecasting

Francine Schevenhoven

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 08.02.2021

Year: 2021

Title: Training of supermodels - in the context of weather and climate forecasting

Name: Francine Schevenhoven

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

The research presented in this thesis forms part of the Synchronization to Enhance Reliability of Climate Prediction (STERCP) European Union Horizon 2020 project, an international collaboration between partners in both Norway and the Netherlands aiming to further develop the climate supermodeling approach.

The main part of my doctoral research has been carried out at the Geophysical Institute of the University of Bergen (UiB). Through the Bjerknes Centre for Climate Research (BCCR), the Bergen-based research collaboration of about 200 climate scientists, I was able to quickly connect with the Nansen Environmental and Remote Sensing Center (NERSC) where I spent a very pleasant part of my working days. The research for this doctoral thesis all started with a master thesis internship at the Royal Netherlands Meteorological Institute (KNMI) in De Bilt, the Netherlands. During my PhD I frequently returned to KNMI and furthermore I was kindly allowed to use their computational and storage resources remotely throughout my entire PhD. Apart from my time in Norway and the Netherlands I spent one month of my PhD at the Institute of Atmospheric Physics (IAP) in Beijing, China and unfortunately (due to the 2020 corona crisis) only a little time at the Data Assimilation Research Centre at the University of Reading, United Kingdom.

Apart from the different institutes, I was involved in the national Research school on Changing climates in the coupled earth system (CHESS), which provided many opportunities to meet other PhD climate scientists in Norway during several courses and conferences. Outside Norway I got the chance to visit many conferences, summer schools and workshops, such as the European Geophysical Union Annual Meeting (EGU), the Society for Industrial and Applied Mathematics on Application of Dynamical Systems conference (SIAM DS), the Climate Informatics (CI) workshop and the Data Assimilation Summer school.

CHESS
Research school on changing climates in the coupled earth system

UNIVERSITAS BERGENSIS

BJERKNES CENTRE
for Climate Research

Acknowledgements

First and foremost I would like to thank the person who sparked my love for supermodeling, my supervisor Frank Selten. I am immensely grateful to you for all your support throughout the years; without you this thesis would not have existed. Secondly, I would like to thank Noel Keenlyside for giving me the opportunity to come to Bergen, for creating such a positive and friendly atmosphere in our group, and for our very pleasant discussions that sharpened my thinking. Thirdly, I want to express my gratitude towards Alberto Carrassi, for your involvement and enthusiasm, your always fast reply to pieces I sent you, and for teaching me so much about research and writing. I would also like to thank the other members of the STERCP project, in particular Wim Wiegerinck and Greg Duane, for all the nice and insightful discussions. Additionally, I had the great pleasure of working with Hailiang Du, I hope I can visit you soon in Durham.

I spent one very enjoyable month of my PhD in Beijing, China, at the Institute of Atmospheric Physics. I am very grateful to Fei Zheng, Lily, Song, and Helen for your hospitality, for showing me so much of the city and for your attempts to teach me Chinese.

Over the years I have had a lovely work environment in our office Målestuen, which I owe to my neighbours Marion, Sébastien, Ashneel, and all my other office mates. In general, I would like to thank everyone at the Geophysical Institute for being such a welcoming and close-knit community. I am very happy that so many of you joined 'kakeklubben', the Norwegian cake club. Thank you all for the amazing cake every Monday and the fun conversations in Norwegian. Tusen takk Kjersti, for din talmodighet og villighet til å lære oss norsk. Norsken min har forbedret seg så mye takket være deg!

Living in Bergen has been a wonderful experience, thanks to many people. I would like to thank especially Marco, Jasmijn, Chinh, and Anne for all the hikes, dinners and drinks, and for having your door always open to me. Clemens, thank you for being such a kind and pleasant roommate, and for letting me stay in a stunningly beautiful apart-

ment. Furthermore, I had a great time helping to establish UiBdoc, the organization for all PhDs and postdocs in Bergen. Many thanks to my fellow board members and other participants for all the social and professional activities, I hope that many events will follow.

I very much appreciated the support of my friends and family in the Netherlands and in France. Thank you for your interest in my work and life in Bergen, and for facing the Bergen weather. Finally, the person who has stood most by my side all these years is my partner Paul. Thank you for your encouragement, your patience, and for visiting me so often. I am looking forward to our future together.

Francine Schevenhoven
Bergen, November 2020

Abstract

Given a set of imperfect weather or climate models, predictions can be improved by combining the models dynamically into a so called 'supermodel'. The models are optimally combined to compensate their individual errors. This is different from the standard multi-model ensemble approach (MME), where the model output is statistically combined after the simulations. Instead, the supermodel can create a trajectory closer to observations than any of the imperfect models. By intervening during the forecast, errors can be reduced at an early stage and the ensemble can exhibit different dynamical behavior than any of the individual models. In this way, common errors between the models can be removed and new, physically correct behavior can appear.

In our simplified context of models sharing the same evolution function and phase space, we can define either a connected or a weighted supermodel. A connected supermodel uses nudging to bring the models closer together, while in a weighted supermodel all model states are replaced at regular time intervals (i.e., restarted) by the weighted average of the individual model states. To obtain optimal connection coefficients or weights, we need to train the supermodel on the basis of historical observations. A standard training approach such as minimization of a cost function requires many model simulations, which is computationally very expensive. This thesis has focused on developing two new methods to efficiently train supermodels. The first method is based on an idea called cross pollination in time, where models exchange states during the training. The second method is a synchronization-based learning rule, originally developed for parameter estimation.

The techniques are developed on low-order systems, such as Lorenz63, and later applied to different versions of the intermediate-complexity global coupled atmosphere-ocean-land model SPEEDO. Here the observations are from the same models, but with different parameters. The applicability of the method to real observations is tested using sensitivity to noisy and incomplete data. The characteristics the individual models should have in order to be combined together into a supermodel are identified, as well as which physical variables should be connected in a supermodel, and which ones should not. Both training methods result in supermodels that outperform both the individual models and the MME, for short term predictions as well as long term simulations. Fur-

thermore, we show that the novel use of negative weights can improve predictions in cases where model errors do not cancel (for instance, all models are too warm with respect to the truth). A crucial advantage of the proposed training schemes is that in the present context relatively short training periods suffice to find good solutions. Although the validity of our conclusions in the context of real observations and model scenarios has yet to be proved, our results are very encouraging. In principle, the methods are suitable to train supermodels constructed using state-of-the art weather and climate models.

List of papers

I. Schevenhoven, F. J., and F. M. Selten, (2017), An efficient training scheme for supermodels, *Earth System Dynamics, 8(2)*, 429-438, doi:10.5194/esd-8-429-2017

II. Selten, F. M., F. J. Schevenhoven, and G. S. Duane, (2017), Simulating climate with a synchronization-based supermodel, *Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(12)*, 126,903, doi:10.1063/1.4990721

III. Schevenhoven, F., F. Selten, A. Carrassi, and N. Keenlyside, (2019), Improving weather and climate predictions by training of supermodels, *Earth System Dynamics, 10(4)*, 789-807, doi:10.5194/esd-10-789-2019

IV. Schevenhoven, F., and A. Carrassi, Training supermodels with noisy and sparse observations, *Manuscript in preparation*

V. Schevenhoven, F., and H. Du, A comparison of different approaches to combine models dynamically, *Manuscript in preparation*

Contents

Chapter 1

Introduction

1.1 Weather and climate modeling

Should I take my raincoat with me or not? The relevance of a good weather forecast is easy to grasp. With knowledge about specific (extreme) weather events, one can prepare oneself and try to prevent physical or property damage. The importance of a good climate forecast has become more recognized in the last decades. Our climate is not stable, and currently we are experiencing a clear change. It is well established that climate change is real. However, the magnitude of the change and the impact on society are still uncertain, and depend also on the actions that are taken.

To help to decrease the uncertainty, scientists continuously work to improve weather and climate models. The climate system is very complex, with interactions across scales over many orders of magnitude. The first climate models from the '70s consisted of only a large scale atmospheric component. Later, an ocean and land component were added, and many other components with increasingly smaller scales such as sea ice, aerosols, the carbon cycle, atmospheric chemistry etc. Since the models have a limited resolution, not all physical processes can be fully resolved. In these cases, modelers parametrize the effects of these processes. The different choices for parametrizations often result in different behavior of the models. Since there are still many unknown parameters and approximations in weather and climate models, it requires a huge knowledge and computational effort to model and tune all of these aspects. We are still a long way from being to resolve all important processes, and so huge uncertainties will remain for long-time to come.

Apart from the imperfect models, also the real-world observations have their imperfections. Using these observations as initial condition for a model forecast, automatically results in forecast errors. To assess and possibly reduce the problem of both imperfect models and imperfect observations, ensemble techniques have been developed in weather forecasting and climate prediction. Typically, ensembles are constructed by

perturbing the initial conditions and using different models. In this way a probabilistic forecast, a Multi-Model-Ensemble (MME), is generated to account for forecast uncertainties.

1.2 Multi-model ensemble (MME)

An example of a MME can be found in the assessments of the Intergovernmental Panel on Climate Change (IPCC). As can be seen in Fig. 1.1, there is no full consensus among the models on the amount of warming in response to the same increase in greenhouse gas concentrations. Often it helps to average across models of the MME, as model errors tend to average out. This is the case when some models are biased warm while other models are biased cold, compared to the 'correct' forecast. Although we do not know the correct forecast, we often expect errors to be randomly distributed, and so averaging would reduce them. In the IPCC reports, models are generally considered equal when combined. The individual performance of the models is not taken into account, 'bad' models get the same weight as 'good' models. Furthermore, it is not taken into account that models often share a common component, and hence common errors because they have not been developed independently (*Collins et al.*, 2013). Despite this rather simple approach, the multi-model mean has been demonstrated to be quite useful in the evaluation of the Coupled Model Intercomparison Project 5 (CMIP5) models. In addition, not only the mean from a multi-model ensemble is useful to get a better forecast, also the spread between the model predictions within the ensemble gives information about the uncertainty (*Collins et al.*, 2013).

Although the IPCC uses the multi-model approach in its assessments, the question remains why exactly and under which conditions the multi-model approach is beneficial. *Weigel et al.* (2008) tried to develop a theoretical framework to show the benefit of the multi-model mean. They concluded on the basis of Gaussian toy models that the multi-model approach can indeed enhance prediction skill and outperform the best individual model, provided that the individual models are overconfident. This means that the individual model ensemble forecasts have a too small range, while being centered at the wrong value. In that case the multi-model combination can widen the ensemble spread and move the ensemble mean towards the truth. Moreover, as long as it widens the ensemble spread in a right direction, also a 'bad' model can contribute towards the multi-model combination. *Weigel et al.* (2008) remark that this MME approach is better than simple ensemble inflation methods, where the ensemble spread is widened by a multiplication factor to a realistic spread in case the models are overconfident, which could destroy potential predictability.

Figure 1.1: Surface air temperature change in 2081-2100 for a collection of individual models, displayed as anomalies with respect to the period 1986-2005 (Figure 12.9 from the IPCC Fifth Assessment Report 2013, Collins et al. (2013)

The IPCC approach of an equally weighted mean might be improved by allowing weights to be non-identical for different models and variables, as well as allowing them to vary in space and time. This is the idea of a so-called 'superensemble' (*Krishnamurti et al.*, 2016). In *Krishnamurti et al.* (2016) the superensemble consists even of more than 10 million weights. Also negative weights are allowed in order to remove bias. An important innovation is that models can also be combined in terms of physical processes such as cumulus parameterization schemes. *Krishnamurti and Sanjay* (2003) designed a single unified model, with one new weighted parametrization scheme based on the parametrization schemes of the individual models. However, a full multi-model ensemble seems to further improve the prediction, due to the large reduction in systematic errors of the models.
Although the superensemble method of *Krishnamurti et al.* (2016) results often in better predictions, and while the results can be used to get a better understanding of error growth in individual models as well, the method suffers from some significant shortcomings. The weights for the superensemble are determined during a training phase, on the basis of the root mean squared error (RMSE) of the models with historical observations. If during this training phase certain events such as extrema have not been seen, then the statistical weights will probably not give a good forecast for these events (*Krishnamurti et al.*, 2016). Since the superensemble is a statistical approach rather than a dynamical one, it is very well possible that the dynamical behavior of the models has changed in the forecast phase. The statistically optimal weights from the training do not need to be optimal for the forecast (*Krishnamurti et al.*, 2016). An example of this is given in *Knutti et al.* (2010), where it is shown that models with a large historical temperature bias can give the same scenario for future warming as models with a small bias, implying that a statistical combination of the models on the basis of past performance does not need to be significant for the future. It would be better to go back to the root of the model runs and reduce the errors in the earliest possible stage of the runs by combining the models dynamically. Intervening this early can correct the dynamics of the individual models, resulting in physically better justified runs. This is precisely what supermodeling attempts to achieve.

1.3 Summary of Chapter 1 and outlook

Calculating a multi-model mean in order to combine different imperfect weather and climate models can be helpful in the case the models can compensate for each other's systematic errors. However, the MME approach is a statistical one, where the individual runs are completed individually and only combined afterwards. This means that

the average of the outputs does not necessarily represent a possible dynamical solution. Since the combined output has not been trained to improve the dynamics, it does not guarantee more reliable out of sample projections. Instead of training a statistical model combination, one can combine models dynamically in order to reduce model uncertainty. This dynamical combination of models is called a supermodel.
The training of a supermodel is the topic of this thesis. Before training, first the structure of the supermodel needs to be defined. The next chapter lays this foundation and summarizes the development of supermodeling throughout the years.

Chapter 2

Supermodeling

2.1 Toward supermodeling: Interactive ensemble

In the previous chapter we have seen that the standard MME approach is to combine the outputs of independently run models. Training on historical observations does not necessarily produce a well-trained MME forecast. Our climate system is a non-autonomous system due to the anthropogenic forcing, thus the optimal weights to combine historical model means might not be optimal to combine future model means, if the dynamics of the models are not appropriate. The interactive ensemble is a crucial new approach to combine models to better represent dynamics. An interactive ensemble is a multi-model ensemble where *during* the run the results of the different models are combined. The new model states formed thereby are used to produce new initial conditions with which the models can continue their run.

2.1.1 Multiple identical atmospheric models combined with one ocean model

The novel idea of intervening during a model run stems from *Kirtman and Shukla* (2002). In their paper, they address the error growth within a coupled atmosphere-ocean model. As long as a state-of-the-art Atmospheric Global Circulation Model (AGCM) receives a correctly prescribed sea surface temperature (SST), this boundary forcing is so strong that the interannual variability, especially above the tropical oceans, remains well modeled to a large extent despite any errors in the AGCM (*Richter et al.*, 2018). The same holds for an Ocean Global Circulation Model (OGCM) driven with heat fluxes and winds derived from historical reanalysis. Problems arise when the atmosphere is coupled to an ocean model. In a coupled atmosphere-ocean model the ocean model receives heat, moisture and momentum fluxes from the atmosphere. The other way around the atmosphere receives the SST from the ocean model, and uses it to compute the heat and moisture fluxes that also affect the atmosphere. Errors intrinsic to the AGCM (OGCM) lead to errors in the simulation of the heat fluxes (SST). In this

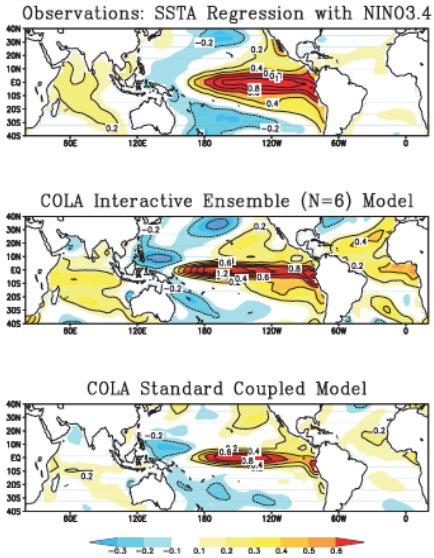


Figure 2.1: Linear regression between NINO3.4 time series and global SSTA (Kirtman and Shukla, 2002).

way, coupling an OGCM to an AGCM leads to an error in the SST, which feeds back onto the atmosphere, causing error growth. Poor parameterizations are major sources of such errors. This can lead in some cases (as in *Kirtman and Shukla* (2002)) that the AGCM simulates too vigorous internal atmospheric variability. This will affect the simulation of coupled ocean atmospheric variability, such as El Niño. The interactive ensemble is a way to reduce this effect. Since the internal variability of the atmosphere is operating at a weather time scale, it is straightforward to cancel out this variability by just averaging over multiple atmospheric trajectories; this strengthens the relative importance of atmospheric variations related to the SST. Hence *Kirtman and Shukla* (2002) created an ensemble out of one AGCM with six different synoptically independent initial conditions. The OGCM sees the ensemble average of the fluxes. In this case, the interactive ensemble performs clearly better than the standard technique of one atmosphere coupled to one ocean model in terms of ENSO variability, and global teleconnections associated with ENSO. See Fig. 2.1 for an example of the correlation between NINO3.4 time series and global SSTA for the interactive ensemble and a standard coupled model. Even today AGCMs may simulate too strong internal variability (*Scaife and Smith*, 2018), and the interactive ensemble approach could still be useful to improve the simulation of climate. The interactive ensemble is a good example of reducing the error by intervening during a model run.

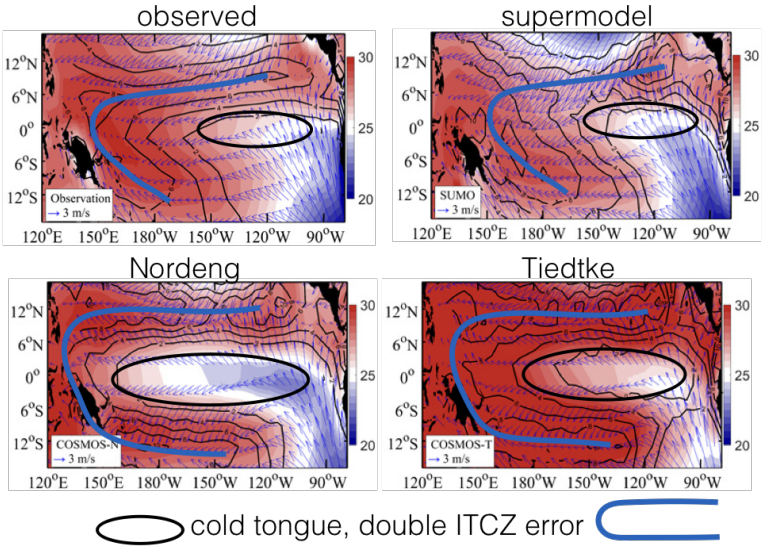


Figure 2.2: SST, precipitation and wind climatology (Shen et al., 2017).

2.1.2 Supermodeling: different atmospheric models coupled to one ocean model

Noise can be cancelled out by using the same atmospheric models, as we have seen in *Kirtman and Shukla* (2002), and this can improve simulated variability when the relative importance of atmospheric noise to coupled ocean-atmosphere variability is too large. But what if the climatology itself suffers from substantial errors? Then it could be useful to have different models that can compensate for each other's errors, as we have seen in the MME approach in Chapter. 1. Unfortunately, many climate models are subjected to the same type of model error, for example the double Intertropical Convergence Zone (ITCZ) and cold tongue bias in the tropical Pacific (*Lloyd et al.*, 2011; *Bellenger et al.*, 2014). Combining these models afterwards, as in MME, will hence not result in a better estimation of the climatology. Interestingly, these systematic errors in climatology do not necessarily have the same cause (*Shen et al.*, 2017). If the models have not taken too many time steps yet, the evolution of the model error still can be described as linear (*Carrassi and Vannitsem*, 2016) and the error in a certain variable does not impact other variables that much yet. This means that if the errors in climatology come from different sources, it might be possible to cancel out errors that emerge, if we correct the model errors in the beginning of the climatological run when the errors are still univariate.

The approach of *Shen et al.* (2016) is to use two different atmospheric models in an interactive ensemble. They used two different AGCMs, two versions of the COSMOS model (*Jungclaus et al.*, 2006), but with different convection schemes, *Tiedtke* (1989)

(COSMOS(T)) and *Nordeng* (1994) (COSMOS(N)). Both versions of the COSMOS model show a double ITCZ pattern (see Fig. 2.2). The interaction between the atmospheres and the ocean was built in the same way as in *Kirtman and Shukla* (2002), meaning the individual heat, moisture and momentum flux from both atmospheres were averaged and given to the ocean model. Instead of an equally weighted flux as in *Kirtman and Shukla* (2002), a weighted flux was calculated. The weights were trained by a simplex method (*Nelder and Mead*, 1965), where the error was minimised between 30 year climatological SST runs of the interactive ensemble and SST observations over a period of 30 years. The resulting combination of different models, a supermodel, (called SUMO in *Shen et al.* (2016)) showed indeed much better climatological behaviour in the tropical Pacific (see Fig. 2.2). The double ITCZ error was alleviated and in general there was an improvement of equatorial Pacific dynamics and the ENSO-induced anomalies. The reason for this improved behaviour compared to the individual models COSMOS(N) and COSMOS(T) lies within the fact that the models have different convection schemes and thereby simulate different wind and equatorial upwelling patterns. Combining these effects can compensate errors in the momentum fluxes, such that in SUMO a single ITCZ can be formed (*Shen et al.*, 2017).

2.1.3 Summary of Section 2.1

An interactive ensemble can improve the simulation of climate compared to the standard MME (see Fig. 2.3 for a schematic difference between the interactive ensemble and MME), for example when the internal variability is overestimated compared to the coupled dynamics, as in *Kirtman and Shukla* (2002). In this case, it was enough to remove the noise from the atmospheric models, but *Shen et al.* (2016) showed the usage of an interactive ensemble with different atmospheric models, a supermodel, in order to improve the climatology. With intervening during the simulation, errors can be reduced at an early stage and the ensemble can exhibit different dynamical behavior than any of the individual models. In this way, common errors between the models can be removed and new, more physically realistic behavior can appear.

2.2 A supermodel: synchronization between models

2.2.1 An atmospheric connected supermodel

Since the climatological SST of SUMO in *Shen et al.* (2016) is better than the resulting SST from any multi-model ensemble combination with weights between -1 and 2 (*Shen et al.*, 2016), there must have been a non-linear interaction between the COSMOS models during the run of SUMO that impacts the final result. The effect of these inter-

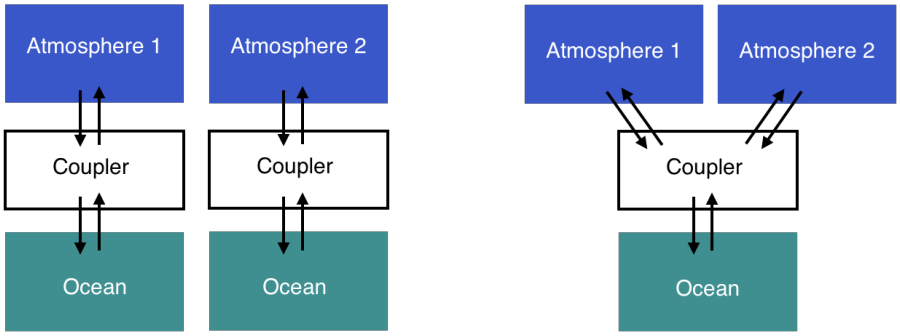


Figure 2.3: Difference between the MME approach (left) and the interactive ensemble approach with different atmospheric models, as in the supermodel of Shen et al. (2016) (right). The arrows indicate the exchange between the models during the simulation.

actions becomes mainly clear in the tropical Pacific, since the climatology of SUMO in this region is significantly different than the climatology of the individual imperfect models. In other areas of the globe the effect is less clear, there the SUMO climatology could be the result of an MME approach. In order to have effective interaction over the full state space, the models should exchange more variables, and this can lead to more different behaviour compared to the individual models. In *Shen et al.* (2016) only the atmospheric fluxes were combined and given to the ocean. A natural next step is to exchange variables between the atmospheric models themselves. Figure 2.4 shows the structure of this supermodel with the next level of sophistication, note the exchanges between the atmospheres as compared to the interactive ensemble of Fig. 2.3. Both atmospheres calculate their own values, exchange them, and agree which combination of values would be best to proceed with. This 'agreeing' can be better described in terms of synchronization. By synchronization we mean that the different runs are correlated in time, in contrast to independent model runs such as in MME. In a supermodel, individual model runs are substantially influenced by other model runs during the simulation, such that the effects of individual physical processes are at least partly synchronized.

2.2.2 Type of synchronization

Communication between models, and therefore synchronization to some extent between the models, is a feature that distinguishes the supermodel from the MME and hence allows the supermodel to be superior to the MME. Synchronization allows compensation of errors in dynamics, and if the models are synchronized enough, a new supermodel trajectory can be formed. The exact type of synchronization that is desir-

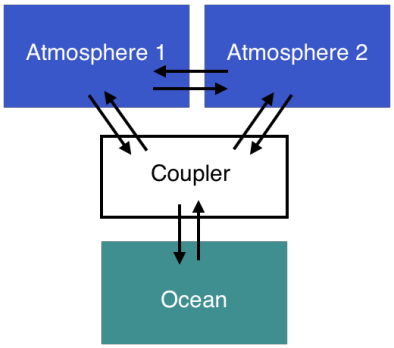


Figure 2.4: The next step in supermodeling: to combine the atmospheric models as well. The arrows indicate the exchange between the models during the simulation.

able in a supermodel depends on the individual models and also depends on the type of improvement that the supermodel is expected to deliver, as will be clarified in the next section. Nevertheless, a link can be made between the synchronization within a supermodel and synchronization in data assimilation. Within data assimilation there is a need for a consensus between a model and observations, while in a supermodel models should form a consensus with other models, but the idea of agreement remains the same. One way of achieving this is by diffusive coupling, also called nudging. This coupling has been applied to a Lorenz96 system and observations of the same system, in *Szendro et al.* (2009). More important than exact synchronization is the ability of the model to run freely after the coupling has been switched off (*Szendro et al.*, 2009). Also in supermodeling this is the case, first and foremost the coupling between the different models should result in a qualitative similarity with the natural system, the 'truth'.
An example of the qualitatively improved behaviour compared to the standard MME was seen in *Kirtman and Shukla* (2002) and *Shen et al.* (2016). Although only the fluxes were combined, there was some form of sychronization between the individual models. In *Shen et al.* (2016) it is mentioned that this synchronization is strongest in the tropical Pacific area, due to the fact that in that region the atmosphere is most sensitive to changes in SST. Then, also in the interactive ensemble of *Kirtman and Shukla* (2002) there must have been synchronization in that area. Since in *Kirtman and Shukla* (2002) the models were identical, one would expect even more synchronization. On the other hand, the models cannot have been synchronized fully, since then again there would have been effectively only one atmospheric model left, and so atmospheric noise would not be reduced. Hence this must be a case of partial synchronization, where the atmospheric models are relatively well synchronized in the tropical Pacific area, but display different behavior compared to fully synchronized models.

For some applications, one would like to have more synchronization between the individual models. If the individual models are not synchronized among themselves, such as in the case of a MME, averaging the individual model trajectories into one new trajectory cancels out the independent variability in the trajectories. The trajectory resulting from MME is smoother than the individual model trajectories, because there is a non zero probability that the individual trajectories are at different phases while their average is taken. One could compare this variance reduction to the variance of the sample mean of independent variables. If the variance of the independent variables is equal to σ^2, then the variance of the sample mean of N variables is only equal to $\frac{\sigma^2}{N}$. The averaged MME trajectory does not extend over as many possible states in the phase space as the supermodel trajectory. MME is also not meant to construct a new trajectory, the motivation is to improve statistics like the mean and higher order moments. If the supermodel on the other hand consists of synchronized models, variance reduction is minimal and a rich attractor structure can be simulated. The trajectory that emerges in this case can not only give an improvement for statistics like mean and variance, also the sequence of events can be of much value.

To which extent is synchronization possible in a supermodel? Weather and climate models exhibit chaotic behavior. Chaotic behaviour means that even for identical systems, if we start integrating from very close but different initial conditions, the trajectories will diverge exponentially fast from each other, hence the prediction error will increase over time (*Lorenz*, 1963; *Vannitsem*, 2017). Despite this non-predictable behaviour, it is still possible to synchronize identical deterministic chaotic models among themselves perfectly, such that the systems are at exactly the same state at the same time (*Pecora et al.*, 1997). Even more so, perfect synchronization can be achieved with relatively limited observations (*Duane et al.*, 2007). In supermodeling however, we work with non-identical models, the individual models do not share exactly the same set of equations. Therefore identical synchronization is not possible within a supermodel.

In case the models differ only in parameter values, *Afraimovich et al.* (1986) described the possibility of generalized synchronization. This means that for some systems a one-to-one smooth mapping exists from one model to the other. If this mapping is known, knowing one model state suffices to know the other model state. A disadvantage is that producing this mapping is often not possible (*Pecora et al.*, 1997; *Weber et al.*, 2015). Although producing this mapping might not be feasible, we can give some constraints to the amount of synchronization necessary for a supermodel forecasting on a weather timescale. By combining different models into a supermodel, the model trajectories should be synchronized enough such that the supermodel is as chaotic as the true atmo-

sphere on a weather timescale. The chaotic behaviour of the supermodel should occur during the full length of the forecast. A common phenomenon within synchronization theory is namely on-off intermittency. This means that periods of synchronization between the models are interrupted by periods of bursting away from each other, which can lead to unexpected behaviour, as we shall see in Sect. 2.4.

2.2.3 Summary of Section 2.2

Continuing on the path of the hierarchy of supermodels, not only different atmospheric fluxes, but also variables between the atmospheric models themselves can be combined. This will increase the synchronization among models, and this in turn can help to improve the dynamics of the models. The exact amount of synchronization is dependent on the goal of improvement. If there is enough synchronization between the models the resulting supermodel can give not only improved statistics, which is the aim of the MME, but also improved trajectories. Hence not only the likelihood of certain states can be seen, also the development from state to state, since the dynamics are better represented.

2.3 Different forms of a supermodel

Let us formally write the model equations of a weather or climate model i as

$$\dot{\mathbf{x}}_i = \mathbf{f}_i(\mathbf{x}_i, \mathbf{p}_i) \tag{2.1}$$

where $\mathbf{x}_i \in \mathbb{R}^n$ is a high-dimensional state vector, $\mathbf{f}_i : \mathbb{R}^n \to \mathbb{R}^n$ a non-linear evolution function depending on the state $\mathbf{x}_i$, and on a number of adjustable parameters $\mathbf{p}_i \in \mathbb{R}^m$. In practice, weather and climate models generally differ in the representation of the climate state, i.e. the phase space where $\mathbf{x}_i$ is defined, the evolution function and parameter values. Supermodeling so far has mainly been performed in a simplified context, where the models share the same evolution function, $\mathbf{f}$, and the same phase space, so that $\mathbf{x}_i \in \mathbb{R}^n$ for all i. However, the models differ in the parameters, $\mathbf{p}_i \neq \mathbf{p}_j$ if $i \neq j$. We will furthermore denote the *truth* as given by the model $\mathbf{f}$ with a specific set of parameters. The connections between the models are as yet univariate, meaning a certain element of the state vector of one model can only be coupled to the corresponding element of the other model. Next, the possible different forms of a supermodel are presented.

2.3.1 Connected supermodel

In *Szendro et al.* (2009), nudging terms were used to achieve synchronization between the observations and the model. The same approach within supermodeling is a possi-

bility, which is called a *connected* supermodel. The time-derivative for model i within a connected supermodel is given by:

$$\dot{\mathbf{x}}_i = \mathbf{f}(\mathbf{x}_i, \mathbf{p}_i) + \sum_{j \neq i} \mathbf{C}_{ij}(\mathbf{x}_j - \mathbf{x}_i). \tag{2.2}$$

Note if the synchronization between the models tends to be perfect, the nudging term $\sum_{j \neq i} \mathbf{C}_{ij}(\mathbf{x}_j - \mathbf{x}_i)$ disappears. Since the individual models do not necessarily synchronize perfectly, the supermodel solution based on N imperfect models is defined as

$$\mathbf{x}_s = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i. \tag{2.3}$$

The nudging terms $\sum_{j \neq i} \mathbf{C}_{ij}(\mathbf{x}_j - \mathbf{x}_i)$ in Eq. 2.2 push the state of each model to the state of the others at every time step. The size of the nudging coefficients $\mathbf{C}_{ij} \in \mathbb{R}^{n \times n}$ reflects the strength of the coupling between the models. They have the form of diagonal matrices and can thus be written as $\mathbf{C}_{ij} = \text{diag}(\mathbf{c}_{ij})$ with $\mathbf{c}_{ij} \in \mathbb{R}^n$. The diagonal form reflects the fact that each model state vector component is nudged towards the same component of the other model. The approach can be extended to be multivariate allowing for cross nudging, but this will require careful scaling of the variables. For appropriate connections the models fall into a synchronized motion (*Pecora and Carroll*, 1990). Note that the states will be close for strong connections so that smoothing and loss of variance due to the averaging in Eq. 2.3 will be limited. The supermodel solution depends on the relative strengths of the connection coefficients. Different sets of connection coefficients lead in general to different supermodel solutions.

2.3.2 Weighted supermodel

When the size of the connection coefficients is increasing toward infinity, the connected supermodel converges toward a weighted supermodel. A weighted supermodel based on N imperfect models is given by:

$$\dot{\mathbf{x}}_i = \mathbf{f}(\mathbf{x}_s, \mathbf{p}_i) \tag{2.4a}$$

$$\dot{\mathbf{x}}_s = \sum_{i=1}^{N} \mathbf{W}_i \dot{\mathbf{x}}_i, \tag{2.4b}$$

where $\mathbf{x}_s \in \mathbb{R}^n$ represents the supermodel state vector and diagonal matrices $\mathbf{W}_i = \text{diag}(\mathbf{w}_i)$ with $\mathbf{w}_i \in \mathbb{R}^n$ denote the weights. In the weighted supermodel the states are imposed to be perfectly synchronized at the weighting step, when the models are combined.

2.3.3 Conditions for equivalence between a connected and weighted supermodel

In *Wiegerinck et al.* (2013) a sketch of a proof is given that for $\mathbf{C}_{ij(i\neq j)} \to \infty$ a connected supermodel converges to a weighted supermodel. A reformulated and extended sketch of the proof is given in the following:

Lemma 1. *Combine the connected supermodel equations as in Eq. 2.2 into one equation for all N models, for each state variable l:*

$$\dot{\boldsymbol{x}}^l = \boldsymbol{f}(\boldsymbol{x}^l, \boldsymbol{p}) + \boldsymbol{L}^l \boldsymbol{x}^l \tag{2.5}$$

where $\boldsymbol{L}^l \in \mathbb{R}^{N\times N}$ *consists of all nudging coefficients* $c^l_{ij} \geq 0$ *between model i and j for state variable l:*

$$\begin{pmatrix} -\sum_{i\neq 1} c^l_{1i} & c^l_{12} & \cdots \\ c^l_{21} & -\sum_{i\neq 2} c^l_{2i} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{2.6}$$

Without loss of generality, omit state variable l and rewrite Eq. 2.5 to

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}) + \boldsymbol{L}\boldsymbol{x}. \tag{2.7}$$

Then the eigenvalues λ_i *of* $\boldsymbol{L}$ *are given by:* $0 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots$.

Proof. First we prove that 0 is an eigenvalue of matrix $\mathbf{L}$. Multiply $\mathbf{L}$ by $\mathbf{x} = (1, 1, \cdots)^T$. Then

$$\mathbf{L}\mathbf{x} = \begin{pmatrix} -\sum_{i\neq 1} c_{1i} + c_{12} + \cdots \\ c_{21} - \sum_{i\neq 2} c_{2i} + c_{23} + \cdots \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \end{pmatrix} \tag{2.8}$$

Hence $\lambda_1 = 0$ is an eigenvalue of $\mathbf{L}$.
Define $\mathbf{v}^i = (v^i_1, v^i_2, \cdots)$ as an eigenvector of $\mathbf{L}$, corresponding to eigenvalue λ_i, with $1 < i \leq N$. We want to show that λ_i cannot be larger than 0. Each eigenvector $\mathbf{v}^i$ has a maximal element $v^i_k := \max\{v^i_1, \cdots, v^i_N\}$. In case of multiple v^i_j that have the maximal value, choose one without loss of generality. For row k of $\mathbf{L}\mathbf{x}$ the following holds:

$$c_{k1}v^i_1 + c_{k2}v^i_2 + \cdots + c_{k(k-1)}v^i_{k-1} - \sum_{j\neq k} c_{kj}v^i_k + c_{k(k+1)}v^i_{k+1} + \cdots = \lambda_i v^i_k \tag{2.9}$$

Assume first $v^i_k \geq 0$. Since v^i_k is the largest element and all $c_{ki} \geq 0\ \forall i \neq k$ it holds that

$$c_{k1}v^i_1 + c_{k2}v^i_2 + \cdots + c_{k(k-1)}v^i_{k-1} + c_{k(k+1)}v^i_{k+1} + \cdots \leq \sum_{j\neq k} c_{kj}v^i_k \tag{2.10}$$

Hence $\lambda_i v^i_k \leq 0$, hence $\lambda_i \leq 0$.
If we assume $v^i_k \leq 0$, then

$$c_{k1}v^i_1 + c_{k2}v^i_2 + \cdots + c_{k(k-1)}v^i_{k-1} + c_{k(k+1)}v^i_{k+1} + \cdots \geq \sum_{j\neq k} c_{kj}v^i_k \tag{2.11}$$

and hence $\lambda_i v^i_k \geq 0$, hence $\lambda_i \leq 0$. □

Theorem 1. *If the connection coefficients $c_{ij} \to \infty$ for all i, j with $i \neq j$ and with the current value $c_{ij} > 0$, and matrix $\boldsymbol{L}$ is mixing and diagonalizable, then Eq. 2.2 converges to Eq. 2.4 with weights $\boldsymbol{w} = (w_1, w_2, \ldots)$ the left eigenvector of $\boldsymbol{L}$ corresponding to eigenvalue 0.*

Sketch of a proof. In the following lines we give a sketch of the proof of the theorem. First we show that if all connection coefficients $c_{ij(i\neq j)}$ of $\mathbf{L}$, that are not equal to zero, go to infinity, the connected supermodel synchronizes to one solution for all models. For simplicity we show a one dimensional case, where we assume the eigenvalues of $\mathbf{L}$ to be real. Define u as the difference between two arbitrary models i and j: $u = x_i - x_j$. Hence

$$\dot{u} = \dot{x_i} - \dot{x_j} \tag{2.12a}$$
$$= f^i(x_i) - f^j(x_j) + c_{ij}(x_j - x_i) - c_{ji}(x_i - x_j) \tag{2.12b}$$
$$= f^i(x_i) - f^j(x_j) - (c_{ij} + c_{ji})u \tag{2.12c}$$

Furthermore we assume $f : X \to \mathbb{R}$ is a bounded continuous function, which is realistic from a physical perspective. Hence there exists a real number M such that $|f(x)| \leq M \; \forall x \in X$. We can derive

$$|u| \leq \left| \frac{d}{c_{ij} + c_{ji}} + k \exp(-(c_{ij} + c_{ji})t) \right|. \tag{2.13}$$

with d and $k < \infty$. Hence, if all connection coefficients $c_{ij}, c_{ji} \to \infty$, then u converges to zero, thus the models synchronize.
In the next part, we show that in the case of N models, we can write the synchronized solution in terms of the left and right eigenvector of $\mathbf{L}$ corresponding to eigenvalue 0. For $\mathbf{L}$ diagonalizable, there exists a basis of eigenvectors. Therefore we can write $\mathbf{x}$ as a sum of right eigenvectors $\mathbf{b}_i$ and coefficients $\alpha_i(t)$:

$$\mathbf{x}(t) = \sum_i \alpha_i(t)\mathbf{b}_i. \tag{2.14}$$

Since the left and right eigenvector corresponding to different eigenvalues are orthogonal, we can obtain coefficient α_i from the inner product of $\mathbf{x}$ with the corresponding left eigenvector $\mathbf{a}_i$ of $\mathbf{L}$. We choose $\mathbf{a}_i$ such that $\mathbf{a}_i\mathbf{b}_i = 1$. Then $\mathbf{a}_i\mathbf{x}(t) = \mathbf{a}_i\alpha_i(t)\mathbf{b}_i = \alpha_i(t)\mathbf{a}_i\mathbf{b}_i = \alpha_i(t)$. Multiplying both sides of Eq. 2.7 by $\mathbf{a}_i$ results in:

$$\dot{\alpha_i}(t) = \mathbf{a}_i\mathbf{f}(\mathbf{x}, \mathbf{p}) + \mathbf{a}_i\mathbf{L}\mathbf{x} \tag{2.15a}$$
$$= \mathbf{a}_i\mathbf{f}(\mathbf{x}, \mathbf{p}) + \lambda_i\mathbf{a}_i\mathbf{x} \tag{2.15b}$$
$$= \mathbf{a}_i\mathbf{f}(\mathbf{x}, \mathbf{p}) + \lambda_i\alpha_i. \tag{2.15c}$$

We rewrite $\mathbf{L}$ as $\mathbf{L} = k\mathbf{M}$ with $k \to \infty$ instead of $c_{ij} \to \infty$. Then

$$\dot{\alpha}_i(t) = \mathbf{a}_i \mathbf{f}(\mathbf{x}, \mathbf{p}) + k\tilde{\lambda}_i \alpha_i. \tag{2.16}$$

Equation 2.16 has the same form as Eq. 2.12, such that $\alpha_i(t)$ converges to zero if $k\tilde{\lambda}_i \to -\infty$. Since $\mathbf{L}$ is mixing (all models are directly or indirectly connected with each other), $\mathbf{L}$ has a dominant eigenvalue: $\lambda_1 = 0$ (Theorem of Perron-Frobenius). Since according to Lemma 1, all eigenvalues are smaller or equal to zero, $\lambda_i < 0 \; \forall i \in \{2, N\}$. All modes corresponding to negative eigenvalues will vanish, and therefore the solution for $\mathbf{x}$ will converge to:

$$\mathbf{x}(t) = \alpha_1(t)\mathbf{b}_1 \tag{2.17}$$

Define the left eigenvector of $\mathbf{L}$ corresponding to $\lambda_1 = 0$ as $\mathbf{a}_1 = (w_1, w_2, w_3, ...)$ with $\mathbf{a}_1$ normalized such that $\sum_i w_i = 1$. Then $\mathbf{a}_1\mathbf{b}_1 = 1$ and $\alpha_1 = \sum_i w_i x_i = x_s$, with s denoting the synchronized supermodel. This results in the synchronized states $\mathbf{x}(t) = \alpha_1(t)\mathbf{b}_1 = (x_s(t), x_s(t), \ldots)$

To obtain the synchronized state dynamics, multiply Eq. 2.7 with the left eigenvector $\mathbf{a}_1$. The term with $\mathbf{L}$ vanishes, and we are left with: $\dot{x}_s = \sum_{i=1}^{N} w_i \dot{x}_i = \sum_{i=1}^{N} w_i f(x_s, p_i)$, corresponding to Eq. 2.4 for the weighted supermodel.

2.3.4 Connected or weighted?

Should either a connected or weighted supermodel be preferred? It is difficult to draw a firm conclusion, because the answer depends on the specific type of applications. Nevertheless some advantages and disadvantages of each approach can be named. According to *Wiegerinck et al.* (2013), a connected supermodel allows for more flexibility in case the ensemble trajectories are not perfectly synchronized. For example, if the models are stuck in a certain regime if one calculates a fixed weighted average supermodel state, but the individual models in a loosely connected supermodel are able to escape. On the other hand, the optimal size of the connection coefficients after training the supermodel on the basis of observations is typically quite large in examples of small scale systems such as Lorenz63 (*Wiegerinck et al.*, 2013). The larger the coefficients, the stronger the models converge on a synchronized trajectory, which can be described by a weighted superposition of the models as we have seen in the previous subsection. The possible flexibility advantage of the connected supermodel has not been observed yet. For some training applications, near perfect synchronization, as imposed in a weighted supermodel, is also required, as we shall see later in the results. Ideally, one would require in principle as much synchronization as possible, if the individual models are good enough to obtain one supermodel that closely describes the

true system.
That said, creating this near-perfect supermodel is still utopian. A weighted average for every variable might not be feasible because of disrupting physical balances. A potential solution is to construct a hybrid model, where every time step a weighted average is calculated for the variables for which it is possible. For the 'difficult' variables for which it is not possible, there could be a softer approach with nudging instead of a weighted average. This approach will not disrupt the balances but still keep the variables from the different models in the neighborhood of each other, and this should be an improvement compared to not connecting the variables at all.

2.3.5 Summary of Section 2.3

In our simplified context of models sharing the same evolution function and phase space, we can define a connected or a weighted supermodel. A connected supermodel uses nudging to bring the models closer, while in a weighted supermodel all models receive as new state the same weighted average of the individual model states, thus synchronization is imposed. If the nudging in a connected supermodel is strong enough, it will converge to a weighted supermodel. Depending on the type of synchronization needed or possible, either a connected or weighted supermodel is favorable.

2.4 Synchronization: what, where and when

Using a connected or weighted supermodel influences the degree of synchronization among models, as described in the previous section. Other aspects that influence the amount of synchronization are the variables being exchanged and how often this takes place. In a real-world climate context, this is partly dependent on the amount of observational data available. Observational data are used to obtain an optimal combination of the models to best simulate the observations. Another factor that could limit the exchange between the models is the capacity of high performance computers.

2.4.1 Which variables should be exchanged?

As mentioned in the previous section, not all variables are suitable to be combined. For example humidity is a variable that is known to be troublesome. On one hand, it behaves like a tracer, being advected around by the winds. On the other hand, it can have a strong impact on the vertical stability, triggering convective events with strong mixing of heat and moisture in the vertical and the release of latent heat in the atmosphere. In the SPEEDO model (*Severijns and Hazeleger*, 2010) for example, a step function determines whether convection will take place. Assume we have two models, where in

only one of them convection takes place at a particular grid point and time step. After the new model states have been calculated, the variables are exchanged and a new humidity profile is defined for both models. The humidity state imposed by the exchange will impact the vertical stability and may trigger strong convective events. We found that these imbalances either make the models crash, or make the training and forecast difficult because of the numerical shocks they introduce: inconsistencies that could induce non-characteristic convective outbursts. Issues with humidity are known in data assimilation too (*Liu and Kalnay*, 2005). Humidity has large variability in small scales and large temporal variations. These two aspects make it also more difficult for training a supermodel.
Although humidity seems to be a difficult variable to exchange between the models, one needs enough variables to be exchanged in order to maintain enough synchronization between the models. Luckily, variables such as temperature and wind are more continuous in time and space, thus making them easier to use in synchronization of the models.

2.4.2 Where in space should the models be combined?

To reduce computation effort, it could be useful to couple only a part of the state vector, in case coupling is necessary but full coupling too expensive. Coupling only a part of the state vector can also be very useful in case not enough observations are available for training in a certain area.

Gridpoint representation

Suppose our model is numerically solved using a gridpoint representation. Usually, the observations we use to train the supermodel, are not available at each grid point, and very commonly they are not even located on the same grid. If the supermodel is constructed with global connections (i.e., the connection coefficient has the same value for every grid point), it is straightforward to also apply these connections to the areas where no observations are available, if this coupling is required. However, it is not straightforward to train coefficients that vary regionally if data are not available at all grid locations. In this case, one could choose to not couple the regions where no observations are available, as long as these regions are sufficiently synchronized (through constraining the model elsewhere). It is difficult to obtain synchronized models when only certain geographical parts are nudged, such as the tropics, the polar regions, only land or only sea, since then certain physical processes are not connected. In these cases, one could use some form of interpolation or a more advanced form of multivariate data assimilation to determine connections in the areas without observations.

Spectral representation

A possible reduction in the number of connections is to exchange less spectral coefficients, as discussed in *Hiemstra et al.* (2012). It is shown that in a QG-model (*Marshall and Molteni*, 1993), in the case of identical models, only the first eleven out of a total of thirty wavenumbers, describing the largest scales, need to be nudged towards each other to obtain complete synchronization. It is also possible to only combine intermediate and small scales. However, the fewer large scales are nudged, the more wavenumbers are needed for full synchronization.
Another way of limiting information exchange is by only nudging in certain preferred phase space directions. In *Hiemstra et al.* (2012), EOF vectors are used to define these directions. The first EOF vector is in the direction where most of the variance is explained. Nudging along the dominant EOF vectors could be efficient since they efficiently span the space embedding of the attractor. Nudging in a direction orthogonal to the attractor is not very productive, since one of the characteristics of an attractor is that model trajectories converge quickly on the attractor, without any form of nudging. Hence this direction can be omitted. *Hiemstra et al.* (2012) found that using EOF vectors as basis vectors is more efficient than connecting original state variables in the context of Lorenz63 and the QG-model. In *Duane et al.* (2006), the possibility of bred vectors as basis vectors for the nudging is discussed. Bred vectors are useful to determine likely directions of forecast error (*Toth and Kalnay*, 1997). *Duane et al.* (2006) notice that seemingly any basis that captures the essential physics in each local region is necessary and sufficient for synchronization.

2.4.3 How often should models exchange?

In Eq. 2.2 and Eq. 2.4 the supermodel is defined as a combination of the time-derivatives of the imperfect models. Therefore, the models are combined every time step. However, the observations used to train the model may not be available every time step. Should then the imperfect models in the supermodel be combined as frequently as that the observations are available during training? And how should this combination be constructed?
In the case of a connected supermodel, one could use so-called snapshot nudging. This means that the nudging term in Eq. 2.2 is active only when an observation is available. This is achieved by appending a δ-function equal to zero, except when an observation is encountered. In the case of a weighted supermodel one could simply combine the states of the imperfect models instead of the tendencies. This can be done as often as desired. Since we started from the simplified assumption that the models share

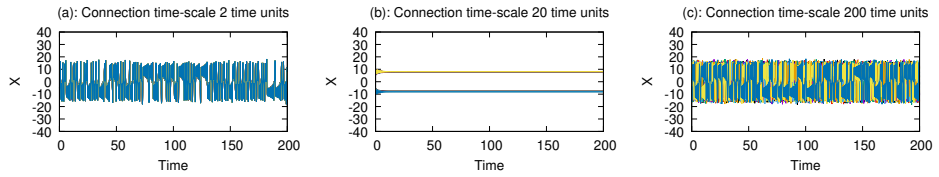


Figure 2.5: The x-component of 22 identical Lorenz63 systems nudged towards each other with different nudging strengths, for (a): frequent (b): intermediate frequency and (c): infrequent coupling.

the same evolution function and phase space, the time steps of the different models will coincide. Note that combining states in a weighted supermodel is different from combining states in MME, as in MME the states are only combined after the runs have been performed, while in a weighted supermodel the combined state is calculated with a certain frequency during the run. This combined state is then used as an initial condition for each model to continue its trajectory.

If the observations are too sparse, it could happen that combining the models with this same frequency leads to *oscillation death*. Oscillation death means that the supermodel has a significant lower variance than the imperfect models individually. Figure 2.5 shows oscillation death in the case of a connected supermodel, with identical standard Lorenz63 systems starting from different initial conditions. If the connection is strong enough, all models will fully synchronize (Fig. 2.5a). If the connection is too weak (Fig. 2.5c), the models do not synchronize at all. With a connection strength in between, it can happen that the models suddenly collapse to one or two seemingly fixed points (Fig. 2.5b), sometimes interrupted by periods bursting apart. Similar 'oscillation death' can also be seen in a weighted Lorenz63 supermodel. Figure 2.6 shows an example with three different imperfect models, that were only combined every 25 time steps of 0.01. Apparently in Lorenz63 it is possible that the model tendencies exactly counteract each other when either the connection is not strong enough or not frequent enough. Whether more complex models are also easily affected by oscillation death is still an open question. This unexpected behavior due to combining the models not frequent enough is something to avert. On the other hand, if the observations for training have a certain frequency in time, and the imperfect models exchange information more often during the supermodel forecast run, the connection coefficients or the weights of the supermodel might not be optimal.

2.4.4 Summary of Section 2.4

This section discussed which variables have to be exchanged between models in order to synchronize their simulations, and where in space and how often the variables need

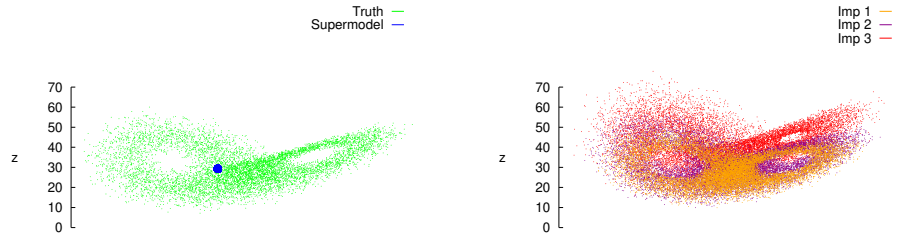


Figure 2.6: The true and supermodel attractor (left) and the three imperfect model attractors (right).

to be exchanged. Variables such as temperature and wind do not seem to pose many problems since they are relatively continuous. A variable like humidity on the other hand seems more difficult to exchange. From a computational point of view or because of spatial lack of observations, it can be impossible to combine every gridpoint or every wavelength for a certain variable. In this case, synchronization can be still possible if the necessary physical processes are captured within the elements that are coupled. Coupling the models as frequent as possible in time however seems valuable, to avoid issues such as oscillation death.

2.5 Training of supermodels

The basis for a supermodel is the approach to synchronize the models (in terms of exchanged variables, and their spatial and temporal exchange frequency). In this way a supermodel can be seen as a time dependent graph, where the connected vertices exchange directly information with each other during a supermodel time step. Then the length of the edges defines the connection strength. For a connected supermodel the values of the edges are just the values of the connection coefficients, for a weighted supermodel one could convert to connection coefficients, keeping in mind that the correspondence is only exact in the limit of infinitely large connection coefficients. The value of the connection coefficients or the weights does not have to be positive by definition. If the individual models cannot compensate for each other, negative weights can help to further improve the predictions. An example are the heat fluxes in *Shen et al.* (2016), where one of the models contributed with its heat flux multiplied by a negative weight to the supermodel. In this thesis we will further explore the possibilities for negative weights. The aim of this thesis is to train supermodels: to define the supermodel structure and the values of the connection strengths.

2.5.1 Cost function

Previous training approaches often involved minimization of an appropriate cost function, like many optimization problems. However, the cost function value depends on the model solution, and often multiple evaluations are necessary which makes it a costly approach.
In *Shen et al.* (2016) a simplex-method is used, the Nelder-Mead algorithm (*Nelder and Mead*, 1965). With this simplex method a cost function is calculated for different weights. In *Shen et al.* (2016) more than 300 computations of the cost function and hence more than 300 model runs were necessary to obtain a sufficient minimum. Since the interactive ensemble was trained on minimizing the error in SST climatology, one run corresponded to a simulation of 40 years, which made the training computationally very expensive.
A less expensive cost function approach is used by *van den Berge et al.* (2011). A connected supermodel is constructed that consists of imperfect model versions of small scale systems such as Lorenz63. The cost function is minimized with the Fletcher-Reeves-Polak-Ribiere-Conjugate Gradient method (*Fletcher and Reeves*, 1964). Comparing this method to the Nelder-Mead algorithm is difficult, since the latter one does not rely on first-order derivatives as Conjugate Gradient methods do. Also the imperfect models of *van den Berge et al.* (2011) and *Shen et al.* (2016) are very different. Both minimization methods however need in the order of 100 model evaluations. The main computational difference lies within the fact that in *van den Berge et al.* (2011) the cost function is only calculated for short term time intervals. This is even enhanced by adding a γ^t factor, where t is time, to the cost function, with $\gamma < 1$ such that differences between the observations and the supermodel close to the time of the initialization have a stronger impact on the connection coefficients than later on during the run. In *Mirchev et al.* (2012) the same cost function approach is used, but instead of the γ factor a factor $w(t)$ is used, which is during the run equal to $\alpha^t, \alpha < 1$, except for the very last time instant of each run, where $w(t) = 1$. This is to ensure that the supermodel trajectory does not deviate too quickly from the observations.

2.5.2 Duration of the training time

Once the training method has been chosen, also the training duration needs to be decided. *Wiegerinck and Selten* (2017) showed that in a driven Lorenz63 system (*Wiegerinck and Basnarkov*, 2013) and a Quasi-Geostrophic (QG) atmospheric model (*Marshall and Molteni*, 1993), a weighted supermodel trained on only short timescales can result in a worse climatology than the individual models or a posterior ensemble.

For example, the forecasts used to train the QG model were merely one day, while most of the weather phenomena are captured within three or four days. One would therefore expect that training using longer forecasts of several days is necessary to obtain an improved supermodel weather forecast. *Wiegerinck and Selten* (2017) illustrated nicely that training of the supermodel is dependent on the purpose of the supermodel, and that short-term training does not always imply improved long-term behaviour.

2.5.3 Summary of Section 2.5

In order to train a supermodel, one needs first to define the structure of a supermodel: connected or weighted, which models are synchronized, which variables are exchanged, with which spatial and temporal exchange frequency? The next step is to train the values of the connection strengths. The previously used training approach of minimizing a cost function is computationally very expensive, since it often requires a lot of model runs. Furthermore, if one would like to have a supermodel that performs well in the long-term, it is not always sufficient to train only on short-term, and this makes training even more computationally demanding. Hence, there is a need for new training methods that are computationally much more efficient than the cost function approaches so far. The development of such methods is the objective of this thesis.

Chapter 3

Objectives

The main objective of this thesis is to develop efficient algorithms to train supermodels: a combination of models that outperforms the constituent individual models and the standard multi-model ensemble approach in the context of weather and climate predictions. Previously, supermodels were trained by algorithms that tried to minimize a cost function, and this required many model simulations. The computational cost of running state-of-the-art climate models is huge, hence there is a need for more efficient training algorithms.

The first aim is therefore to develop new efficient training schemes. First the methods are developed and tested on low-order dynamical systems such as Lorenz63 and a quasi-geostrophic model (Paper I). We start with a 'perfect model approach', such that the 'observations' to train the supermodels are noise free and continuously available in time. In reality, observations are not perfect and only available sparsely in time. We therefore investigate the effectiveness of the methods in the context of noisy and sparse in time observations and adapt the methods when required (Paper IV).

Having developed the training schemes in the context of the simpler systems, we proceed with larger scale systems and construct the first supermodel out of a global coupled atmosphere-ocean-land model of intermediate complexity, SPEEDO (Paper II). As this is the first time a supermodel is constructed by combining different atmospheric components solving the moist primitive equations, therefore the second aim is to determine which physical variables should be connected in a supermodel and which variables should not.

Intuitively, the supermodel approach can only work if the individual models can compensate for each others errors, for example when one model overestimates the temperature and the other model underestimates. Therefore it is important to make precise which characteristics the individual models should have in order to be successfully combined into a supermodel (Paper I and III). In case the different individual models cannot compensate each others errors as required, we address the question whether

negative connections or weights can help to improve the supermodel solution (Paper III).
The supermodeling approach is not the only approach to combine different models dynamically. Another possibility is to 'cross pollinate' model trajectories (*Smith*, 2001; *Du and Smith*, 2017). In this approach, model states are not combined into a supermodel state, but instead different models are selected to be used for different geographic regions. Data assimilation is then used to combine the best model states from the different regions into new initial conditions for the models to continue their simulation with. We compare both approaches in the context of a small scale Lorenz96 system (Paper V).

To summarize, the main objectives are:

- To develop new efficient training schemes for supermodels that are not based on minimization of a cost function.
- To use the training schemes to train a supermodel based on different versions of the global coupled atmosphere-ocean-land model SPEEDO and compare the methods.
- To show that the supermodels can outperform the individual imperfect models as well as the standard multi-model ensemble approach.
- To identify more precisely the characteristics that the individual models should have in order to be combined together into a supermodel.
- To investigate which physical variables should be connected in a supermodel, and which ones not.
- To explore the options of negative connections between the models within a supermodel.
- To investigate the effectiveness of the training methods in the context of noisy and sparse in time available observations, adapt the methods if necessary and make them suitable for training state-of-the-art models.
- To compare the supermodel approach to the cross pollination approach of *Du and Smith* (2017).

Chapter 4

Summary of results

In this thesis efficient training methods are developed to optimally combine different models dynamically into a so-called supermodel with better weather and climate forecast skill than the individual models. The training is the stage of the process where the relative weights of each model within the supermodel are determined based on observational data. In Paper I a new training scheme is developed based on Cross Pollination in Time (CPT), in which model trajectories are 'crossed', such that a larger ensemble of trajectories can be constructed. The method is first tested on small scale dynamical systems. In Paper II the 'synch rule', a new learning rule originally intended for parameter estimation, is introduced and tested on the global atmosphere-ocean-land model SPEEDO of intermediate complexity. In Paper III, CPT is successfully applied to SPEEDO and using the synch rule, the novel concept of negative connections between models is introduced. The results in Papers I to III were obtained using noise-free observations of the full systems. On the other hand, in Paper IV both methods are adapted to train supermodels on the basis of noisy and sparse observations. In Paper V the supermodeling approach is compared to another method called CPT II, also based on the original CPT, to let models exchange information during the simulation. The advantages and disadvantages of both approaches to combine models dynamically are discussed.

Paper I: An efficient training scheme for supermodels
Francine J. Schevenhoven and Frank M. Selten
Earth System Dynamics, **8**, 429-438 (2017)

In this study a new training method, based on the idea of cross pollinating different model trajectories, is introduced. It is assumed that we have an observed trajectory, called the 'truth'. The training phase of CPT starts from an observed initial condition

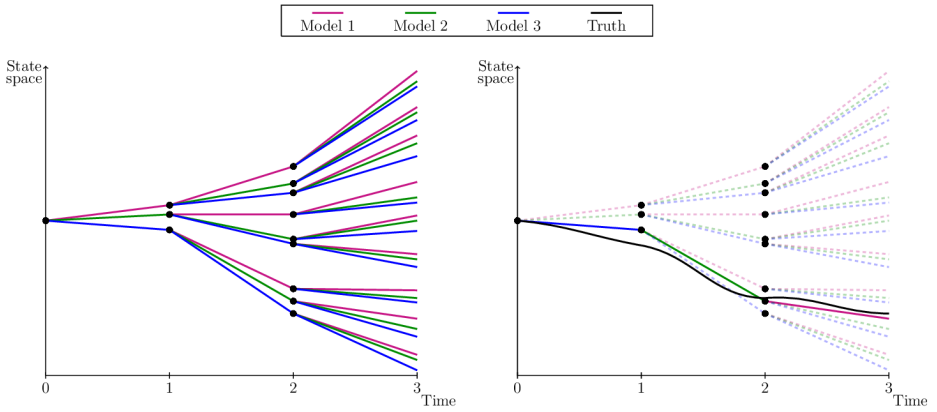


Figure 4.1: CPT in general (left) and CPT in the context of supermodeling (right).

in state space. From this initial state, the imperfect models run for a certain period each ending in a different state. From these endpoints all models run again (Fig. 4.1 (left)). For training a supermodel only those predictions that remain closest to the truth are continued, the others are discarded. This results in a CPT trajectory (Fig. 4.1 right). In a sufficiently long training phase, we can determine the frequency that a model is selected to keep the CPT close to the truth. These frequencies determine the weights in the supermodel. This CPT method leads to a dynamical combination of forecast models (a weighted supermodel) with superior short-term forecast quality and improved climatology. The method is applied to the Lorenz63 system and a quasi-geostrophic model. Observations are obtained from a 'perfect' model. The imperfect models are versions of the same model with perturbed parameters. The CPT training is very efficient, since only one or possibly a few training iterations are necessary to obtain converged weights. Furthermore, CPT is based on short-term trajectories, although it turned out that the errors in the climatology are also greatly reduced. The results show that a supermodel with weights trained by CPT can give significantly better predictions than a supermodel consisting of the same imperfect models with equal weights. The imperfect model parameters should form a so-called 'convex hull' in order for the supermodel to outperform the individual models: the imperfect model parameters surround the perfect parameter values. The convex hull implies that an additional imperfect model with worse prediction skill than the other models might still contribute to a superior supermodel solution, as long as it can compensate the errors in the other models.

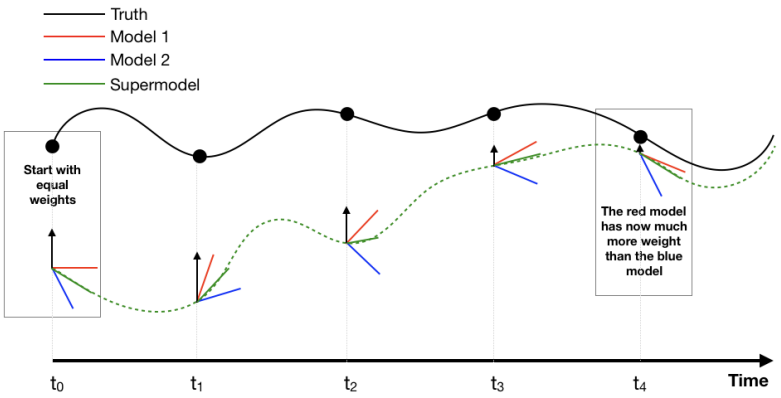


Figure 4.2: Synch rule training. The supermodel weights are updated such that the supermodel synchronizes with the truth.

Paper II: Simulating climate with a synchronization-based supermodel
Frank M. Selten, Francine J. Schevenhoven and Gregory S. Duane
Chaos, **27**,126903 (2017)

This paper explores the options of using a synchronization-based learning algorithm (the synch rule) which was originally intended for parameter estimation, to train a supermodel. During training, the weights are updated according to the rule such that the supermodel synchronizes with the observations (see Fig. 4.2). The synch rule is based on the covariance between the observation-supermodel synchronization error and the inter-model state or tendency difference. In the experiments, the perfect model approach is used, in which the perfect and imperfect models are different parametric versions of the SPEEDO global climate model, an atmosphere model coupled to a land and an ocean/sea-ice model with about 250,000 degrees of freedom. Connection terms are introduced that synchronize the imperfect models on a common solution, referred to as a connected supermodel. The resulting supermodels have a climatology and a climate response to a CO_2 increase in the atmosphere that is closer to the perfect model as compared to the imperfect models and the standard multi-model ensemble average.

Paper III: Improving weather and climate predictions by training of supermodels
Francine Schevenhoven, Frank Selten, Alberto Carrassi and Noel Keenlyside
Earth System Dynamics, **10**, 789-807 (2019)

We apply both CPT and the synch rule method to different parametric versions of SPEEDO to construct a weighted supermodel. The SPEEDO model has five prognostic variables: temperature, vorticity, divergence, specific humidity and surface pressure. Exchanging specific humidity and surface pressure deteriorate the supermodel solution, probably because the exchange leads to imbalances and fast spurious adjustments. On the other hand, a perfect SPEEDY atmosphere only fully synchronizes with the observations when at least temperature, vorticity and divergence are nudged to the observations. Therefore, in a weighted supermodel, at least these variables need to be exchanged. The weights are trained using data from the perfect model with both CPT and the synch rule training methods. Both training methods yield supermodels that outperform the individual imperfect models in short-term forecasts as well as in long-term climate simulations. Furthermore, the advantage of the supermodeling approach compared to the MME is shown. In one experiment, all imperfect models have a warmer and wetter climatology than the truth, such that a MME with positive weights will also result in a warmer and wetter climatology. The trained weighted supermodel however has a climatology close to the truth. This is due to the fact that the model errors are compensated at an early stage and not a posteriori as in a MME, where model errors have propagated spatially across the globe and across the different meteorological fields. In another experiment we explored the use of negative weights in order to improve predictions in the case that model errors do not compensate; i.e., both imperfect models have parameter perturbations and climatological errors of the same sign. A supermodel with negative weights trained using the synch rule results in stable and credible simulations in which the short-term forecast errors as well as climatological errors are reduced with respect to the imperfect models.

Paper IV: Training supermodels with noisy and sparse observations
Francine Schevenhoven and Alberto Carrassi
Manuscript in preparation

In this paper we show the potential of the CPT and synch rule training methods to train a weighted supermodel on the basis of noisy and sparse in time observations. In Paper III, both methods were able to improve weather and climate predictions in a perfect observations setting by using different parametric versions of the global coupled atmosphere–ocean–land model SPEEDO for both the imperfect models. In this work, the

observations have been perturbed with different levels of Gaussian noise. Furthermore, the observations are not available every model time step of 15 min, but only every hour, 6 hours or 24 hours. Still, with some adaptations to the training methods, both methods are able to give adequate weights. Another result of this paper is that not only the synch rule, but also CPT is adapted such that it can handle negative weights as well.

Paper V: A comparison of different approaches to combine models dynamically
Francine Schevenhoven and Hailiang Du
Manuscript in preparation

In this manuscript we compare different approaches to cross-pollinate model trajectories. The main difference between our approach and CPT II of *Du and Smith* (2017) is in the construction of the communication between the models. In the supermodeling approach, forecasts are improved by weighting different model states into a new supermodel state, while in the approach of *Du and Smith* (2017) for each region in state space only one model is chosen to use, and data assimilation is performed to assimilate the new states into each of the individual models. Both methods are applied to different parametric versions of the Lorenz96 system with noisy observations and result in forecasts of similar quality that clearly outperform the individual model forecasts. Another experiment in this study uses a two-layer Lorenz96 system as perfect model and one-layer Lorenz96 systems as imperfect models. In the presence of this unresolved scale error the CPT supermodel training is still able to provide weights of good quality. This is promising for future applications with state-of-the-art models and real observations.

Chapter 5

Discussion and Future work

In this thesis I have shown that by combining models dynamically into a so-called supermodel, weather and climate predictions can be improved. The focus in the thesis is on the development of efficient algorithms to train such a supermodel on the basis of observations. I used two different forms for the supermodel, either (1): a connected supermodel in which the individual models are nudged towards each other, or (2): a weighted supermodel, in which a new supermodel state is calculated, which is a weighted superposition of the individual model states. I introduced a new training scheme based on Cross Pollination in Time (CPT) and adapted the already existing synchronization rule (synch rule) for parameter estimation for training. The methods were first applied to small dynamical systems such as Lorenz63 and second were applied to different versions of a climate model of intermediate complexity, the global atmosphere-ocean-land model SPEEDO, and noisy and sparse in time observations. The resulting trained supermodels outperformed the individual models, and an optimally weighted Multi-Model Ensemble (MME) as well. This shows the advantage of combining models dynamically instead of only statistically such as with MME. I identified which physical variables should be exchanged among the models, such as temperature and wind, and which variables are not suitable to combine dynamically, such as humidity and surface pressure. Furthermore, I developed the convex hull principle to determine which imperfect models should be combined. I discovered that a 'bad' model with poor forecast quality is able to improve the supermodel provided it can compensate the erroneous behavior of the other models. If the convex hull cannot be constructed with positive weights, I showed that negative weights can result in stable and plausible simulations with mitigated errors.

Both CPT and the synch rule are very efficient training methods. The synch rule requires one year of simulation to obtain stable weights or connection coefficients. The training period for CPT is even as short as a few weeks. Despite training on weather time scale, the supermodels also showed improved behavior on climatological time

scale, which is in line with the research of for instance *Rodwell and Palmer* (2007), that climatological errors due to uncertainty in the model parameters develop quickly during the first few days of weather forecasts. Nevertheless, it can be necessary to extend the training period if not all model errors originate from these 'fast' physics, for example when the imperfections in the slower ocean and land components of the climate model are of importance. For longer training periods, both CPT and the synch rule trajectories need to be nudged towards the observations, in order to remain in the same phase as the observations. To further enable the trajectories to stay close to the observations I presented other ideas for future work such as increasing the ensemble size of possible trajectories or defining the distance between the observations and the trajectories not only in terms of RMSE, but in terms of systematic errors, such as an erroneous double Intertropical Convergence Zone (ITCZ) for instance.

In principle CPT and the synch rule are both suitable for training a supermodel consisting of several state-of-the-art models. The application of these methods is however not straightforward. State-of-the-art models require much more computational time than the SPEEDO model used in this thesis. The difference in computational cost of a supermodel compared to a MME is the exchange between the models during the simulation. This exchange can be very expensive, for example, if new restart files need to be written and read in order to let the models continue from a new supermodel state. Combining the models every model time step is therefore not realistic. On the other hand, models should be connected frequently enough, such that they remain in the same phase, otherwise the supermodel trajectory will suffer from variance reduction. Hence a certain balance needs to be found in the frequency of exchange among the models.

Another possible complicating factor is that state-of-the-art models can differ in resolution and time step, hence making it more difficult to simply replace one model state by another, as performed in this thesis. A possible solution is to create a common state space, and to use techniques from data assimilation to transform the individual model states towards the common space. This transformation will inevitably increase errors, hence in future perhaps a more direct form of nudging one model state to the other would be more efficient. Furthermore, in a supermodel an artificial model state is created, that may disrupt physical balances and is therefore not suitable for the individual models unless strongly adapted by data assimilation. Under these circumstances another possibility to explore is to use CPT in the sense of *Du and Smith* (2017), such that each geographic region is represented by only one model. Then again the models can be connected every cross pollination time by data assimilation techniques, such as pseudo-orbit data assimilation as used in *Du and Smith* (2017).

Apart from a different resolution and time step, state-of-the-art models can also differ

in numerical discretization and physics, which could complicate the training process compared to the parametric error only in SPEEDO, since it might be more difficult to follow the observations during training. A small experiment was performed in the context of Lorenz96, to show that CPT also gave good results when the observations came from a higher resolution model than the imperfect models in the supermodel. Additionally, the experiments with sparse in time and noisy observations showed that even under these circumstances both the CPT and the synch rule trajectories were able to follow the observations if adaptations were made such as nudging towards the observations. These results give us the confidence that both CPT and the synch rule are in principle suitable for state-of-the-art models. The preliminary work has been performed, the training methods are ready to be applied to train a supermodel that can truly improve the current state-of-the-art weather and climate predictions.

I

Paper I

An efficient training scheme for supermodels

Francine J. Schevenhoven and Frank M. Selten

Earth System Dynamics, **8**, 429-438 (2017)

An efficient training scheme for supermodels

Francine J. Schevenhoven[1,2], and Frank M. Selten[1]

[1]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[2]Geophysical Institute, University of Bergen, Bergen, Norway

Abstract

Weather and climate models have improved steadily over time as witnessed by objective skill scores, although significant model errors remain. Given these imperfect models, predictions might be improved by combining them dynamically into a so-called "supermodel". In this paper a new training scheme to construct such a supermodel is explored using a technique called cross pollination in time (CPT). In the CPT approach the models exchange states during the prediction. The number of possible predictions grows quickly with time and a strategy to retain only a small number of predictions, called pruning, needs to be developed. The method is explored using low-order dynamical systems and applied to a global atmospheric model. The results indicate that the CPT training is efficient and leads to a supermodel with improved forecast quality as compared to the individual models. Due to its computational efficiency, the technique is suited for application to state-of-the art high-dimensional weather and climate models.

1 Introduction

Weather and climate models remain imperfect despite continuous model development. For example at middle to high latitudes, the simulated zonal wind stress maximum averaged over an ensemble of state-of-the-art climate models lies 5 to 10° equatorward of observationally based estimates, which means that on average the midlatitude winds are too strong in the current models (*IPCC*, 2013).

Improving the models is a large research effort. A demanding aspect is that there are many uncertain parameters and approximations because not all physical processes are explicitly resolved. To model and tune all of these aspects requires a huge computational effort. Even if the optimal solution can be achieved, imperfections remain due to the complexity of the climate system with interactions across scales over many orders of magnitude. In order to improve predictions, it often helps to average across model outcomes as model errors tend to average out. *Branicki and Majda* (2015) provide some evidence that this multi-model ensemble method (MME) does indeed improve predictions under certain conditions. However, it is not straightforward which imperfect models and what weights should be used for the MME forecast. Because of this,

almost all operational MME predictions are based on equal weights.

In contrast to the standard MME, an alternative approach is to let models exchange information during the simulation which leads to new solutions. If the models complement each other, these solutions potentially stay closer to the observed trajectory than the trajectories of the imperfect models individually. Hence, both the short-term predictability and the climate statistics will improve. The MME approach only combines trajectories from an ensemble of models after the simulation. This can lead to improved estimates of, for instance, the true mean state. It cannot, however, produce trajectories that remain closer to observed trajectories as combining trajectories of different models leads to smoothing.

A successful approach of combining models is found in *van den Berge et al.* (2011), where combining models into one large supermodel (SUMO) by prescribing connections between model equations is proposed. The connection coefficients are learned from historical observations. The optimization of the coefficients is achieved by minimizing a cost function. In *Wiegerinck et al.* (2013), it is noted that the size of these coefficients is typically very large. If the connection coefficients are large enough, the system will quickly synchronize into a joint state. This joint state can be described as a weighted superposition of the imperfect models referred to as weighted SUMO.

Since the minimization of a cost function can be computationally very expensive, we propose a new procedure in this paper to construct such a weighted superposition of imperfect models. The weights are learned from observed trajectories. This new learning process is based on an idea proposed by *Smith* (2001): cross pollination in time (CPT). CPT "crosses" different model trajectories in order to create a larger solution space with trajectories that potentially follow the observed evolution more closely.

Our training method for a weighted supermodel is developed using the Lorenz 63 system (*Lorenz*, 1963) following the perfect-model approach. The model with standard parameter values generates observations and imperfect models are created by perturbing the parameter values. Next, we apply the method to a more chaotic and realistic global atmospheric model with 1449 degrees of freedom by *Marshall and Molteni* (1993).

Section 2 of this paper explains the training by cross pollination. Applications of the method are described in Sect. 3 for the Lorenz 63 system and in Sect. 4 for the global atmospheric model. The final section discusses the results and provides an outlook to apply the developed approach to state-of-the-art models.

2 Training the supermodel

We assume that we have an observed trajectory, called the "truth". The training phase of CPT starts from an observed initial condition in state space. From this initial state, the imperfect models run for one time step each ending in a different state. (See Fig. 1a.) From these endpoints all models run again. Continuing this process leads to a rapid increase in the number of predictions with time. A larger region of the state space thus can be explored. In order to retain only a small number of predictions, a pruning step is required. We choose to continue only those predictions that remain closest to the truth, the others are discarded, as is depicted in Fig. 1b.

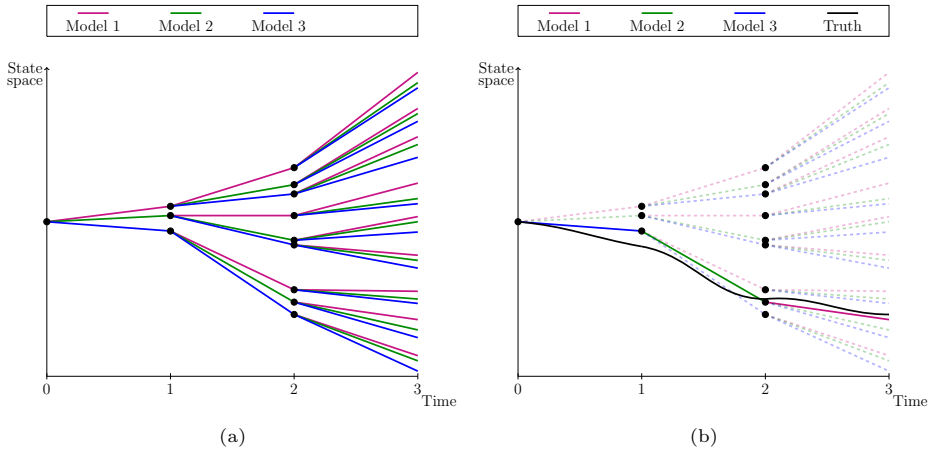


*Figure 1: Cross pollination in time for 3 models, without pruning (**a**) and with pruning (**b**).*

2.1 Determining weights

In the training phase, for each model, it is counted how often for a particular variable its prediction remains closest to the truth. The probabilities thus obtained can be used as weights for the corresponding time derivatives of the variables. This superposition of weighted imperfect models forms a supermodel which potentially has improved prediction skill.

2.2 Iterative method

In order to obtain convergence towards a supermodel that reflects the truth in the best possible way, the training is carried out iteratively. The first iteration step leads to a first estimate of the weights of the supermodel. In the second iteration step, this supermodel is added as an extra imperfect model. In the subsequent iteration steps, the

previously obtained supermodel is replaced by the newly obtained supermodel. If the added supermodel is closer to the truth than the initial imperfect models, the constructed trajectory in the CPT procedure receives fewer contributions from the initial imperfect models. Ideally, learning stops when the supermodel remains closer to the truth than the individual imperfect models for all time steps during the training.

3 Results Lorenz 63

In the Lorenz 63 system a chaotic attractor appears for certain parameter values. The attractor has the shape of a butterfly and each "butterfly wing" contains an unstable fixed point at its center, around which the trajectories alternately revolve in an unpredictable pattern. The differential equations of the system contain system parameters σ, ρ, β. The state space is described by coordinates x, y, z (Eq. 1).

$$\dot{x} = \sigma(y - x) \tag{1a}$$

$$\dot{y} = x(\rho - z) - y \tag{1b}$$

$$\dot{z} = xy - \beta z \tag{1c}$$

The standard parameter values are $\sigma = 10, \rho = 28$ and $\beta = 8/3$. Numerical solutions are obtained by using a fourth-order Runge-Kutta time stepping scheme, with a time step of 0.01.

The observed trajectory is generated by the model with these standard parameter values. Two different imperfect models are created with parameter values that deviate about 30% from the standard parameter values, as denoted in Table 1. In the Appendix it is explained why only two different imperfect models are considered and how the imperfect parameter values are chosen.

Table 1: Standard and perturbed parameter values for the Lorenz 63 system.

Model	σ	ρ	β
Truth	10	28	8/3
Model 1	12.25	19	3.3
Model 2	7.5	35	1.9

The behavior of these imperfect models is quite different from the truth as can be seen in Fig. 2. Two stable fixed points characterize the attractor of model 1. Model 2 has a chaotic attractor that resembles the truth, but its mean is shifted towards higher z values.

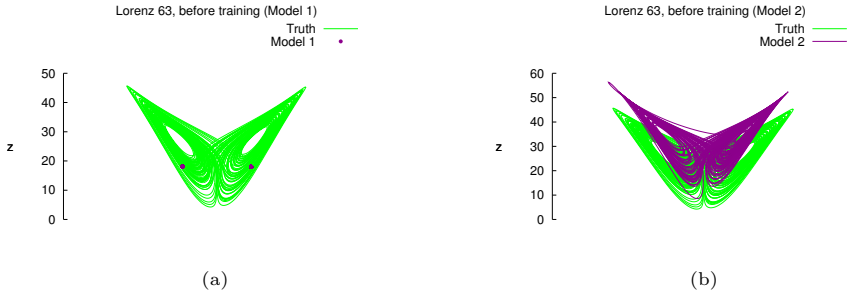


Figure 2: Trajectories of the imperfect models (purple), together with the true trajectory (green).

The training period T is chosen to be 200 time steps, enough to revolve about two times around the unstable fixed points. The number of iterations is 100. The same part of the attractor is used for training in every iteration.
The weights w_i, $i \in \{1,2\}$ that are the result of the training phase are listed in Table 2. They determine the superposition of the imperfect models (Eq. 2). For all three coordinates x, y, z, they sum up to 1.

$$\dot{x}_{super} = \sum_{i=1}^{2} w_i^x \dot{x}_i \tag{2a}$$

$$\dot{y}_{super} = \sum_{i=1}^{2} w_i^y \dot{y}_i \tag{2b}$$

$$\dot{z}_{super} = \sum_{i=1}^{2} w_i^z \dot{z}_i \tag{2c}$$

After 45 iterations, the weights for $\dot{y}$ and $\dot{z}$ do not change anymore. The weights for $\dot{x}$ after 100 iterations are still not constant, but the values differ only from the third decimal onwards.

Table 2: Weights of the supermodel of the Lorenz63 system.

Model	w_i^x	w_i^y	w_i^z
$i=1$	0.5248	0.4385	0.5491
$i=2$	0.4752	0.5615	0.4509

In the case of the Lorenz 63 system, the superposition of the imperfect Lorenz 63 models again forms a Lorenz 63 system, because the parameter values σ, ρ, β appear linearly in the differential equations. Hence, the supermodel is a Lorenz 63 system for which the parameter values can be calculated. The supermodel parameters are almost

perfect, as is shown in Table 3. This is possible because for all three perturbed parameters, one of the models has an imperfect parameter value smaller than the standard parameter value and the other model has one that is larger (Table 1). Hence, for each of the parameters, one can find a linear combination of the imperfect parameter values with positive weights whose sum is equal to 1, that represents the standard parameter value (Eq. 3).

$$\sigma_{super} = \sum_{i=1}^{2} w_i^x \sigma_i \tag{3a}$$

$$\rho_{super} = \sum_{i=1}^{2} w_i^y \rho_i \tag{3b}$$

$$\beta_{super} = \sum_{i=1}^{2} w_i^z \beta_i \tag{3c}$$

Table 3: Parameter values of the truth and the supermodel.

Model	σ	ρ	β
Truth	10	28	8/3
Supermodel	9.993	27.983	2.669

If this supermodel is integrated for a long time period, the attractor of the supermodel and the truth look quite similar, as can be seen in Fig. 3.

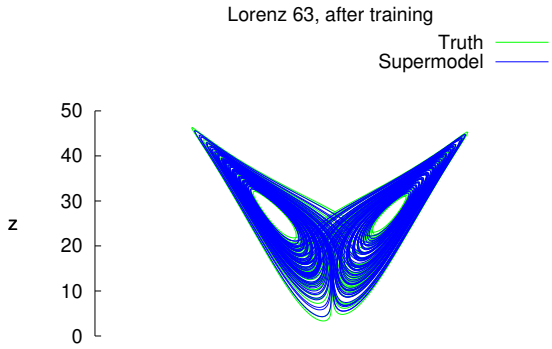


Figure 3: Trajectory of the supermodel (blue), together with the true trajectory (green).

3.1 Climate measures

Straightforward measures to compare the attractor of the supermodel and the truth are the mean, standard deviation and covariance. The calculation of these statistics is based

on 500 runs of 5000 time steps. The error estimation of a 95% confidence interval is also calculated. In Table 4, it can be seen that the statistics of both the true and the supermodel attractor are very similar. In particular, the standard deviations for each of x, y and z are the same up to the first decimal. The largest differences are in the covariance between x, z and y, z. However, these differences are within the 95% uncertainty intervals and are thus not significant. The sizes of all confidence intervals for both the truth and the supermodel are almost identical.

Table 4: The mean, standard deviation (SD) and covariance for the truth and the supermodel. The 95% *error estimation is given in brackets.*

	Truth	**Supermodel**
Mean x	0.073 (0.099)	0.033 (0.098)
Mean y	0.073 (0.099)	0.034 (0.098)
Mean z	23.552 (0.012)	23.528 (0.012)
SD x	7.843 (0.010)	7.844 (0.009)
SD y	8.939 (0.011)	8.942 (0.010)
SD z	8.618 (0.012)	8.623 (0.012)
Cov. xy	61.529 (0.150)	61.547 (0.148)
Cov. xz	0.189 (0.266)	0.057 (0.268)
Cov. yz	0.247 (0.336)	0.109 (0.334)

3.2 Forecast quality

Along with the measures of the climate statistics of the models, a measure for the quality of the "weather prediction" can also be constructed. This measure reflects the forecast quality of the models on shorter time scales. The squared Euclidean distance between the true trajectory and the trajectory of a model with a slightly perturbed initial condition is calculated and averaged over a number of forecasts, as shown in Fig. 4. On the true attractor, this value converges for large enough a forecast time T to a value corresponding to the average distance between two arbitrary states. This distance is used to normalize the measure of the forecast quality.

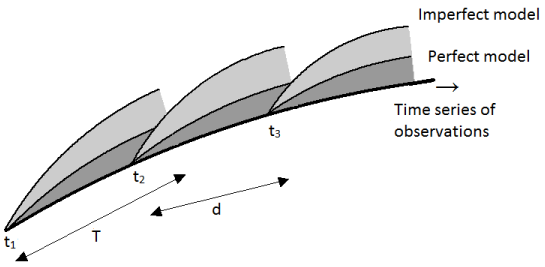


Figure 4: Measure of the forecast quality. At times t_i a short integration of time T starts from observed initial conditions and slightly perturbed conditions. The fixed time interval between times t_i is denoted by d.

The initial perturbation is chosen in the order of 10^{-1}. The number of forecasts is equal to 1000, and the distance between the initial states d is 10 time steps. Figure 5 shows that the ability of the supermodel and the true model to predict the observed truth is about the same. In comparison, the imperfect models lose their prediction skill very quickly.

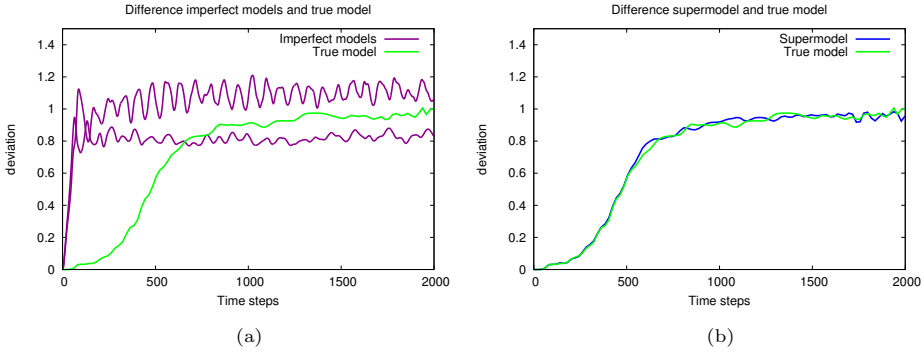


Figure 5: Forecast quality of the imperfect Lorenz 63 models (purple) and the supermodel (blue) compared to the true Lorenz 63 model (green).

4 Results for a quasi-geostrophic model

Given the encouraging results from the Lorenz 63 system, the CPT method is next applied to a more complex model with 1449 degrees of freedom: a three level quasi-geostrophic (QG) global atmosphere model developed by *Marshall and Molteni* (1993). The model solves the quasi-geostrophic potential vorticity equation on the sphere using

a spectral method with spherical harmonic functions. A triangular T21 truncation is used. The performance of this model is quite realistic. According to *Corti et al.* (1997), the simulation of teleconnections and blockings in the Pacific and Atlantic regions is "surprisingly accurate". The evolution of the quasi-geostrophic potential vorticity at the three levels is given by

$$\dot{q_1} = \mathcal{J}(\psi_1, q_1) - D_1(\psi_1, \psi_2) + S_1, \tag{4a}$$

$$\dot{q_2} = \mathcal{J}(\psi_2, q_2) - D_2(\psi_1, \psi_2, \psi_3) + S_2, \tag{4b}$$

$$\dot{q_3} = \mathcal{J}(\psi_3, q_3) - D_3(\psi_2, \psi_3) + S_3, \tag{4c}$$

where q is the potential vorticity, ψ the streamfunction, $D(\psi)$ a linear operator that represents dissipative terms and S a constant potential vorticity (PV) source. The index i refers to the pressure level. Here, the potential vorticity is defined as

$$q_1 = \nabla^2\psi_1 - R_1^{-2}(\psi_1 - \psi_2) + f, \tag{5a}$$

$$q_2 = \nabla^2\psi_2 + R_1^{-2}(\psi_1 - \psi_2) - R_2^{-2}(\psi_2 - \psi_3) + f, \tag{5b}$$

$$q_3 = \nabla^2\psi_3 + R_2^{-2}(\psi_2 - \psi_3) + f(1 + \frac{h}{H_0}), \tag{5c}$$

where f is the Coriolis parameter, R_1 and R_2 are the Rossby radii of deformation of the 200-500 hPa and 500-800 hPa layer, respectively, h is the orographic height, and H_0 is a scale height. To create different imperfect models, three parameter values are varied:

- τ_E timescale in days of the Ekman damping (τ_E in Eq. (A11) of *Marshall and Molteni* (1993))
- R_1 Rossby radius of deformation of the 200-500 hPa layer
- R_2 Rossby radius of deformation of the 500-800 hPa layer

Four different imperfect models are used for the CPT training phase; their parameters are denoted in Table 5. The imperfect values of the Rossby radii of deformation are chosen to differ by only a few thousandths from the true value since even a small deviation leads to very different behavior of the model. Numerical solutions are obtained by using a fourth-order Runge-Kutta time stepping scheme, with a time step of 1/36 day.

Table 5: Parameter values of the imperfect QG-models.

Model	τ_E	R_1	R_2
Truth	2.0	0.1150	0.0720
Model 1	1.5	0.1165	0.0705
Model 2	1.5	0.1130	0.0725
Model 3	2.4	0.1130	0.0705
Model 4	2.4	0.1165	0.0725

The training period T is 100 time steps, which corresponds to an integration period of about 3 days. Most of the development of weather systems can be captured within 3 days. The number of iterations is 20. With every iteration, a new part of the attractor is used for training by continuing the observed trajectory to get a better sampling of the attractor.

In Table 6 the resulting weights for the different levels are shown. After 20 iterations, the weights are not completely converged, they differ a few percent per iteration, but there is no increasing or decreasing trend. Note that at the 200 hPa level the superposition of models consists solely of model 1 and model 2. The only parameter with imperfections affecting this level is R_1, and the imperfect value of this parameter is equal for models 1 and 4. The same holds for models 2 and 3. Since at every time step in the CPT training, every model receives the same state, the tendencies of models 1 and 4 are the same at this level and the same holds for models 2 and 3. Therefore the corresponding weights of models 3 and 4 are 0 since these are never chosen during the CPT training.

Table 6: Weights of the imperfect QG-models at 200, 500 and 800 hPa.

Model	w_i^{200}	w_i^{500}	w_i^{800}
$i=1$	0.653	0.217	0.093
$i=2$	0.347	0.459	0.235
$i=3$	0.000	0.157	0.215
$i=4$	0.000	0.167	0.457

In an additional experiment we left out the imperfect model with the poorest short- and long-term predictability in order to test the hypothesis that the addition of a relatively bad model can still improve the quality of the supermodel solution. The same imperfect models are used. The model with the poorest predictability is model 1 (Table 7), so the supermodel is constructed out of models 2, 3 and 4. Note that these three models still span the same uncertainty range in the three parameters. The same CPT training phase is applied.

The CPT training provides weights that determine a superposition of models that is capable of following observed trajectories more closely. But to what extend do the values of these weights matter? Is training really necessary? In order to assess this we evaluated the quality of a supermodel with equal weights given to each imperfect model in the superposition.

4.1 Climate measures

As measure of the long-term behavior of the quasi-geostrophic model we choose to compare the geostrophic winds of the different models. The potential vorticity calculated by the model determines these winds. The true model, imperfect models and supermodel are integrated over 900 days in a perpetual winter simulation.
As statistical measure (RMSE), the errors in the 900-day average wind strength at the 200, 500 and 800 hPa level at each location are averaged over the globe:

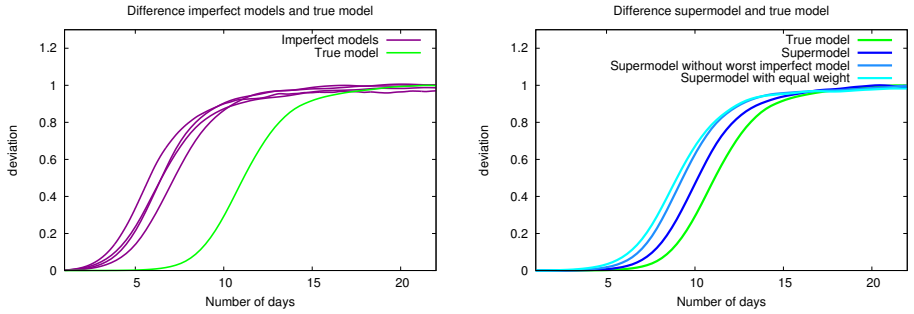
$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\|\mathbf{u}_i^{truth}\| - \|\mathbf{u}_i^{mod}\|)^2},$$

with i denoting the grid point, u the zonal wind, v the meridional wind and N the total number of grid points.

Table 7: The root mean squared error of the wind strength (ms^{-1}) over 900 winter days. For the true model, the average RMSE is given. The value for which 95% of the RMSE values is below that value is given in brackets.

Model	**800 hPa**	**500 hPa**	**200 hPa**
Model 1	1.92	1.95	2.27
Model 2	1.80	1.37	2.31
Model 3	1.10	0.90	1.79
Model 4	1.42	1.36	2.06
True model	0.48 (0.65)	0.78 (0.92)	1.66 (2.05)
Supermodel	0.45	0.80	1.77
Supermodel equally weighted	1.56	1.51	2.63
Supermodel without worst imperfect model	1.42	1.38	2.09

We take a Monte Carlo approach to assess the uncertainty of the RMSE values. For 98 different initial conditions, a trajectory of 900 days is computed with the true model. Then the RMSE is calculated for these trajectories with respect to one other true trajectory of observations. The 95% percentile of these values is listed in Table 7. This table reveals that, with respect to this climate measure, the supermodel is indistinguishable from the true model. The RMSE values of the imperfect models are significantly larger. Note that the supermodel was not trained to reproduce the observed mean state, but apparently training on a 3-day time scale is sufficient.
The RMSE values of the supermodel without the inclusion of the worst model are comparable with the values of the second-best imperfect model (model 4). The RMSE values of the equally weighted supermodel are even worse.



(a) (b)

Figure 6: Forecast quality of imperfect QG models (purple), the QG supermodel (dark blue), the QG supermodel generated without the worst imperfect model in forecast quality (medium blue) and the QG supermodel generated with equal weights (light blue) compared to the forecast quality of the true QG model (green).

4.2 Forecast quality

As was done for the Lorenz 63 system, the forecast quality can be measured by calculating the mean squared error between the true trajectory and the trajectory of a model with a slightly perturbed initial condition and then averaging this over a number of forecasts. The mean squared error is taken over all three levels and all spectral coefficients. The number of forecasts is 100, and the distance d between the initial states is 1000 time steps. In *LORENZ* (1969) it is mentioned that an initial perturbation in the smallest length scale leads to large errors in all scales after 2 weeks. For that reason the initial perturbation is chosen in the order of 10^{-4} as it leads to an almost complete loss of predictability after 14 days.

The forecast quality of the supermodel is not as good as that of the true model, but the supermodel greatly improves the predictability as compared to the imperfect models (Fig. 6).

The forecast quality of the supermodel without the inclusion of the worst model also improves compared to the imperfect models, but is clearly not as good as the forecast quality of the supermodel with inclusion of this worst model (Fig. 6). Thus, inclusion of relatively bad models can still contribute towards a superior supermodel.

The equally weighted supermodel turns out also to perform better than the imperfect models (Fig. 6), but significantly worse than the supermodel with the weights trained by CPT. Hence, training does add value to the quality of the supermodel.

5 Conclusions

In this study we have demonstrated that a new training method based on cross pollination in time leads to a dynamical combination of forecast models (a weighted supermodel) with superior forecast quality and improved climatology. The CPT training is based on short-term trajectories only, but it turned out that the errors in the climatology are also greatly reduced. The results indicate that a supermodel with weights trained by CPT can give significantly better predictions than a supermodel consisting of the same imperfect models with equal weights.

State-of-the-art models are far more complex than the examples from this paper, but in principle the approach is applicable to state-of-the-art models as well. With an increased number of uncertain parameters, it is to be expected that more imperfect models are required to construct a supermodel with improved prediction skill. This will increase the amount of computation time, but if during the CPT training, the number of trajectories is pruned back to a single prediction, the computational cost of CPT grows only proportionally to the number of imperfect models.

In this study the imperfect models differed in parameter values only but were structurally identical. In reality, imperfect state-of-the-art weather models differ in structure, generally solving different equations on different grids using different numerical methods. In this case, methods from data assimilation might be used in order to cross states between models, as is done by *Du and Smith* (2017). Alternatively, a common state space might be defined, with models projecting their states into this common state space and the CPT training limited to this common state space.

In the case when a supermodel solution hardly improves the prediction skill as compared to the imperfect models, one might experiment with the introduction of an additional imperfect model that has in some sense the "opposite" error behavior as compared to the other imperfect models. This additional imperfect model can have worse prediction skill, but it might still contribute to a superior supermodel solution. For the quasi-geostrophic atmosphere model in this study, it was demonstrated that a model with poor forecast quality still contributed towards an improved supermodel.

A remarkable result of this study is that even if only a relatively small part of the attractor is used for training, the method results in a supermodel with improved climatology. There is evidence in *Rodwell and Palmer* (2007) that climatological errors develop quickly during the first few days of weather forecasts, implying that a short-term training can reduce climatological errors. If this result carries over to the state-of-the-art models, then computationally expensive long climate simulations as in *Shen et al.* (2016) can be avoided during training. Using relatively short integrations only can still

improve the climatology of a supermodel.

As indicated above, there are several ways to apply and further develop the CPT training methodology presented in this study. It is not only applicable to weather and climate models, but also to numerical models of other complex systems, for example economical or biological models. Cross pollination in time as applied in this paper is a promising approach for combining models dynamically in order to further improve predictions.

Data availability

No data sets were used in this article.

Appendix A

The supermodeling approach only works well if the imperfect models are on "opposite" sides of the truth. We took this into account in the choice of the imperfections. The imperfect values of the parameters and the number of imperfect models is based on the convex hull principle. In one dimension this convex hull principle basically says that if there is one parameter value σ_1 smaller than the true value σ and one parameter value σ_2 larger than the true value, there are positive weights w_1, w_2 such that a linear combination $w_1\sigma_1 + w_2\sigma_2$ is exactly equal to σ. In case of Lorenz63, the equations for $\dot{x}, \dot{y}$ and $\dot{z}$ each contain only one parameter that appears linearly in the equation. Since we apply different weights for the different equations for $\dot{x}, \dot{y}$ and $\dot{z}$, per equation we need only two imperfect models to be able to reconstruct exactly the true parameter value with positive weights. This convex hull principle can be extended to more dimensions:

Definition 1. *Let $\boldsymbol{x}^1, ..., \boldsymbol{x}^k$ be vectors in $\mathbb{R}^n$ and let $\lambda_1, ..., \lambda_k$ be nonnegative scalars whose sum is unity.*
***(a)** The vector $\sum_{i=1}^{k} \lambda_i \boldsymbol{x}^i$ is said to be a **convex combination** of the vectors $\boldsymbol{x}^1, ..., \boldsymbol{x}^k$.*
***(b)** The **convex hull** of the vectors $\boldsymbol{x}^1, ..., \boldsymbol{x}^k$ is the set of all convex combinations of these vectors.*

In this definition, the vectors $\mathbf{x}^i, i \in 1, ...k$ represent the imperfect parameter values x per model i and $\lambda_i, i \in 1, ...k$ the corresponding weights. This convex hull generalizes the 'in between' concept for one dimension. To be able to reproduce the n-dimensional vector $\mathbf{x}$, it has to lie inside the convex hull of vectors $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^k$.
We can write this as a matrix-vector equation, where the last row indicates that the sum of the weights has to be equal to 1 and the vector $\mathbf{x}$ represents the true parameter values:

$$\begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^{n+1} \\ x_2^1 & x_2^2 & \cdots & x_2^{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^{n+1} \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \lambda_{n+1} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{pmatrix}$$

For parameter vectors of size n, we have $n+1$ constraints, since also the sum of the weights has to equal 1. Hence we know that to be able to reproduce the true parameter vector $\mathbf{x}$, for n parameters that appear linearly in one differential equation for a state variable, $n+1$ linearly independent vectors of these parameters are needed which form a convex hull around the true parameter vector.

For the quasi-geostrophic model, the imperfect parameters do not appear linearly in the equations. Therefore choosing the parameter perturbations such that they form a convex hull around the true parameter values will not necessarily result in a model that reproduces the truth. Nevertheless, in practice we found that this approach still worked well. In this case, choosing the imperfect parameter values on opposite sides of the truth created "opposite" behaviour such that the imperfect models could compensate for each other.

Competing interests

The authors declare that they have no conflict of interest.

Edited by: C. Franzke
Reviewed by: G. Hu and one anonymous referee

References

Branicki, M., and A. J. Majda (2015), An information-theoretic framework for improving imperfect dynamical predictions via multi-model ensemble forecasts, *Journal of Nonlinear Science*, *25*(3), 489–538, doi:10.1007/s00332-015-9233-1. 1

Corti, S., A. Giannini, S. Tibaldi, and F. Molteni (1997), Patterns of low-frequency variability in a three-level quasi-geostrophic model. climate dynamics 13(12):883904 dee dp, uppala sm, simmons aj, berrisford, in *Verlag Donges JF, Zou Y, Marwan N, Kurths J (2009a) The backbone of the climate network. EPL (Europhysics Letters) 87(4):48,007 Donges JF, Zou Y, Marwan N, Kurths J (2009b) Complex networks in climate dynamics. The European Physical Journal-Special Topics*, Springer. 4

Du, H., and L. A. Smith (2017), Multi-model cross pollination in time, *Physica D, submitted*. 5

IPCC (2013), *Climate Change 2013: The Physical Science Basis*, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.) Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp]. 1

Lorenz, E. (1963), Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, *20*, 130–140. 1

LORENZ, E. N. (1969), The predictability of a flow which possesses many scales of motion, *Tellus*, *21*(3), 289–307, doi:10.1111/j.2153-3490.1969.tb00444.x. 4.2

Marshall, J., and F. Molteni (1993), Toward a dynamical understanding of planetary-scale flow regimes, *Journal of the Atmospheric Sciences*, *50*(12), 1792–1818. 1, 4, 4

Rodwell, M. J., and T. N. Palmer (2007), Using numerical weather prediction to assess climate models, *Q.J.R. Meteorol. Soc.*, (133), 129–146, doi:10.1002/qj.23. 5

Shen, M.-L., N. Keenlyside, F. Selten, W. Wiegerinck, and G. S. Duane (2016), Dynamically combining climate models to "supermodel" the tropical pacific, *Geophysical Research Letters*, *43*(1), 359–366, doi:10.1002/2015GL066562, 2015GL066562. 5

Smith, L. A. (2001), *Nonlinear Dynamics and Statistics*, chap. Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems, pp. 31–64, Mees, Alistair I., Birkhäuser Boston, Boston, MA, doi:10.1007/978-1-4612-0177-9_2. 1

van den Berge, L. A., F. M. Selten, W. Wiegerinck, and G. S. Duane (2011), A multi-model ensemble method that combines imperfect models through learning, *Earth System Dynamics*, *2*(1), 161–177, doi:10.5194/esd-2-161-2011. 1

Wiegerinck, W., M. Mirchev, W. Burgers, and F. Selten (2013), Consensus and synchronization in complex networks, chap. Supermodeling Dynamics and Learning Mechanisms, pp. 227–255, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-642-33359-0_9. 1

Paper II

II

Simulating climate with a synchronization-based supermodel

Frank M. Selten, Francine J. Schevenhoven and Gregory S. Duane

Chaos, **27**,126903 (2017)

Simulating climate with a synchronization-based supermodel

Frank M. Selten[1], Francine J. Schevenhoven[2,3], and Gregory S. Duane[2,4]

[1]Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
[2]Geophysical Institute, University of Bergen, Bergen, Norway
[3]Bjerknes Centre for Climate Research, Bergen, Norway
[4]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, USA

Abstract

The SPEEDO global climate model (an atmosphere model coupled to a land and an ocean/sea-ice model with about 250.000 degrees of freedom) is used to investigate the merits of a new multi-model ensemble approach to the climate prediction problem in a perfect model setting. Two imperfect models are generated by perturbing parameters. Connection terms are introduced that synchronize the two models on a common solution, referred to as the supermodel solution. A synchronization-based learning algorithm is applied to the supermodel through the introduction of an update rule for the connection coefficients. Connection coefficients cease updating when synchronization errors between the supermodel and solutions of the "true" equations vanish. These final connection coefficients define the supermodel. Different supermodel solutions, but with equivalent performance, are found depending on the initial values of the connection coefficients during learning. The supermodels have a climatology and a climate response to a CO2 increase in the atmosphere that is closer to the truth as compared to the imperfect models and the standard multi-model ensemble average, showing the potential of the supermodel approach to improve climate predictions.

Complex numerical codes are being used to predict the behavior of real-world phenomena like the climate or the economy. In this study, we demonstrate that predictions can be improved by forming an ensemble of inter-connected different imperfect climate models that synchronize on a common solution, referred to as the supermodel solution. This supermodel solution depends on the connections. The connections are trained using observations of the truth such that the supermodel minimizes synchronization errors when nudged to trajectories of the truth. This is the first time that the potential of the supermodel approach is demonstrated in the context of a complex global climate model. Due to its computational efficiency, the synchronization-based learning approach is applicable to state-of-the-art climate models with millions degrees of freedom and historical observations of the Earth's global climate system.

1 Introduction

Global climate models are complex numerical codes that integrate coupled sets of ordinary differential equations (ODEs) with prescribed time-dependent forcing terms in time in order to produce projections of our future climate. It is commonly found that a multi-model averaged climatology is closer to the observed climatology, which is defined as the average over 30 years and is referred to as the climate normal. Model estimates tend to be distributed around the truth, and therefore, averaging across models helps in reducing errors in the simulated mean state (*Weigel et al.*, 2008). Although improved climate statistics are useful, climate adaptation and impact studies often require climate trajectories as input (*Hazeleger et al.*, 2015). However, averaging trajectories from multiple models without synchronization leads to undesired smoothing and variance reduction.

Here we follow an alternative, synchronization-based approach that produces improved climate trajectories by combining climate models dynamically. This approach was inspired by a study in which two different atmosphere models were coupled to a single ocean model leading to improved climate simulations when the ocean exchanged heat with only one atmosphere model and momentum with the other (*Kirtman et al.*, 2003). Here, we dynamically combine models by introducing connection terms into the model equations that nudge the state of one model to the state of every other model in the ensemble, effectively forming a new dynamical system with the values of the connection coefficients as additional parameters. For appropriately chosen connections, the models synchronize on a common solution that depends on the values of the connection coefficients (*Hiemstra et al.*, 2012; *Lunkeit*, 2001). This solution is referred to as the supermodel solution (*van den Berge et al.*, 2011). During a learning phase, the supermodel is nudged to an observed trajectory and the connection coefficients are adjusted by update rules that depend on the synchronization error. The connection coefficients cease to update when the synchronization is perfect.

So far, the supermodel approach was pioneered using relatively simple dynamical systems only (*Duane*, 2015; *Mirchev et al.*, 2012; *van den Berge et al.*, 2011) or with very limited inter-model connections in a global climate model context (*Shen et al.*, 2016). Here, we demonstrate the potential of a fully connected supermodel constructed from versions of a complex global climate model. This model, SPEEDO (*Severijns and Hazeleger*, 2010) is described in section 2 which is followed by a discussion of the supermodel implementation using SPEEDO in section 3. The synchronization-based learning is explained in section 4 and is applied to the SPEEDO supermodel in section 5. In section 6 we discuss the merits of the supermodel approach in relation to the stan-

dard multi model ensemble approach. We conclude this paper with a summary of open issues in section 7 and discuss the application of the synchronization-based supermodel approach for state-of-the-art weather and climate models.

2 SPEEDO climate model

The SPEEDO climate model consists of an atmospheric component (SPEEDY) that exchanges information with a land (LBM) and an ocean-sea-ice component (CLIO) using coupling routines (Fig. 1). The coupling routines perform re-gridding operations between the computational grids of the different modules. A detailed description of SPEEDO can be found in *Severijns and Hazeleger* (2010). The atmospheric model SPEEDY solves the primitive equations on a sphere using a spectral method (*Molteni*, 2003). The primitive equations are derived from the Navier-Stokes equations with suitable approximations for atmospheric flow at scales larger than a few kilometers (*Zdunkowski and Bott*, 2003). The spectral expansion into spherical harmonics is truncated at total wavenumber 30 which corresponds to a spatial resolution at the equator of about 700 km. It has 8 vertical levels and relatively simple representations of radiation, convection, clouds, precipitation and turbulent heat, water, and momentum exchange at the surface. The solar radiation follows the daily and seasonal cycle. In principle the model consists of 31680 coupled ODE's for the spectral coefficients of the two horizontal wind components U (east-west) and V (north-south), temperature T and specific humidity q at the 8 vertical levels and the log of surface pressure p_s. Calculations in physical space are performed on a Gaussian grid with approximately 3.75 degree spacing (48×96 grid-cells). Speedy exchanges water and heat at the 2115 land points of the land model LBM that uses three soil layers and up to two snow layers to close the hydrological cycle over land and a heat budget equation that controls the land temperatures. The horizontal discretisation is the same as for the atmosphere model. The land surface reflection coefficient for solar radiation is prescribed using a monthly climatology. Each land bucket has a maximum soil water capacity. The runoff is collected in river-basins and drained into the ocean at specific locations of the major river outflows. SPEEDY exchanges heat, water, and momentum with the ocean model CLIO (*Goosse and Fichefet*, 1999). CLIO solves the primitive equations on a computational grid of 3° horizontal resolution and 20 unevenly spaced layers in the vertical. It has a rotated grid over the North Atlantic ocean in order to circumvent the singularity at the pole. It has a free-surface and is coupled to a three-layer thermodynamic-dynamic sea-ice model. The sea-ice model takes into account the heat storage in the snow-ice system and calculates the changes of snow and ice thickness in response to surface and bot-

tom heat fluxes. In the computation of ice-dynamics, sea ice is considered to behave as a viscous-plastic continuum as it moves under the action of winds and ocean currents. In total, CLIO has about 200.000 degrees of freedom. The SPEEDO equations can be written as

$$\dot{\mathbf{a}} = \mathbf{f}^a(\mathbf{a};\mathbf{p}^a) + \mathbf{g}^a(\mathbf{e}^h, \mathbf{e}^w, \mathbf{e}^m) \tag{1a}$$

$$\dot{\mathbf{o}} = \mathbf{f}^o(\mathbf{o};\mathbf{p}^o) + \mathbf{g}^o(\mathcal{P}^o\mathbf{e}^h, \mathcal{P}^o\mathbf{e}^w, \mathcal{P}^o\mathbf{e}^m, \mathcal{P}^o\mathbf{r}) \tag{1b}$$

$$\dot{\mathbf{l}} = \mathbf{f}^l(\mathbf{l};\mathbf{p}^l) + \mathbf{g}^l(\mathcal{P}^l\mathbf{e}^h, \mathcal{P}^l\mathbf{e}^w, \mathbf{r}) \tag{1c}$$

Here, we formulate the model in terms of ordinary differential equations (ODEs) on a grid, instead of the more usual partial differential equations (PDEs), to be explicit about the numerical scheme and also for consistency with the ODE scheme for learning inter-model connections which is presented in Section IV. The bold lowercase characters represent vectors with **a** the atmospheric state vector, **o** the ocean/sea-ice state vector, **l** the land state vector, $\mathbf{e}^h$ the heat exchange vector between the atmosphere and the surface, $\mathbf{e}^w$ the water exchange vector, $\mathbf{e}^m$ the momentum exchange vector and **r** the river outflow vector describing the flow of water from land to ocean. The exchange vectors depend on the state of the atmosphere and the surface. The projection operators $\mathcal{P}$ represent the re-gridding operations between the computational grids. These operations are conservative so that the globally integrated heat and water loss of the atmosphere at any time at the surface equals the integrated heat and water gain of the land and ocean. The non-linear functions **f** represent the cumulative contribution of the modelled physical processes to the change in the climate state vector and depend on the values of the parameter vectors **p**. Some of these parameters go through a daily and/or seasonal cycle and/or have a spatial dependence like the reflectivity of the surface. The non-linear functions **g** describe how the exchange of heat, water, and momentum between the subsystems affects the change of the climate state vector.

3 Supermodel

In this study, we consider the SPEEDO climate model with standard parameter values as "truth" and create imperfect models of this truth by perturbing parameter values in the atmospheric component. A supermodel is formed by connecting two imperfect atmosphere models through linear nudging terms that nudge the state of model 1 to model 2 and vice versa (see Fig. 2 and Eq. 2) and couple them to the same ocean and land model. Both atmosphere models receive the same state information from the ocean and land model and each calculates its own water, heat, and momentum exchange. The ocean and land model receive the exchanges from both atmospheres and use the

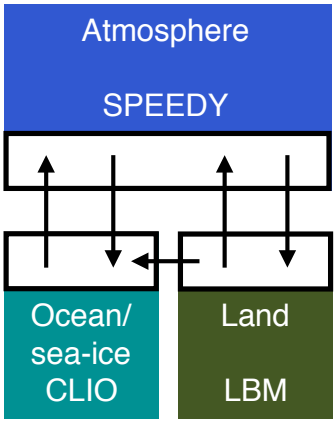


Figure 1: Schematic representation of the SPEEDO climate model. The atmosphere needs surface characteristics (temperature, roughness, reflectivity and soil moisture) in order to calculate the exchange of heat, water, and momentum. Coupler software communicates this information between the components and interpolates between the computational grids.

average of the two as input. The supermodel state $\mathbf{a}_s$ at any time is defined as the average of the two model states. For moderate inter-model nudging, it is found that the atmospheres synchronize nearly completely, and therefore, by taking the averaged state, no significant spatial or temporal smoothing is introduced. The SPEEDO supermodel equations are given by

$$\dot{\mathbf{a}}_1 = \mathbf{f}^a(\mathbf{a}_1;\mathbf{p}_1^a) + \mathbf{g}^a(\mathbf{e}_1^h,\mathbf{e}_1^w,\mathbf{e}_1^m) - \mathbf{C}_{12}[\mathbf{a}_1 - \mathbf{a}_2] \tag{2a}$$

$$\dot{\mathbf{a}}_2 = \mathbf{f}^a(\mathbf{a}_2;\mathbf{p}_2^a) + \mathbf{g}^a(\mathbf{e}_2^h,\mathbf{e}_2^w,\mathbf{e}_2^m) - \mathbf{C}_{21}[\mathbf{a}_2 - \mathbf{a}_1] \tag{2b}$$

$$\dot{\mathbf{o}} = \mathbf{f}^o(\mathbf{o};\mathbf{p}^o) + \mathbf{g}^o(\mathcal{P}^o\overline{\mathbf{e}^h},\mathcal{P}^o\overline{\mathbf{e}^w},\mathcal{P}^o\overline{\mathbf{e}^m},\mathcal{P}^o\mathbf{r}) \tag{2c}$$

$$\dot{\mathbf{l}} = \mathbf{f}^l(\mathbf{l};\mathbf{p}^l) + \mathbf{g}^l(\mathcal{P}^l\overline{\mathbf{e}^h},\mathcal{P}^l\overline{\mathbf{e}^w},\mathbf{r}) \tag{2d}$$

$$\mathbf{a}_s = \frac{1}{2}[\mathbf{a}_1 + \mathbf{a}_2] \tag{2e}$$

where $\mathbf{C}$ denotes a diagonal matrix with connection coefficients on the diagonal, the subscripts refer to the respective models, and the overbar denotes an average over the two models.

The atmospheric components of the supermodel equation can be written as

$$\begin{aligned}\dot{\mathbf{a}}_s = &\frac{1}{2}[\mathbf{f}^a(\mathbf{a}_1;\mathbf{p}_1^a) + \mathbf{f}^a(\mathbf{a}_2;\mathbf{p}_2^a)] + \\ &\frac{1}{2}[\mathbf{g}^a(\mathbf{e}_1^h,\mathbf{e}_1^w,\mathbf{e}_1^m) + \mathbf{g}^a(\mathbf{e}_2^h,\mathbf{e}_2^w,\mathbf{e}_2^m)] + \\ &\frac{1}{2}[\mathbf{C}_{12} - \mathbf{C}_{21}][\mathbf{a}_2 - \mathbf{a}_1]\end{aligned} \tag{3}$$

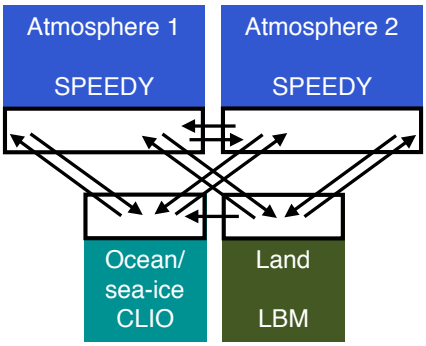


Figure 2: Schematic representation of the SPEEDO climate supermodel. The two imperfect atmosphere models exchange water, heat, and momentum with the perfect ocean and land models. The ocean and land models send their state information to both atmosphere models. The atmosphere models exchange state information with each other.

From this equation, a number of interesting observations can be made. For equal connection coefficients ($\mathbf{C}_{12} = \mathbf{C}_{21}$) the last term vanishes and the supermodel solution becomes equal to the average of both imperfect model solutions. If the synchronization is perfect ($\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_s$), then the supermodel solution obeys the averaged imperfect model equations with equal weights $\frac{1}{2}$. Solving the weighted averaged equations is referred to as weighted supermodeling (*Wiegerinck and Selten*, 2017) as distinct from connected supermodeling. For unequal connection coefficients ($\mathbf{C}_{12} \neq \mathbf{C}_{21}$), with, for instance, model 1 more strongly nudged to model 2 than vice versa, the last term is non-zero and systematically pushes the connected supermodel solution every time-step toward the state of model 2 (see Fig. 3).

Negative connection values imply that the model solutions are driven apart. This is undesired as the aim is to synchronize the models on a common solution. Without this constraint, the model with the negative coefficient will be driven away from the other model solution. However, the two solutions could still remain close together, if the other model gets a positive connection coefficient with larger value and chases the model with the negative coefficient. In this study we will restrict to positive connection coefficients only.

4 Synchronization-based Learning

The supermodel solution (Eq. 2) depends on the choice of the connection coefficient values $\mathbf{C}$. A learning algorithm that extends synchronization of states to synchronization of parameters is applied in order to train the supermodel to follow trajectories from the truth more closely as explained in section 4.2.

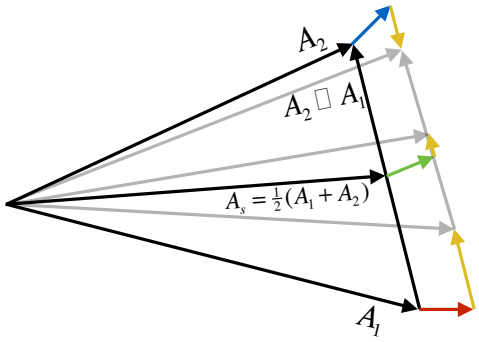


Figure 3: Graphical representation of one time-step of the connected supermodel. Black arrows denote state vectors at initial time, grey arrows one time-step later. The change in the supermodel state (green vector) is the averaged change of model 1 (red vector) and model 2 (blue vector) plus a vector due to the nudging terms (yellow vectors) pointing in the direction of the model 2 state, assuming that the model 1 state is more strongly pushed to the model 2 state than vice versa.

4.1 From state synchronization to parameter synchronization

Suppose we have two coupled dynamical systems:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\mathbf{p}) \tag{4a}$$

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y};\mathbf{q}) - \mathbf{K}(\mathbf{y}-\mathbf{x}) \tag{4b}$$

where $\mathbf{p}$ and $\mathbf{q}$ are vectors of parameters, $\mathbf{K}(\mathbf{y}-\mathbf{x})$ is a nudging term that couples the two systems, and $\mathbf{K}$ is a diagonal matrix of nudging coefficients. We will assume that the equations are linear in the parameters $\mathbf{p}$.

Suppose that when the two systems are identical, i.e. when $\mathbf{p} = \mathbf{q}$, the two systems synchronize, that is, as $t \to \infty$, $\mathbf{y}(t) \to \mathbf{x}(t)$. We want to find some rule for varying the parameters $\mathbf{q}$, i.e. a dynamical equation such that even if the two systems are not identical, $\mathbf{p} \neq \mathbf{q}$, the systems will still synchronize, and the parameters will become equal, that is, as $t \to \infty$, $\mathbf{y}(t) \to \mathbf{x}(t)$, and $\mathbf{q}(t) \to \mathbf{p}$. The problem is that of "adaptive observers" in the electrical engineering literature (*Besançon et al.*, 2006; *Zhang*, 2002). First, for concreteness, before stating the general rule for parameter estimation, we show how such a rule might be derived for the simple case of two connected Lorenz systems:

$$\dot{x_1} = \sigma(x_2 - x_1) \qquad \dot{y_1} = \sigma(y_2 - y_1) - c(y_1 - x_1) \tag{5a}$$

$$\dot{x_2} = \rho x_1 - x_2 - x_1 x_3 \qquad \dot{y_2} = \tilde{\rho} y_1 - y_2 - y_1 y_3 \tag{5b}$$

$$\dot{x_3} = -\beta x_3 + x_1 x_2 \qquad \dot{y_3} = -\beta y_3 + y_1 y_2 \tag{5c}$$

where the subscripts refer to the state vector elements.

If $\tilde{\rho} = \rho$, the two dynamical systems are identical and are known to synchronize for ap-

propriately chosen coupling constant c (*Pecora and Carroll*, 1990; *Yang et al.*, 2006): As $t \to \infty$, $(y_1(t), y_2(t), y_3(t)) \to (x_1(t), x_2(t), x_3(t))$, and $\mathbf{e}(t)^2 \to 0$, where the synchronization error $\mathbf{e} := (y_1 - x_1, y_2 - x_2, y_3 - x_3)$. We claim that that in the case of non-identical systems, with $\tilde{\rho} \neq \rho$, we can still arrange for synchronization if we can allow $\tilde{\rho}$ to vary as a new dynamical variable, specifically introducing the dynamical equation for $\tilde{\rho}$:

$$\begin{aligned} \dot{\tilde{\rho}} &= -(y_2 - x_2) y_1 \\ &= -e_2 y_1 \end{aligned} \tag{6}$$

To see why Eq. 6 implies $(y_1, y_2, y_3, \tilde{\rho}) \to (x_1, x_2, x_3, \rho)$, consider the Lyapunov function $L := (y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2 + (\tilde{\rho} - \rho)^2$. If we can show $L(t) \to 0$, we will have the desired state and parameter synchronization. Consider the time derivative

$$\dot{L} = \dot{L}_0 + 2\dot{\tilde{\rho}}(\tilde{\rho} - \rho) \tag{7}$$

where $L_0 := \mathbf{e}^2$ is the part of the Lyapunov function formed from state errors alone. The key point is that the time derivative of L_0 differs from its value for $\tilde{\rho} = \rho$, because the dynamical equation for y_2 is different. Specifically, in the derivative

$$\begin{aligned} \dot{L}_0 &= \frac{d}{dt}\left((y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2\right) \\ &= 2(y_1 - x_1)\frac{d}{dt}(y_1 - x_1) + 2(y_2 - x_2)\frac{d}{dt}(y_2 - x_2) \\ &\quad + 2(y_3 - x_3)\frac{d}{dt}(y_3 - x_3) \end{aligned} \tag{8}$$

(where we do not include parameter error explicitly), every term on right hand side is exactly of the same form as in the case of equal parameters, except for the term that contains y_2. The time-derivative factor in that term can be written as:

$$\begin{aligned} \frac{d(y_2 - x_2)}{dt} &= (\tilde{\rho} - \rho) y_1 + \rho y_1 - y_2 - y_1 y_3 \\ &\qquad - \rho x_1 + x_2 + x_1 x_3 \\ &= (\tilde{\rho} - \rho) y_1 + \frac{d(y_2 - x_2)}{dt}|_{\tilde{\rho}=\rho} \end{aligned} \tag{9}$$

where the last term is the value that the time-derivative would have if the parameters were equal. Substituting Eq. 9 into Eq. 8 gives

$$\begin{aligned} \dot{L}_0 &= 2(y_1 - x_1)\frac{d}{dt}(y_1 - x_1) + 2(y_2 - x_2)\frac{d}{dt}(y_2 - x_2)|_{\tilde{\rho}=\rho} \\ &\quad + 2(y_3 - x_3)\frac{d}{dt}(y_3 - x_3) + 2(y_2 - x_2)(\tilde{\rho} - \rho) y_1 \\ &= \dot{L}_0|_{\tilde{\rho}=\rho} + 2(y_2 - x_2)(\tilde{\rho} - \rho) y_1 \end{aligned} \tag{10}$$

Inserting Eq. 6 and Eq. 10 into Eq. 7, we have

$$\begin{aligned}\dot{L} &= \dot{L}_0|_{\tilde{\rho}=\rho} + 2(y_2 - x_2)(\tilde{\rho} - \rho)y_1 \\ &\qquad - 2(y_2 - x_2)(\tilde{\rho} - \rho)y_1 \\ &= \dot{L}_0|_{\tilde{\rho}=\rho}\end{aligned} \tag{11}$$

Trajectories of the coupled identical systems monotonically approach synchronization after some point in time, with $\dot{L}_0|_{\tilde{\rho}=\rho} < 0$. So, by Eq. 11 the same can be said of the trajectories of the non-identical systems, for which $\dot{L} < 0$. The reason is that the dynamical equation Eq. 6 is constructed so that changes in the time-derivative of L due to *explicit inclusion of parameter error in the Lyapunov function* will exactly cancel the changes due to the *implicit effect on the states through the dynamical equations*.

The above argument can be generalized to any pair of dynamical systems that are known to synchronize, for given coupling, when the parameters are identical. Consider two Lyapunov functions of state error and parameter error:

$$L_0(\mathbf{e}) := \mathbf{e}^2 = \sum_i e_i^2 \tag{12a}$$

$$L(\mathbf{e}, \mathbf{r}) := \mathbf{e}^2 + \mathbf{s}^2 = \sum_i e_i^2 + \sum_j s_j^2 \tag{12b}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{x}$, and $\mathbf{s} = \mathbf{q} - \mathbf{p}$. Because L_0 and L vanish only when all arguments vanish, if we can show either that $L_0 \to 0$ or that $L \to 0$, we have synchronization. If $L \to 0$, we also have parameter matching. We seek a dynamical equation for $\mathbf{q}$ such that there is a simple relationship between the Lyapunov function for the case of unequal parameters and that for the case of equal parameters. Since we already know $L_0|_{\mathbf{p}=\mathbf{q}} \to 0$, because the identical systems synchronize, we then have $L \to 0$ as well.

To find a suitable parameter update rule, $\dot{\mathbf{q}} = \mathbf{u}(\mathbf{x})$, we consider the time derivative of L:

$$\begin{aligned}\dot{L} &= \dot{L}_0 + 2\sum_j s_j \dot{s}_j \\ &= 2\sum_i e_i \dot{e}_i + 2\sum_j (q_j - p_j)\dot{q}_j\end{aligned} \tag{13}$$

recalling that $\dot{p}_i = 0$. We seek to decompose $\dot{e}_i$ in Eq. 13 as the sum of the value for a system with $\mathbf{q} = \mathbf{p}$ and a correction term due to the fact that $\mathbf{q} \neq \mathbf{p}$:

$$\dot{e}_i = \dot{e}_i|_{\mathbf{q}=\mathbf{p}} + \sum_j (q_j - p_j)\frac{\partial f_i(\mathbf{y};\mathbf{p})}{\partial p_j} \tag{14}$$

The partial derivative with respect to parameter p_j is the co-factor of that parameter in the ith dynamical equation, since the parameter only appears linearly. Inserting Eq. 14 in Eq. 13 yields:

$$\dot{L} = \dot{L}_0|_{\mathbf{q}=\mathbf{p}} + 2\sum_i e_i \sum_j (q_j - p_j)\frac{\partial f_i(\mathbf{y};\mathbf{p})}{\partial p_j} + 2\sum_j (q_j - p_j)\dot{q}_j \tag{15}$$

If we choose a parameter adaptation rule:

$$\dot{q}_j = -\sum_i e_i \frac{\partial f_i(\mathbf{y};\mathbf{p})}{\partial p_j} \tag{16}$$

then the last two terms in Eq. 15 cancel and we have

$$\dot{L} = \dot{L}_0|_{\mathbf{q}=\mathbf{p}} \tag{17}$$

which we claim is enough to give $L \to 0$, and hence synchronization, in the unequal-parameter case. That is because we already know that the Lyapunov function for the equal-parameter case is monotonically decreasing for some finite region of state space, i.e. $\dot{L}_0|_{\mathbf{q}=\mathbf{p}} \leq 0$ in this region, and $\dot{L}_0|_{\mathbf{q}=\mathbf{p}} = 0$ only if $\mathbf{x} = \mathbf{y}$. We make an additional assumption of "high-quality synchronization", needed for complete rigor but commonly observed: We assume that after some time there is no bursting away from the synchronization manifold, there is a finite distance from the manifold below which all points belong to the attractive region, so once a trajectory enters the region it cannot leave, since a decreasing L_0 implies decreasing distance from the manifold. Then $\dot{L} \leq 0$ in this region as well, implying $\mathbf{y}(t) \to \mathbf{x}(t)$, and $\mathbf{q}(t) \to \mathbf{p}$ as desired. (Strictly speaking, $\dot{L} \leq 0$ could imply that L converges to a positive constant value, and not to 0, but since $\dot{L} = \dot{L}_0|_{\mathbf{q}=\mathbf{p}} = 0$ only if $\mathbf{x} = \mathbf{y}$, the strict inequality holds except possibly on the synchronization subspace $\mathbf{x} = \mathbf{y}$, which is not dynamically invariant except when $\mathbf{q} = \mathbf{p}$.) We have thus given all the elements of a proof that if the two systems synchronize (in the "high-quality" sense) when parameters are identical, then both states and parameters will synchronize when the dynamical equations are augmented with Eq. 16.

The rule Eq. 16 can be generalized a bit: If we start with a more general Lyapunov function $L(\mathbf{e},\mathbf{s}) = \mathbf{e}^2 + \sum_j \frac{1}{\delta_j} s_j^2$, which is positive definite for arbitrary positive constants δ_j, we can derive a rule:

$$\dot{q}_j = -\delta_j \sum_i e_i \frac{\partial f_i(\mathbf{y};\mathbf{p})}{\partial p_j} \tag{18}$$

in place of Eq. 16. A still more general form of the parameter adaptation rule was proved in *Duane et al.* (2007).

4.2 Synchronization-based learning of inter-model connections

We apply the parameter adaptation rule Eq. 18 to the inter-model connections in a supermodel based on two imperfect SPEEDY atmospheres. The configuration is depicted in Fig. 4 with corresponding equations:

$$\dot{\mathbf{a}}_0 = \mathbf{f}^a(\mathbf{a}_0;\mathbf{p}_0^a) + \mathbf{g}^a(\mathbf{e}_0^h,\mathbf{e}_0^w,\mathbf{e}_0^m) \tag{19a}$$

$$\dot{\mathbf{a}}_1 = \mathbf{f}^a(\mathbf{a}_1;\mathbf{p}_1^a) + \mathbf{g}^a(\mathbf{e}_1^h,\mathbf{e}_1^w,\mathbf{e}_1^m) - \mathbf{C}_{12}[\mathbf{a}_1-\mathbf{a}_2] - \mathbf{K}[\mathbf{a}_1-\mathbf{a}_0] \tag{19b}$$

$$\dot{\mathbf{a}}_2 = \mathbf{f}^a(\mathbf{a}_2;\mathbf{p}_2^a) + \mathbf{g}^a(\mathbf{e}_2^h,\mathbf{e}_2^w,\mathbf{e}_2^m) - \mathbf{C}_{21}[\mathbf{a}_2-\mathbf{a}_1] - \mathbf{K}[\mathbf{a}_2-\mathbf{a}_0] \tag{19c}$$

$$\dot{\mathbf{o}} = \mathbf{f}^o(\mathbf{o};\mathbf{p}^o) + \mathbf{g}^o(\mathcal{P}^o\mathbf{e}_0^h,\mathcal{P}^o\mathbf{e}_0^w,\mathcal{P}^o\mathbf{e}_0^m,\mathcal{P}^o\mathbf{r}) \tag{19d}$$

$$\dot{\mathbf{l}} = \mathbf{f}^l(\mathbf{l};\mathbf{p}^l) + \mathbf{g}^l(\mathcal{P}^l\mathbf{e}_0^h,\mathcal{P}^l\mathbf{e}_0^w,\mathbf{r}) \tag{19e}$$

$$\mathbf{a}_s = \frac{1}{2}[\mathbf{a}_1+\mathbf{a}_2] \tag{19f}$$

$$\dot{\mathbf{C}}_{\mu\nu} = \mathbf{u}_{\mu\nu} \tag{19g}$$

where $\mathbf{u}_{\mu\nu}$ denote parameter update rules for the connection coefficient values of model μ nudged to model ν. The two imperfect atmosphere models ($\mathbf{a}_1$ and $\mathbf{a}_2$) are nudged to the truth ($\mathbf{a}_0$) with fixed nudging strength $\mathbf{K}$. Truth is represented by the SPEEDY model with standard parameter values ($\mathbf{a}_0$). The ocean and land model send their state information to all atmosphere models but receive the water, heat and momentum exchange from the true atmosphere only. The imperfect atmosphere models exchange state information and are nudged to each other state. Note that intermittent nudging of models to reality accomplishes the task of *data assimilation* in numerical weather prediction (*Duane et al.*, 2006; *Yang et al.*, 2006), so during training the supermodel effectively does continuous data assimilation. The parameter adaptation rule for a supermodel is obtained by forming the parameter vector $\mathbf{q}$ from the set of connection

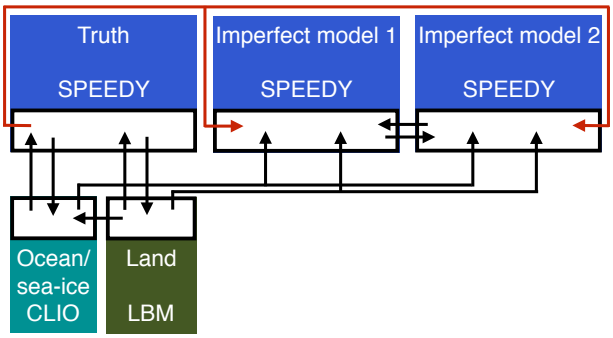


Figure 4: Schematic representation of the SPEEDO climate supermodel during synchronization-based training.

coefficients $C_{\mu\nu,ii}$, for μ and ν ranging over the labels of the separate models, and the index i over the dimension of the state vector (*Duane*, 2015). If one assumes that the truth is a supermodel for "correct" values of $\mathbf{C}_{\mu\nu}$, and that the supermodel will synchronize with this truth, then application of the rule Eq. 18 to the supermodel Eq. 19 using $e_i = a_{s,i} - a_{0,i}$ gives:

$$\dot{C}_{\mu\nu,ii} = \delta_{\mu\nu,i}[a_{\mu,i} - a_{\nu,i}][a_{s,i} - a_{0,i}] \tag{20}$$

where index i is the index of the state vector and the adaptation rates $\delta_{\mu\nu,i}$ are arbitrary constants. In principle, the adaptation rates can be chosen different for each element of the state vector in the inter-model connections, but in this study, we will choose a single adaptation rate for all elements and drop the subscripts from here on. The rule Eq. 20 has a simple interpretation: the time integral of the right-hand side gives the temporal covariance between truth-supermodel synchronization error, $(a_{s,i} - a_{0,i})$, and the inter-model nudging term, $(a_{\mu,i} - a_{\nu,i})$. It indeed makes sense to adapt the inter-model nudging strength, for a given pair of corresponding variables, depending on the sign and magnitude of this covariance. The connection coefficients cease updating when this covariance is zero and/or the synchronization error vanishes. In principle one could consider allowing different connection strengths for each state vector element. For SPEEDY this would imply adapting $N(N-1)$ times 31680 coefficients with N the number of imperfect models in the supermodel. In this study we will impose spatial invariance on the connection coefficients and only consider dependence on the physical variable that is being nudged in order to keep the number of adjustable coefficients relatively small. The nudging is applied to the velocity and temperature fields only, not to surface pressure and atmospheric humidity. It turns out that synchronization of the total state can be achieved by nudging these three fields only and an advantage is that it requires less communication between the atmospheres during the simulation. This choice for the nudging results in six connection coefficients for the inter-model connections between two imperfect models. The adaptation rules for the connection

coefficients become

$$\dot{C}^T_{\mu\nu} = \delta \sum_{k=1}^{8} \sum_{i=1}^{96} \sum_{j=1}^{48} [T_\nu(\lambda_i,\phi_j,p_k) - T_\mu(\lambda_i,\phi_j,p_k)] \times [T_0(\lambda_i,\phi_j,p_k) - T_s(\lambda_i,\phi_j,p_k)] \tag{21a}$$

$$\dot{C}^U_{\mu\nu} = \delta \sum_{k=1}^{8} \sum_{i=1}^{96} \sum_{j=1}^{48} [U_\nu(\lambda_i,\phi_j,p_k) - U_\mu(\lambda_i,\phi_j,p_k)] \times [U_0(\lambda_i,\phi_j,p_k) - U_s(\lambda_i,\phi_j,p_k)] \tag{21b}$$

$$\dot{C}^V_{\mu\nu} = \delta \sum_{k=1}^{8} \sum_{i=1}^{96} \sum_{j=1}^{48} [V_\nu(\lambda_i,\phi_j,p_k) - V_\mu(\lambda_i,\phi_j,p_k)] \times [V_0(\lambda_i,\phi_j,p_k) - V_s(\lambda_i,\phi_j,p_k)] \tag{21c}$$

where λ_i denotes the longitude, ϕ_j the latitude and p_k the pressure level. Integrating Eq. 21 to obtain finite-time changes $\Delta C_{\mu\nu}$, it is seen that the right-hand sides of the resulting equations correspond to the spatial and temporal covariance between the truth-supermodel synchronization error and the inter-model nudging terms. The adaptation rule in this case adjusts the connection coefficient of temperature, for example, between model μ and ν when the temperature difference between model μ and ν spatially and temporally covaries with the truth-supermodel synchronization error in temperature. This procedure makes sense because the inter-model temperature difference is proportional to the inter-model nudging term, and one wants to use more or less inter-model nudging, in a given direction, depending on whether the nudging tends to decrease or increase truth-supermodel synchronization error.

5 Results

5.1 Imperfect models

First, SPEEDO with standard parameter values (the "truth") was integrated for 400 years using present-day atmospheric CO_2 concentrations. The global mean surface temperature of this simulation rises from 13°C to about 14.2° during the first 50 years, remains fairly stable for about 300 years, and subsequently starts to cool during the final 50 years (Fig. 5). Slow cooling trends in the deep ocean are present during the whole simulation and in the end stabilize the ocean and reduce the mixing of heat from the deep ocean to the cold surface waters in the North Atlantic. Consequently, the North Atlantic surface waters cool, Arctic sea-ice expands, and the global mean surface temperature drops.

From this simulation, we selected January 1, 2001, in the middle of the relatively stable

period as the initial condition for the supermodel experiments. We integrated SPEEDO for 40 years for two sets of perturbed parameters (Table 1). The parameters concern parameterized descriptions of horizontal and vertical mixing processes due to unresolved turbulent motions. Imperfect model 1 (red line) warms around 1.4° with respect to the truth, and imperfect model 2 (blue line) cools around 0.5°. The amplitude of these climate differences are not unrealistic as differences of more than a degree in global mean temperature are not uncommon between state-of-the-art climate models. For reference, the perfect model was also integrated for 40 years (green line), referred to as the control simulation. It deviates from the truth due to the sensitive dependence on initial conditions as it was initiated from a perturbed first of January 2001 state.

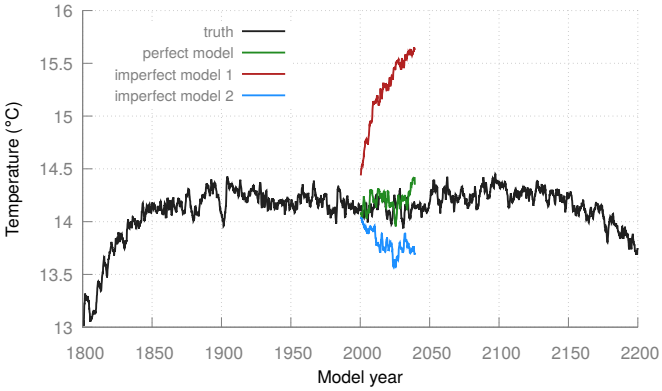


Figure 5: Global mean temperature time-series for SPEEDO with standard parameter values (truth). The perfect model and the two imperfect models were initiated in year 2001 and integrated for 40 years.

	perfect	imperfect 1	imperfect 2
relaxation timescale of convection	6 hours	4 hours	8 hours
relative humidity threshold	0.9	0.85	0.95
momentum diffusion timescale	24 hours	18 hours	30 hours

Table 1: Parameter values in perfect and imperfect models.

5.2 Synchronization

Before we start to connect the two imperfect models and train the connection coefficients, we first have a look at the synchronization errors when the models are nudged to the truth.

During training, synchronization errors determine the updates to the connection coefficients. For effective training, synchronization errors should be significantly larger

for the imperfect models than for the perfect model for given model-to-truth nudging strength **K**. Ideally, the trained supermodel will have synchronization errors close to the perfect model. We investigated the magnitude of the synchronization error in relation to the nudging strength **K** by running the configuration depicted in Fig. 6. The perfect and imperfect models receive the state information from the truth at every time-step and their states are nudged accordingly. Only the truth exchanges water, heat and momentum with the surface (ocean and land) models. The other models receive the state information from the surface models, calculate each their own water, heat and momentum exchange, but this information is not communicated back to the surface models.
The models are initialized from randomly perturbed January 1, 2001 states and integrated for two weeks with **K** set to zero in order to allow the models to de-synchronize. Next **K** is set to a value of 1/24 hours^{-1}, and the integration is continued for the rest of the year. Sensitive dependence on initial conditions and model errors cause a rapid increase of the synchronization error during the first two weeks (Fig. 7a). The error reduces rapidly when **K** is set to 1/24 hours^{-1} after two weeks, levels off within a couple of weeks and remains fairly stable for the remainder of the year. The perfect model does not synchronize perfectly with the truth, but the average error is only 0.01°C. Given daily fluctuations at given locations of tens of degrees, this is a very small synchronization error. The synchronization error is about 10 times larger for imperfect model 1 and 6 times for imperfect model 2. The synchronization error is almost independent of nudging timescale between 24 and 4 h for the perfect model, but for the imperfect models, the errors are reduced by more than 60% over this range (Fig. 7b). Note that nudging only part of the total state vector (the surface pressure field and the humidity field are not nudged) is sufficient to achieve this high degree of synchronization.
Based on these synchronization experiments, we choose a nudging timescale of 24 h. The nudging keeps the imperfect models close to the truth, but at the same time, there is room for a ten-fold reduction in synchronization error by updating the inter-model connection coefficients.

5.3 Learning

In the first learning experiment, the SPEEDO supermodel was initialized at January 1, 2001 in the configuration depicted in Fig. 4 using initial connection coefficient values in Eq. 21 equal to $1/8$ h^{-1}. The nudging timescale corresponding to **K** in Eq. 19 was set to 24 hours as motivated in the previous section. The rate of learning δ in the update rules of Eq. 21 was set to 24000. With these parameter choices, the SPEEDO supermodel was trained for 10 years by integrating Eq. 19 with update rules $\mathbf{u}_{\mu\nu}$ given by

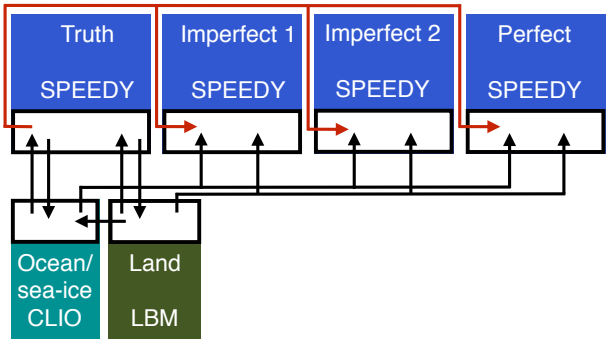


Figure 6: Schematic representation of the SPEEDO configuration for the synchronization experiments.

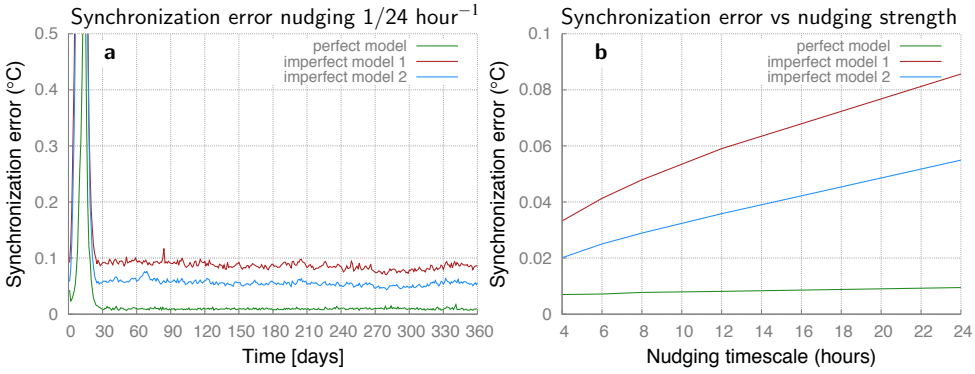


Figure 7: (a) Synchronization error as measured by the root of the globally averaged mean squared surface air temperature difference between model and truth for a nudging timescale of 24 h for the perfect model (green), imperfect model 1 (red), and imperfect model 2 (blue). (b) Time-averaged synchronization error during the final 10 months of the integrations as a function of nudging timescale.

Eq. 21. The six connection coefficients ($C^T_{\mu\nu}$, $C^U_{\mu\nu}$, $C^V_{\mu\nu}$ in Fig. 8ab) converge within the first months to values that remain fairly stable during the remaining 10 years of the training. A weak annual cycle is visible, suggesting that the optimal nudging coefficients have a weak seasonal dependence. It is obvious from the graph that C_{12} and C_{21} lie symmetrical around the initial value for each of the three variables. This is due to the fact that according to the update rules Eq. 21 $\dot{C}_{\mu\nu} = -\dot{C}_{\nu\mu}$. Another consequence is that the asymptotic values depend on the initial value.

In the second learning experiment, the SPEEDO supermodel was initialized with connection strengths equal to $1/24$ h^{-1}. Indeed the asymptotic values are different in this case (Fig. 8cd). In addition the symmetry is broken due to additional constraints that the connection values are not allowed to go below zero or above $1/4$ hours^{-1}. An upper bound is imposed in order to prevent numerical instabilities for too strong nudging and allow some de-synchronization at times when the truth is hard to follow using either imperfect models. The numerical instabilities could be prevented by reducing the time-step but we chose to keep the time-step fixed at 30 minutes and impose an upper bound on the nudging strength.

The learning experiments result in different supermodel solutions as defined by the asymptotic connection values. These are summarized in Table 2. The synchronization error when connected to the truth is similar for both supermodels (Fig. 9), despite the difference in the connection values. The connections of supermodel 1 are about twice as strong as the connections of supermodel 2. The relative strengths between the connections are therefore about the same. This suggests that only the relative strengths matter and that the supermodel solution is invariant under multiplication of all connections by a constant factor provided that the imperfect models synchronize. The initial value of the coefficients during learning selects a particular factor.

It appears that short training periods suffice to train the inter-model connection coefficients, despite the long time-scales present in the climate. During the learning phase, the imperfect atmosphere models receive the true ocean, land, and sea-ice states and are able to learn how the true atmosphere interacts with these states. The imperfections concern fast atmospheric processes (turbulence and convection) and can thus be trained on these time scales. In order to verify that the training of the inter-model connection coefficients does not depend on the ocean/sea-ice state, we repeated the learning experiments starting in year 2151 of the reference simulation. The state of the ocean and sea-ice is different and the thermo-haline circulation in the North-Atlantic basin is about to collapse, causing a rapid cooling of the sea surface temperatures in the North-Atlantic and a growth of the Arctic sea-ice cover. Nevertheless, the training converges on similar connection values (not shown).

The trained supermodels have smaller synchronization errors as compared to both imperfect models (Fig. 9) but not as small as the perfect model. For comparison, we evaluated the synchronization errors of the untrained supermodel with equal weights of $1/24$ h^{-1}. The training has reduced the synchronization error in the east-west component of the wind at 850 hPa by only a small margin (Fig. 9).

It is hoped that the reduction in synchronization error when the supermodel is nudged to truth will be reflected in improved simulations of climate when the supermodel is run freely to simulate climate. If indeed only the relative strengths of the connections matter, then the two supermodel solutions should give similar results. In the next sections we will investigate climate simulations of both supermodel solutions in comparison to the perfect and imperfect models and to the untrained supermodel.

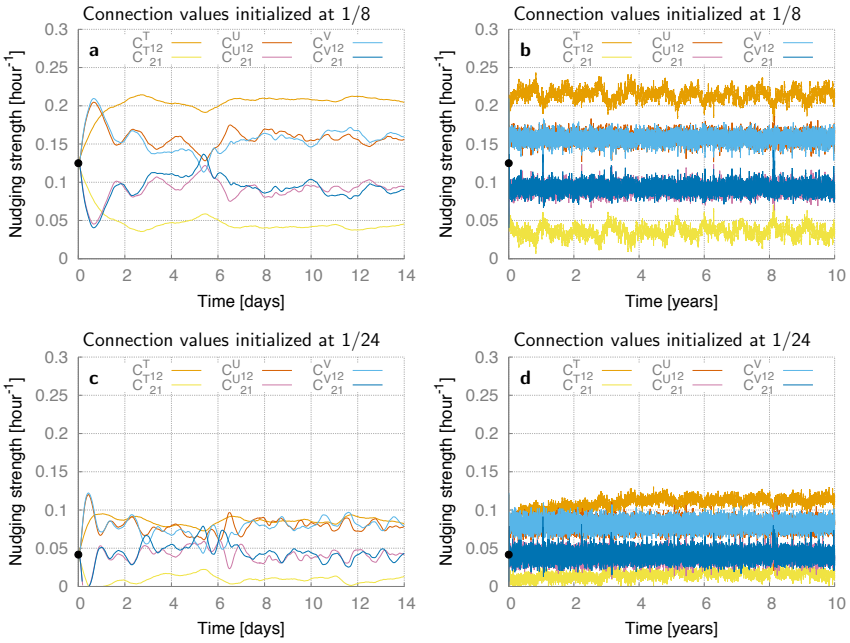


Figure 8: Time-series of the connection coefficients during the training when initialized at $1/8$ h^{-1} (a and b) and $1/24$ h^{-1} (c and d). The first 14 days of the training are plotted in the left panels, right panels display the whole 10 year training period. The black dot denotes the initial values.

	C^T_{12}	C^T_{21}	C^U_{12}	C^U_{21}	C^V_{12}	C^V_{21}
supermodel 1	0.22	0.03	0.17	0.08	0.17	0.08
supermodel 2	0.12	0.01	0.08	0.04	0.08	0.04

Table 2: Connection coefficient values of the two supermodel solutions.

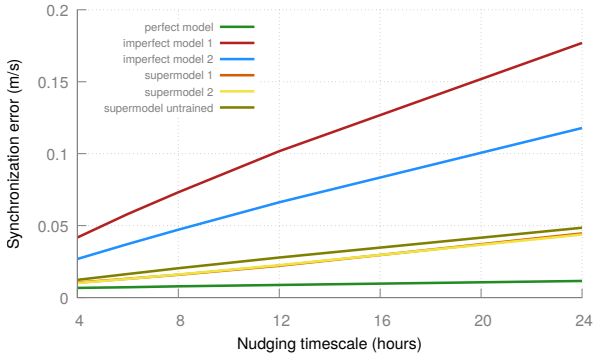


Figure 9: The globally averaged synchronization error in the east-west component of the wind at 850 hPa for different models when connected to the truth for different nudging strengths from similar synchronization experiments as in Fig. 7b.

5.4 Climate

The two trained supermodels and the untrained supermodel are initialized at January 1, 2001 and integrated for 40 years. For both trained supermodels, the evolution of the global mean temperature shows no sign of a drift with respect to the truth or the perfect model, unlike the imperfect models and the untrained supermodel (Fig. 10). The synchronized evolution of the imperfect models in the untrained supermodel produces a global mean temperature close to the average of the global mean temperature of the two imperfect model evolutions.

In the trained supermodels, imperfect model 1 is more strongly nudged to imperfect model 2 for all connections than vice versa (Table 2). In this case Eq. 3 implies that the supermodel solution is systematically pushed away from the averaged solution toward the evolution of imperfect model 2. This makes sense as the warming of imperfect model 1 with respect to the truth is stronger than the cooling of imperfect model 2 (Fig. 10).

Another measure of the quality of the climate simulations are errors in the climatological mean fields. As an example, Fig. 11 shows the error in the time average over the final 30 years of the simulations of the east-west component of the wind at the 200 hPa pressure level (at about 10 km height). The mean winds of imperfect model 1 have errors that reach 5 m/s in the Southern Hemisphere. The error pattern has a rich spatial structure but is to a considerable extent opposite in sign as compared to imperfect model 2. An improved estimate of the true mean winds is obtained by taking the average of both models, commonly referred to as the multi-model average in climate science. The error pattern of the multi-model average is very similar to the

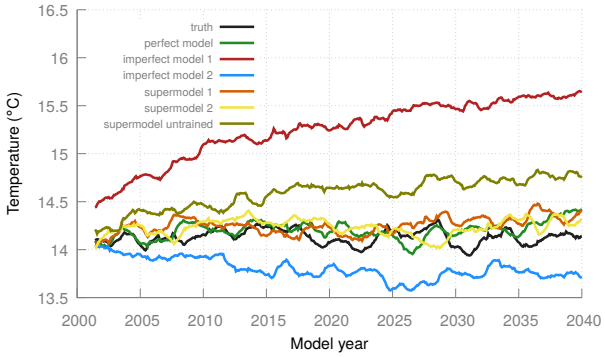


Figure 10: The global mean surface temperature evolution for the truth, the perfect model, the two imperfect models, the two trained supermodels, and the untrained supermodel. All models are initialised at January 1, 2001, of the truth.

error pattern of the untrained supermodel with equal coefficients. Both trained supermodels have smaller errors than the untrained supermodel. The training based on synchronization errors, essentially a training based on short-term prediction errors, has also proved useful for supermodel simulation of long-term climate. For reference, the bottom panel of Fig. 11 shows the errors of the perfect model. Due to sampling uncertainties, the mean state of the truth is not exactly reproduced in a 30 year simulation with the perfect model. The errors in the mean state of the trained supermodels are larger as compared to those of the perfect model, especially in the tropical region. An optimally weighted multi-model average was determined and is also shown in Fig. 11. Adding 0.24 times the mean state of imperfect model 1 and 0.76 times the mean state of imperfect model 2 leads to a similar estimate of the true mean state as provided by either trained supermodels.

Globally averaged errors for a number of different climatological fields paint the same picture (Fig. 12). The untrained supermodel has similar errors to the multi-model average. An optimally weighted multi-model average has smaller errors, comparable to those of the trained supermodels, in the nudged variables and also in variables that are not nudged, like precipitation and cloud-cover. There is still room for some improvement to match the small errors (due to sampling) obtained with the perfect model simulation.

Ideally one should compare attractors in order to make judgements about the quality of the climate simulations instead of just comparing the mean states. For such high-dimensional systems as SPEEDO, the evaluation of the probability density in state space is computationally too expensive since too much data is required in order to

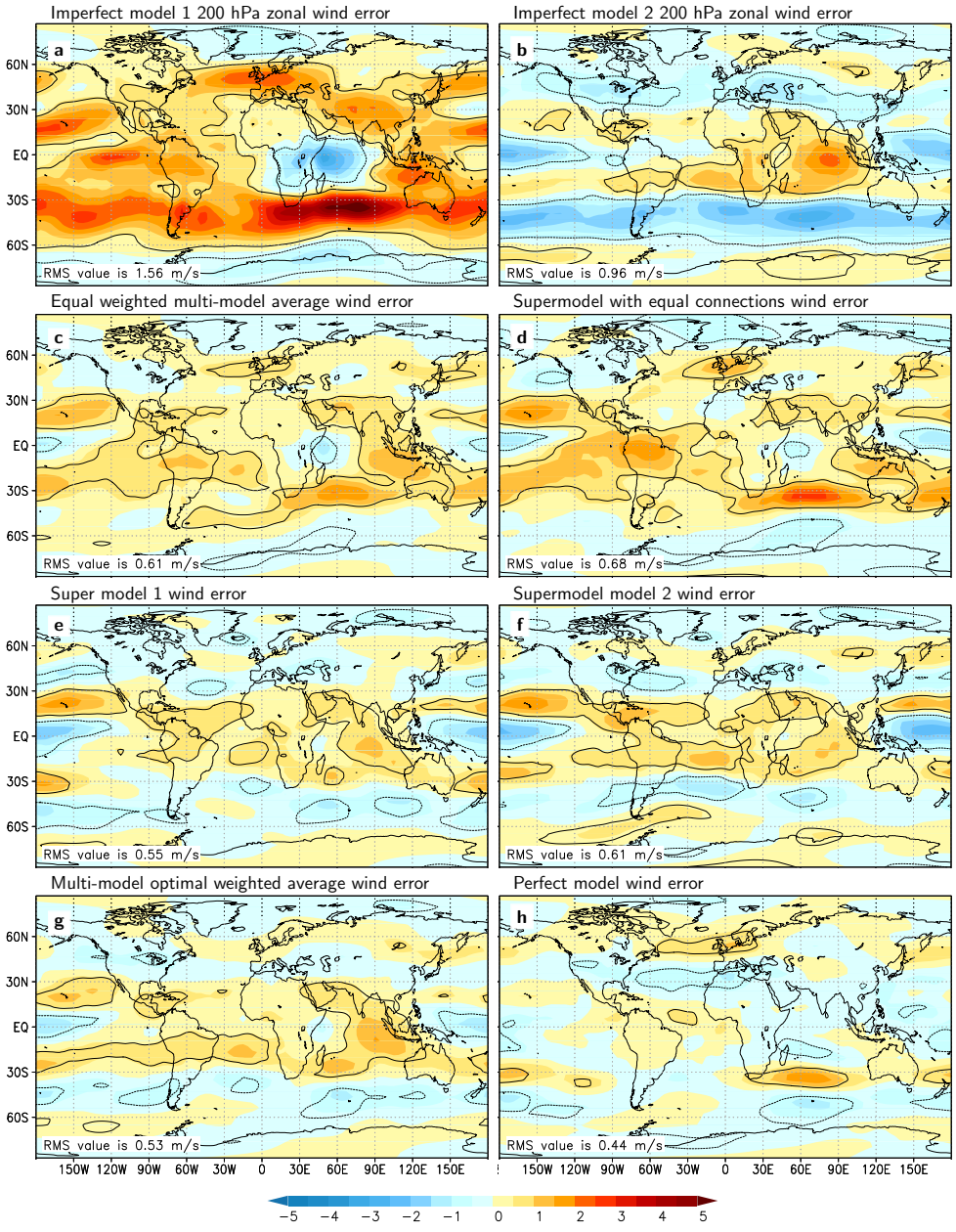


Figure 11: Difference in the east-west component of the wind at the 200 hPa pressure level averaged over model years 2011-2040 for the various models with respect to the truth. Contours denote the areas where the difference is larger than the sampling error at 95% confidence (solid for positive difference and dotted for negative). Positive values imply stronger mean winds blowing eastward. Units: m/s.

yield representative results (*Dool*, 2011), but other statistical properties of the attractors might be compared. As an example, we evaluated the 95% percentile of three-hourly sums of convective rainfall at each location. The results are plotted in Fig. 13. The convective rainfall extremes are largest in tropical areas and in the regions of the extra-tropical storm tracks (panel a). In general, imperfect model 1 overestimates and imperfect model 2 underestimates the convective precipitation extremes (panel c and d). Supermodel 1 simulates the precipitation extremes more accurately. The root mean squared error is 1.2 mm/day as compared to 2.7 and 2 mm/day for imperfect model 1 and 2, respectively.

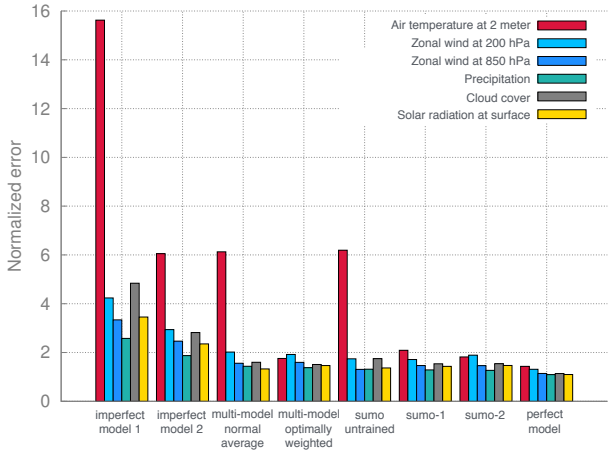


Figure 12: Normalized errors with respect to the truth in various meteorological fields averaged over model years 2011-2040 for the various models. At each location, the difference in the mean state is normalized by 1.96 times the standard deviation of the 30 yearly values, divided by the square root of 30. The normalized errors in this graph correspond to the root of the globally averaged squared normalized differences. With this normalization, the perfect model has errors around value one.

5.5 Climate response

Climate models are commonly used to make projections of the future climate by assuming scenarios for future emissions of greenhouse gasses. Here, we explore whether the trained supermodel is capable of simulating the climate change due to a doubling of the CO_2 concentration. In model year 2041, the CO_2 concentration is doubled and the various models are integrated for 30 years. Global mean temperature time-series are plotted in Fig. 14. The global mean temperature in the supermodel remains close to the truth after the doubling of CO_2. Imperfect model 1 and 2 simulate a similar warming, but the reference state at the onset of doubling is warmer in imperfect model 1 and colder in imperfect model 2. In response to the CO_2 change, also the atmospheric

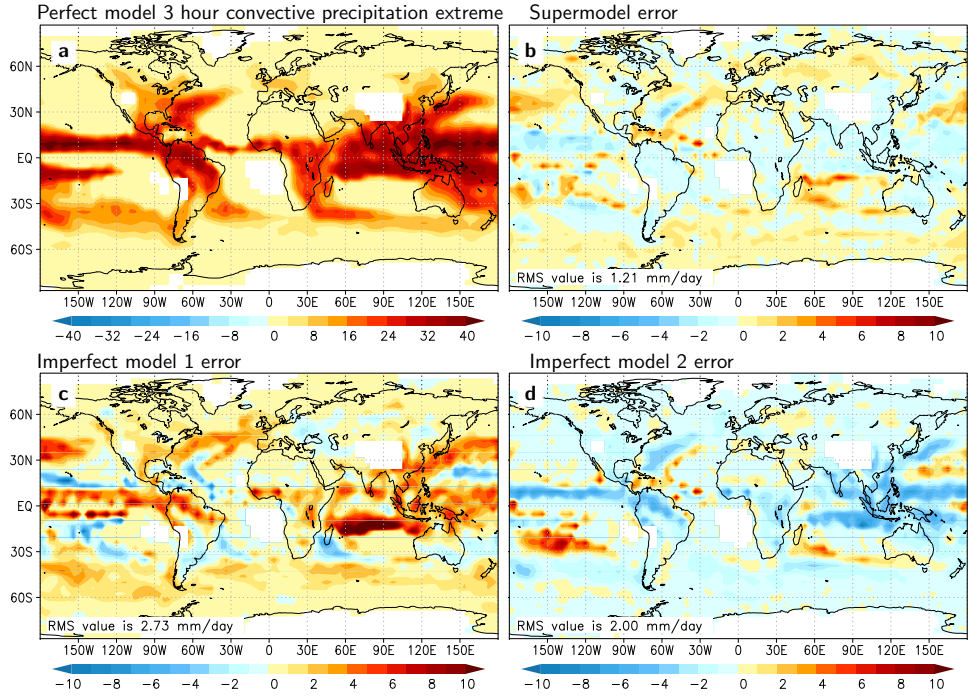


Figure 13: 95% percentile of convective precipitation three-hourly sums (mm/day) in the perfect model (a), and differences with respect to the perfect model for supermodel 1 (b) and both imperfect models (c and d). Calculations are performed for the years 2011-2040.

circulation changes (Fig. 15). The change of the east-west component of the wind at 850 hPa is best simulated by the supermodel, especially in the tropical regions around Indonesia.

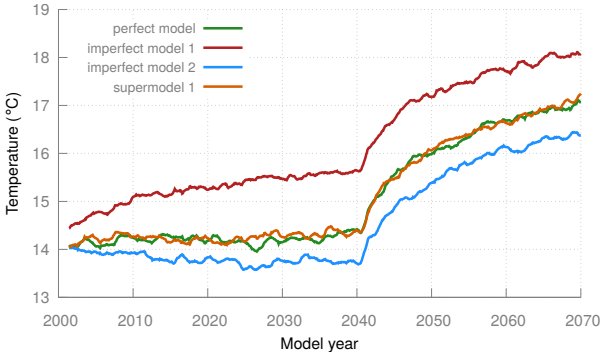


Figure 14: Global mean temperature time-series for the various models. In model year 2041 the CO_2 concentration is doubled and in response the climate starts to warm.

6 Discussion: Supermodeling versus *A Posteriori* Combinations of Model Results

When compared to the optimally weighted multi-model mean, the supermodel yields similar accuracy. However, in addition the supermodel produces actual trajectories that are closer to the true trajectories. Temperature time-series of imperfect model 1 are systematically too warm in most geographical locations and too cold in imperfect model 2. These time-series cannot be averaged to more accurately represent true time series since the time-series are not synchronized and their mean is not a solution of the dynamical equations. In the supermodel the models converge on a synchronized time-series with a more accurate mean temperature. For climate impact studies this is a great advantage since it eliminates the need for bias corrections (correction of time-series in order to remove the error in the mean).

One is interested not only in the mean behavior of the models and supermodel but also in internal variability. The interesting properties of the various attractors are usually captured in probability density functions (pdfs). There is significant ambiguity in methods to combine pdfs of different climate models as an improved estimate of the true pdf. Suppose, as a thought experiment, that one has two different models of the same system, each of which exhibits Gaussian statistics in some variable, but with different means and different variances. That is, suppose that the pdf of some state

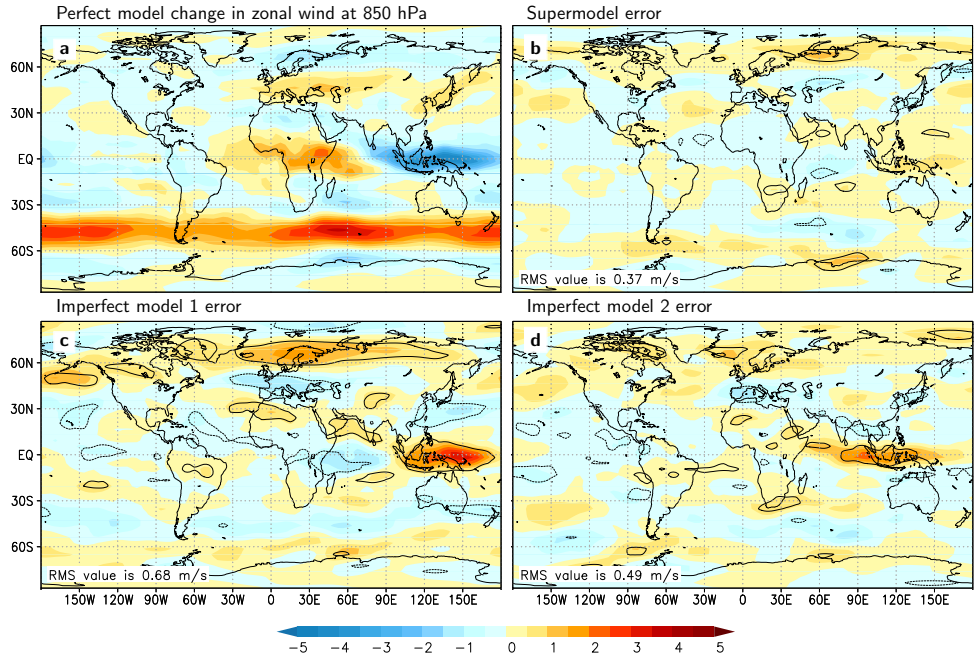


Figure 15: Change in the east-west component of the wind at 850 hPa due to a CO_2 doubling in the perfect model (a) and the error in the simulated wind change for supermodel 1 (b) and for the two imperfect models (c and d). The change is calculated by subtracting the average wind before CO_2 doubling during model years 2016-2040 from the average wind during 2056-2070. The contours indicate regions where the difference is statistically significant at the 95% confidence level. The root of the global mean squared error is plotted in the lower left corner.

variable x is given by pdfs P_1, P_2 in the two models:

$$P_1(x) = N_1 \exp \frac{(x-\mu_1)^2}{2\sigma_1^2} \tag{22a}$$

$$P_2(x) = N_2 \exp \frac{(x-\mu_2)^2}{2\sigma_2^2} \tag{22b}$$

where $N_{1,2}$ are normalization factors. If the means are not greatly separated $|\mu_1 - \mu_2| < \sigma_1, \sigma_2$ and the shapes are similar, $\sigma_1 \approx \sigma_2$, we might guess that the difference is due to some systematic error and that the true distribution is a Gaussian with the average mean:

$$P_m(x) = N_m \exp \frac{(x-\mu_m)^2}{2\sigma^2} \tag{23}$$

with $\mu_m = (\mu_1 + \mu_2)/2$, and $\sigma \approx \sigma_1 \approx \sigma_2$. If we were instead to blindly average the pdfs, we would in general have a non-gaussian distribution with $\sigma > \sigma_1, \sigma_2$, incorrectly inflating the variance. So it might seem that a recipe for intelligently combining pdfs is accessible.

But what if the true distribution were bimodal, with the means of the two modes more widely separated than in the above case of small systematic error, and what if each model, for reasons of its own dynamical imperfections, is biased in favor of one mode? Then, a simple average of pdfs, which would capture the bimodality, would be preferable to the "intelligent" combination described. Without prior information about the form of the true distribution, no general prescription for combining the pdf's is possible. It might not even be possible at all to combine the pdf's of different models into a pdf that reflects the true model well. In *van den Berge et al.* (2011) for example, two periodic attractors and one stable fixed point attractor were used to construct a chaotic supermodel. The supermodel has similar statistics compared to the true chaotic system, while a combination of the pdf's of the non-chaotic systems cannot give a good approximation of the true chaotic pdf. The extent to which such extreme behavior occurs in real climate models is an open question, but the construction of an actual dynamical system is arguably the soundest way to represent the true physical variations within the modelled climate.

The choice of models in a supermodel configuration is also more flexible than in multi-model averaging. In cases where the constituent models are all on one side of the truth, the introduction of an additional imperfect model into the supermodel that is on the opposite side can help to improve the supermodel (*Schevenhoven and Selten*, 2017). But unlike the situation with multi-model averaging, the new imperfect model on the opposite side of the truth, included in the supermodel with significant weight, can be a model with a highly unrealistic long-term climate, such as one with a fixed-point attractor, when run in isolation.

Finally, it should be mentioned that there is one advantage of multi-model runs that naively appears to be lost in supermodeling: the ability to extract spread information from an ensemble of models, as an indicator of model error. But one can easily consider an ensemble of supermodels (*Duane et al.*, 2017), defined by variations in the connection coefficients determined from fluctuations during the learning process or from different learning strategies. The ensemble spread enables to gauge uncertainties due to model errors and due to imperfect knowledge of the initial state. The supermodels have been trained to follow the truth more closely in the neighbourhood of observed trajectories. Therefore, they can be expected to better reproduce the local lyapunov exponents and better represent the divergence of trajectories in medium-range forecasts due to uncertainty in the initial condition.

7 Summary and Concluding Remarks

This is the first time that synchronization-based supermodeling is applied to complex global climate models and its potential to improve climate simulations is demonstrated. The SPEEDO climate model with standard parameter values is regarded as truth; two imperfect models are constructed by perturbing three parameters. Due to the different parameter setting, one model warms with respect to the truth, the other cools. A supermodel is constructed by connecting the temperature and velocity fields of both models through nudging terms, and the supermodel synchronizes on a common solution. Imposing spatial invariance while allowing different connection strengths for the different meteorological fields yields six adjustable connection coefficients. Using the fact that synchronization of states between two connected systems can be extended to synchronization of parameters, when these vary between the two systems, the inter-model connection coefficients within the supermodel are dynamically adjusted, along with the states, so that the supermodel synchronizes with truth, from which it continuously receives data in the learning phase. After a quick adjustment during the first couple of weeks of the training, the coefficients exhibit only small fluctuations around a stable long-term mean value during the ten year training period. These stable, long-term mean values of the coefficients define the supermodel. During a 40 year climate simulation, the supermodel preserves the correct global mean temperature. Moreover, the globally averaged errors in all 30 year mean meteorological fields examined are smaller than the errors in the imperfect models. In addition, the supermodel is able to reproduce the correct warming in response to a doubling of the CO_2 concentration.
The synchronization-based learning algorithm reaches locally optimal values of the connection coefficients. There appears to be a degeneracy, as explained in Section 5.3

in that, depending on initial coefficient values, the algorithm picks from a family of equally good coefficient values - we conjecture that what matters is not the absolute value of the connection coefficients, but their relative strengths. On the other hand, a more refined connection scheme might yield even better results. During learning, the connection values exhibit a weak dependence on the seasonal cycle, for instance, suggesting that a seasonal dependence of the connection strengths might further improve the supermodel climate simulations.

There is no guarantee that the learning algorithm converges on the globally optimal connection coefficients. Other learning approaches based on matching finite segments of the trajectories instead of just the instantaneous states, as in *van den Berge et al.* (2011), or minimization of errors in climate statistics like the mean or the variance over multi-year long trajectories by iterative methods as in *Severijns and Hazeleger* (2005); *Wiegerinck and Selten* (2017), might yield even better supermodel solutions.

In state-of-the-art weather forecasts, models are initialised from observed states that are not on the model attractor. During the forecasts, the trajectories systematically transition to the model attractor, and in a couple of weeks most of the long-term climate errors have developed (*Jung et al.*, 2012). We therefore expect that the training based on short-term prediction errors in this study could also be successfully applied to state-of-the-art weather and climate models. We restricted the evaluation of the supermodel to climate timescales, but we expect that short-term prediction errors are improved with respect to the imperfect models since the learning is based on one time step predictions of the truth. Weather prediction with supermodels remains to be evaluated.

In the present study, only imperfections in the atmospheric component have been considered. During the learning phase, the imperfect atmosphere models receive the true ocean, land, and sea-ice states and can learn how the true atmosphere interacts with these states. The imperfections concern fast atmospheric processes (turbulence and convection) and can thus be trained on these time scales. It remains to be seen how imperfections in the land- and ocean models affect the learning. In principle, the supermodel approach can be extended to include multiple imperfect ocean- and land models with inter-model connections that can be included in the learning.

The magnitude of the synchronization errors between supermodel and truth is only slightly reduced when the learning process is initialized with uniform connections. Nevertheless, the climate properties of the supermodel with the learned coefficients are much better than those of the supermodel with the initial connection coefficients. It seems that even a small reduction of synchronization error in the training phase is heuristically useful for correcting the dynamics of the model, but more work is needed to assess the universality of this behavior.

Nonnegative connection coefficients are imposed during the learning in order to induce synchronization of the imperfect models in the supermodel. One could allow negative coefficients during the learning. In that case one model tries to flee from the other model, but at the same time the other model chases that model at a faster pace and the imperfect model trajectories can still remain approximately in synch. Other regions of phase space can be explored by the supermodel by allowing negative connections, which might lead to an improved supermodel solution.

Although the supermodel is not trained to be able to simulate the correct response of the climate to a perturbation like the doubling of the atmospheric CO_2 concentration, we find this to be the case in this study. The full extent of the robustness of the supermodel solution against variations in ancillary parameters remains to be determined.

In the perfect model approach of the present study, the truth is inside the model class of imperfect models. It remains to be seen how well the supermodel scores when the truth is outside of the imperfect model class, a situation that arises when climate models are used to simulate a much more complex reality (*Wiegerinck and Selten*, 2017).

In applying the supermodel approach to an ensemble of different, state-of-the-art climate models, it must be noted that the different models employ different numerical representations of the various meteorological fields, and especially that they are formulated on different numerical grids. SPEEDO models have been shown to synchronize even when only some meteorological fields are connected, or only some spatial scales. One solution to the problem of different numerical grids would be to transform the grid representation to a spectral representation and do the exchange of state information and nudging in spectral space only for wave-numbers that are well resolved in all participating models in the supermodel, and then transform the nudging tendencies back to the respective grids. Synchronization of the constituent models might be expected, despite the different grid representations.

In the present study we have assumed perfect knowledge of the truth. In reality observations of the truth are incomplete and noisy. The influence of noisy and incomplete observations on the learning of the supermodel remains to be investigated, but we are encouraged by the success of data assimilation for weather prediction under the same circumstances.

We have applied the supermodel approach in the context of simulating the Earths climate, but its application domain is much wider. In any modeling context where different models exist of a complex, real system, like ecological systems or economical systems, where data assimilation from the real systems yields truth-model synchronization, and where enough good quality observational data are present, the supermodel approach potentially leads to more accurate predictions.

Acknowledgments

We wish to acknowledge the support of Camiel Severijns and Paul Hiemstra in coding the SPEEDO supermodel.
This work was partly funded by STERCP (ERC project No. 648982) and Marie Sklodowska-Curie Actions–Individual Fellowship (MSCA-IF, Grant No. 658602).

References

Besançon, G., J. D. León-Morales, and O. Huerta-Guevara (2006), On adaptive observers for state affine systems, *International Journal of Control*, *79*(6), 581–591, doi:10.1080/00207170600552766. 4.1

Dool, H. V. D. (2011), Searching for analogues, how long must we wait?, *Tellus A*, *46*(3). 5.4

Duane, G., W. Wiegerinck, F. Selten, M.-L. Shen, and N. Keenlyside (2017), Supermodeling: Synchronization of alternative dynamical models of a single objective process, *Chapter in Advances in Nonlinear Geosciences, ed: A.A. Tsonis*, doi: 10.1007/978-3-319-58895-7. 6

Duane, G. S. (2015), Synchronicity from synchronized chaos, *Entropy*, *17*(4), 1701–1733. 1, 4.2

Duane, G. S., J. J. Tribbia, and J. B. Weiss (2006), Synchronicity in predictive modelling: a new view of data assimilation, *Nonlinear Processes in Geophysics*, *13*(6), 601–612, doi:10.5194/npg-13-601-2006. 4.2

Duane, G. S., D. Yu, and L. Kocarev (2007), Identical synchronization, with translation invariance, implies parameter estimation, *Physics Letters A*, *371*(5–6), 416 – 420, doi:http://dx.doi.org/10.1016/j.physleta.2007.06.059. 4.1

Goosse, H., and T. Fichefet (1999), Importance of ice-ocean interactions for the global ocean circulation: A model study, *Journal of Geophysical Research: Oceans*, *104*(C10), 23,337–23,355, doi:10.1029/1999JC900215. 2

Hazeleger, W., B. J. J. M. van den Hurk, E. Min, G. J. van Oldenborgh, A. C. Petersen, D. A. Stainforth, E. Vasileiadou, and L. A. Smith (2015), Tales of future weather, *Nature Climate Change*, *5*(2), 107–113. 1

Hiemstra, P. H., N. Fujiwara, F. M. Selten, and J. Kurths (2012), Complete synchronization of chaotic atmospheric models by connecting only a subset of state space, *Nonlinear Processes in Geophysics*, *19*(6), 611–621, doi:10.5194/npg-19-611-2012. 1

Jung, T., M. J. Miller, T. N. Palmer, P. Towers, N. Wedi, D. Achuthavarier, J. M. Adams, E. L. Altshuler, B. A. Cash, J. L. K. III, L. Marx, C. Stan, and K. I. Hodges (2012), High-resolution global climate simulations with the ecmwf model in project athena:

Experimental design, model climate, and seasonal forecast skill, *Journal of Climate*, *25*(9), 3155–3172, doi:10.1175/JCLI-D-11-00265.1. 7

Kirtman, B. P., D. Min, P. S. Schopf, and E. K. Schneider (2003), A new approach for coupled gcm sensitivity studies, *Cola technical report 154 (cola.gmu.edu/pub/ctr/ctr_154.pdf)*, p. 48 pp. 1

Lunkeit, F. (2001), Synchronization experiments with an atmospheric global circulation model, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *11*(1), 47–51, doi:10.1063/1.1338127. 1

Mirchev, M., G. S. Duane, W. K. Tang, and L. Kocarev (2012), Improved modeling by coupling imperfect models, *Communications in Nonlinear Science and Numerical Simulation*, *17*(7), 2741 – 2751, doi:http://dx.doi.org/10.1016/j.cnsns.2011.11.003. 1

Molteni, F. (2003), Atmospheric simulations using a gcm with simplified physical parametrizations. i: model climatology and variability in multi-decadal experiments, *Climate Dynamics*, *20*(2), 175–191, doi:10.1007/s00382-002-0268-2. 2

Pecora, L. M., and T. L. Carroll (1990), Synchronization in chaotic systems, *Phys. Rev. Lett.*, *64*, 821–824, doi:10.1103/PhysRevLett.64.821. 4.1

Schevenhoven, F. J., and F. M. Selten (2017), An efficient training scheme for supermodels, *Earth System Dynamics*, *8*(2), 429–438, doi:10.5194/esd-8-429-2017. 6

Severijns, C. A., and W. Hazeleger (2005), Optimizing parameters in an atmospheric general circulation model, *Journal of Climate*, *18*(17), 3527–3535, doi:10.1175/JCLI3430.1. 7

Severijns, C. A., and W. Hazeleger (2010), The efficient global primitive equation climate model speedo v2.0, *Geoscientific Model Development*, *3*(1), 105–122, doi:10.5194/gmd-3-105-2010. 1, 2

Shen, M.-L., N. Keenlyside, F. Selten, W. Wiegerinck, and G. S. Duane (2016), Dynamically combining climate models to ?supermodel? the tropical pacific, *Geophysical Research Letters*, *43*(1), 359–366, doi:10.1002/2015GL066562. 1

van den Berge, L. A., F. M. Selten, W. Wiegerinck, and G. S. Duane (2011), A multimodel ensemble method that combines imperfect models through learning, *Earth System Dynamics*, *2*(1), 161–177, doi:10.5194/esd-2-161-2011. 1, 6, 7

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Quarterly Journal of the Royal Meteorological Society*, *134*(630), 241–260, doi:10.1002/qj.210. 1

Wiegerinck, W., and F. Selten (2017), Attractor learning in synchronized chaotic systems in the presence of unresolved scales, *Chaos*, *ibid*. 3, 7

Yang, S., D. Baker, H. Li, K. Cordes, M. Huff, G. Nagpal, E. Okereke, J. Villafañe, E. Kalnay, and G. Duane (2006), Data assimilation as synchronization of truth and model: Experiments with the three-variable lorenz system, *Journal of the Atmospheric Sciences*, *63*(9), 2340–2354, doi:10.1175/JAS3739.1. 4.1, 4.2

Zdunkowski, W., and A. Bott (2003), *Dynamics of the Atmosphere: A Course in Theoretical Meteorology*, Cambridge University Press, doi:10.1017/CBO9780511805462. 2

Zhang, Q. (2002), Adaptive observer for multiple-input-multiple-output (mimo) linear time-varying systems, *IEEE Transactions on Automatic Control*, *47*(3), 525–529, doi:10.1109/9.989154. 4.1

Paper III

Improving weather and climate predictions by training of supermodels

Francine Schevenhoven, Frank Selten, Alberto Carrassi and Noel Keenlyside

Earth System Dynamics, **10**, 789-807 (2019)

III

Improving weather and climate predictions by training of supermodels

Francine Schevenhoven[1,2], Frank Selten[3], Alberto Carrassi[4,1,2,5], and Noel Keenlyside[1,2]

[1]Geophysical Institute, University of Bergen, Bergen, Norway
[2]Bjerknes Centre for Climate Research, Bergen, Norway
[3]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[4]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[5]Mathematical Institute, Utrecht University, Utrecht, the Netherlands

Abstract

Recent studies demonstrate that weather and climate predictions potentially improve by dynamically combining different models into a so-called "supermodel". Here, we focus on the weighted supermodel - the supermodel's time derivative is a weighted superposition of the time derivatives of the imperfect models, referred to as weighted supermodeling. A crucial step is to train the weights of the supermodel on the basis of historical observations. Here, we apply two different training methods to a supermodel of up to four different versions of the global atmosphere-ocean-land model SPEEDO. The standard version is regarded as truth. The first training method is based on an idea called cross pollination in time (CPT), where models exchange states during the training. The second method is a synchronization-based learning rule, originally developed for parameter estimation. We demonstrate that both training methods yield climate simulations and weather predictions of superior quality as compared to the individual model versions. Supermodel predictions also outperform predictions based on the commonly used multi-model ensemble (MME) mean. Furthermore, we find evidence that negative weights can improve predictions in cases where model errors do not cancel (for instance, all models are warm with respect to the truth). In principle, the proposed training schemes are applicable to state-of-the-art models and historical observations. A prime advantage of the proposed training schemes is that in the present context relatively short training periods suffice to find good solutions. Additional work needs to be done to assess the limitations due to incomplete and noisy data, to combine models that are structurally different (different resolution and state representation, for instance) and to evaluate cases for which the truth falls outside of the model class.

1 Introduction

1.1 Premises and the multi-model-ensemble

Although weather and climate models continue to improve, they will inevitably remain imperfect (*Bauer et al.*, 2015). Nature is so complex that it is impossible to model all relevant physical processes solely based on the fundamental laws of physics (think, for

instance, about the micro-physical properties of clouds that determine the cloud radiational properties). Progress in predictive power crucially depends on further improving our knowledge and the numerical representation of the physical processes the model is intended to describe. Nevertheless, with the best possible models in hand, more accurate predictions can be obtained by making good use of all of them, thus exploiting multi-model information. In order to reduce the impact of model errors on predictions, it is common practice to combine the predictions of a collection of different models in a statistical fashion. This is referred to as the multi-model ensemble (MME) approach: the MME mean prediction is often more skillful as model errors tend to average out (*Weigel et al.*, 2008), whereas the spread between the model predictions is naturally interpreted as a measure of the uncertainty about the mean (*IPCC*, 2013). Although MME tends to improve predictions of climate statistics (i.e., mean and variance), a major drawback is that it is not designed to produce an improved trajectory that can be seen as a specific climate forecast, given that averaging uncorrelated climate trajectories from different models leads to variance reduction and smoothing.

The foundation of modern weather and climate prediction rests on the assumption that when an estimate of the climate state is at disposal at a particular instance in time, its time evolution can be calculated by a proper application of a numerical discretization of the fundamental laws of physics, supplemented by empirical relationships describing unresolved scales and a complete specification of the external forcing and boundary conditions. Integration in time subsequently yields a predicted climate trajectory into the future, and formally frames the climate prediction endeavor as a mixed initial and boundary conditions problem (see, e.g., *Collins and Allen*, 2002; *Hawkins and Sutton*, 2009). Initial conditions, but also boundary conditions and external forcing are usually estimated by combining data with models via data assimilation techniques (see, e.g., *Carrassi et al.*, 2018, for a review). Errors in the time derivative (i.e. the model error) propagate into errors in the predicted trajectory but model error also affects the model statistics, so that the model and observed mean and variance differ, giving rise to model biases.

An illustrative example of this propagation of model errors is presented in *Rodwell and Jung* (2008) in relation to a change in the model's prescribed aerosol concentrations in the region of the Sahara. Already, within the first few hours of prediction, the different aerosol concentration leads to changing the stability and convection in the region. This in turn changes the upper air divergence and promotes the generation of large-scale Rossby waves that travel horizontally eastward and northward into the Northern Hemisphere during the subsequent week and finally impact the surface air temperatures in Siberia. This example demonstrates that a specific model error can impact model pre-

diction skills on far regions and diverse variables. Furthermore, it suggests that, in order to mitigate or in the best case to prevent model error from growing and affecting the whole model phase space, it is better to intervene at each model computational time step rather than a posteriori by combining outputs after a prediction is completed as in the MME approach.

1.2 Supermodeling

Reducing model errors early in the prediction is precisely what supermodeling attempts to achieve (*van den Berge et al.*, 2011). In a supermodel, different models exchange information during the simulation at every time step and form a consensus on a single best prediction. An advantage over the standard MME approach is that the supermodel produces a trajectory with improved long-term statistics. Improved trajectories are extremely valuable for calculations of the impact of climate on society. For instance, crop yields, spread of diseases and river discharges all depend on the specific sequences of weather events, not just on statistics (*Challinor and Wheeler*, 2008; *Sterl et al.*, 2009; *van der Wiel et al.*, 2019).

The supermodeling approach was originally developed using low-order dynamical systems (*Mirchev et al.*, 2012; *van den Berge et al.*, 2011) and subsequently applied to a global atmosphere model (*Schevenhoven and Selten*, 2017; *Wiegerinck and Selten*, 2017) and to a coupled atmosphere-ocean-land model (*Selten et al.*, 2017). A partial implementation of the supermodeling concept using real world observations was presented in *Shen et al.* (2016). In the original supermodeling concept, model equations are connected by nudging terms such that each model in the ensemble is nudged to the state of every other model at every time step. For appropriate connections, the ensemble of models eventually synchronizes on a common solution that depends on the strength of the connections. For instance, if all models are nudged to a particular model that is not nudged to any other model, the ensemble will follow that particular solution. By training connections on observed data, an optimal solution is found that is produced by the connected ensemble of models. This type of supermodel is referred to as *connected* supermodeling. *Wiegerinck and Selten* (2017) showed that in the limit of strong connections the connected supermodel solution converges to the solution of a weighted superposition of the individual model equations, referred to as a *weighted* supermodel. A crucial step in supermodeling is the training of the connection coefficients (for connected supermodels) or weights (for weighted supermodels) based on data, the observations. The first training schemes of supermodels were based on the minimization of a cost function dependent on long simulations with the supermodel (*Mirchev et al.*, 2012; *Shen et al.*, 2016; *van den Berge et al.*, 2011). Given that iterations, and thus

many evaluations of the cost function, were necessary in the minimization procedure, this approach turned out to be computationally very expensive. *Schevenhoven and Selten* (2017) developed a computationally very efficient training scheme based on cross pollination in time (CPT), a concept originally introduced by *Smith* (2001) in the context of ensemble weather forecasting. In CPT, the models in a multi-model ensemble exchange states during the simulation, generating mixed trajectories that exponentially increase in number in the course of time. As a consequence, a larger area of phase space is explored, thus increasing the chance that the observed trajectory is shadowed within the span of all of the mixed model trajectories. Given the above, CPT training is then based on the selection of the trajectory that remains closest to an observed trajectory. Another alternative efficient approach or training was introduced in *Selten et al.* (2017) to learn the connections coefficients in a supermodel. Their method, hereafter referred to as "synch rule" is based on synchronization and it is inspired by an idea originally proposed in *Duane et al.* (2007) for general parameter learning.

Before supermodeling becomes suitable for the class of large-dimensional state-of-the-art weather and climate models, we need to have training schemes that are computationally suitable for that context. In this paper, we develop, apply and compare CPT and the synch rule to train a weighted supermodel based on the intermediate complexity global coupled atmosphere-ocean-land model SPEEDO (*Severijns and Hazeleger*, 2010). Short-term supermodel prediction skill as well as long-term climate statistics show that both training methods result in supermodels that outperform the individual models. Furthermore, novel experiments with negative weights, as opposed to the standard case of weights larger than or equal to zero, suggest that even when the individual model biases do not compensate for each other an improved supermodel solution can be achieved.

In Sect. 2, the two types of supermodels, *connected* and *weighted*, are introduced in detail. Section 3 describes the global coupled atmosphere-ocean-land model SPEEDO and the construction of a SPEEDO supermodel. The two training strategies are described in Sect. 4 with specific details when applied to the SPEEDO model in Sect. 5. The results of the training are shown in Sect. 6. The final section discusses the results and lists further steps to be taken towards training a supermodel based on state-of-the-art weather and climate models using real-world observations.

2 Weighted and connected supermodeling

To make the supermodeling approach more explicit, we formally write the model equations of a weather or climate model i as

$$\dot{\mathbf{x}}_i = \mathbf{f}_i(\mathbf{x}_i, \mathbf{p}_i), \tag{1}$$

where $\mathbf{x}_i$ is a high-dimensional state vector and $\mathbf{f}_i$ a non-linear evolution function depending on the state $\mathbf{x}_i$ and on a number of adjustable parameters $\mathbf{p}_i$. In practice, weather and climate models generally differ in the representation of the climate state, i.e., the phase where $\mathbf{x}_i$ is defined, the evolution function and parameter values. In this stage of developing the supermodeling approach and training schemes, we simplify the context and focus on a situation where the models share the same evolution function, $\mathbf{f}$, and the same phase space, so that $\mathbf{x}_i \in \mathbb{R}^n$ for all i. However, the models differ in the parameters, $\mathbf{p}_i \neq \mathbf{p}_j$ if $i \neq j$. The approach can be generalized using data assimilation approaches (*Du and Smith*, 2017). We will furthermore denote the "truth" as given by the model $\mathbf{f}$ with a specific set of parameters. An ensemble of imperfect models can be dynamically combined in a *weighted* or *connected* supermodel.

2.1 Weighted supermodeling

A weighted supermodel based on two imperfect models is given by

$$\dot{\mathbf{x}}_1 = \mathbf{f}(\mathbf{x}_s, \mathbf{p}_1) \tag{2a}$$

$$\dot{\mathbf{x}}_2 = \mathbf{f}(\mathbf{x}_s, \mathbf{p}_2) \tag{2b}$$

$$\dot{\mathbf{x}}_s = \mathbf{W}_1\dot{\mathbf{x}}_1 + \mathbf{W}_2\dot{\mathbf{x}}_2, \tag{2c}$$

where $\mathbf{x}_s \in \mathbb{R}^n$ represents the supermodel state vector and diagonal matrices $\mathbf{W}_1 = \mathrm{diag}(\mathbf{w}_1)$ with $\mathbf{w}_1 \in \mathbb{R}^n$ denote the weights. In the weighted supermodel, the states are imposed to be perfectly synchronized. Training a weighted supermodel implies training the weights $\mathbf{w}_i$.

2.2 Connected supermodeling

For completeness and for comparison of the weighted supermodels with the connected supermodel from *Selten et al.* (2017), we introduce the equations for the connected supermodel. A connected supermodel based on two imperfect models is given by

$$\dot{\mathbf{x}}_1 = \mathbf{f}(\mathbf{x}_1, \mathbf{p}_1) - \mathbf{C}_{12}(\mathbf{x}_1 - \mathbf{x}_2) \tag{3a}$$

$$\dot{\mathbf{x}}_2 = \mathbf{f}(\mathbf{x}_2, \mathbf{p}_2) - \mathbf{C}_{21}(\mathbf{x}_2 - \mathbf{x}_1) \tag{3b}$$

$$\dot{\mathbf{x}}_s = \frac{1}{2}(\dot{\mathbf{x}}_1 + \dot{\mathbf{x}}_2), \tag{3c}$$

Note the nudging terms (the rightmost terms in Eq. 3a and Eq. 3b) that push the state of each model to the state of the other at every time step. The size of the nudging terms $\mathbf{C}_{12}$ and $\mathbf{C}_{21}$ reflects the strength of the coupling between the two models. They have the form of diagonal matrices $\in \mathbb{R}^{n\times n}$ and can thus be written as $\mathbf{C}_{12} = \mathrm{diag}(\mathbf{c}_{12})$ with $\mathbf{c}_{12} \in \mathbb{R}^n$. The diagonal form reflects the fact that each model state vector component is nudged towards the same component of the other model. The approach can be extended to be multivariate allowing for cross nudging, but this will require careful scaling of the variables. For appropriate connections, the models fall into a synchronized motion (*Pecora and Carroll*, 1990). Because in general the synchronization will not be perfect due to the different parameter values, the supermodel solution $\mathbf{x}_s$ is defined as the average of the different model states. Note that the states will be close for strong connections so that smoothing and loss of variance due to the averaging will be limited. The supermodel solution depends on the relative strengths of connection coefficients. Training a connected supermodel implies training the value of the connection coefficients.

A connected supermodel allows for more flexibility in the event that the ensemble is not perfectly synchronized (*Wiegerinck et al.*, 2013). In regions of phase space of strong divergence, for instance, one model can pull the ensemble along if it takes a very different trajectory. However, in *Wiegerinck et al.* (2013), it is noted that the size of the connection coefficients after training is typically quite large. The larger the coefficients, the stronger the models converge on a synchronized trajectory, which can be described by a weighted superposition of the models (*Wiegerinck et al.*, 2013). Since for some training applications, perfect synchronization is required as we shall see in Sect. 4, only weighted supermodels are considered in this paper. We do not limit ourselves to combining only two imperfect models into a supermodel; also, combining four imperfect models will be discussed.

3 SPEEDO climate model

The SPEEDO global climate model consists of an atmospheric component (SPEEDY) that exchanges information with a land (LBM) and an ocean-sea-ice component (CLIO) using coupling routines (Fig. 1). The coupling routines perform re-gridding operations between the computational grids of the different modules. A detailed description of SPEEDO can be found in *Selten et al.* (2017); *Severijns and Hazeleger* (2010).

The atmospheric model SPEEDY describes the evolution of the two horizontal wind components U (east-west) and V (north-south), temperature T and specific humidity q at eight vertical levels and the surface pressure p_s. Relatively simple calculations

of heating and cooling rates due to radiation, convective transports, cloud amounts, precipitation and turbulent heat, water and momentum exchange at the surface are performed at a computational grid of approximately 3.75° horizontal spacing (48 ×96 grid cells).

SPEEDY exchanges water and heat with the land model LBM that uses three soil layers and up to two snow layers to close the hydrological cycle over land and a heat budget equation that controls the land temperatures. The horizontal discretization is the same as for the atmosphere model. The land surface reflection coefficient for solar radiation is prescribed using a monthly climatology. Each land bucket has a maximum soil water capacity. The runoff is collected in river-basins and drained into the ocean at specific locations of the major river outflows.

SPEEDY exchanges heat, water and momentum with the ocean model CLIO (*Goosse and Fichefet*, 1999). CLIO describes the evolution of ocean currents, temperature and salinity on a computational grid of 3° horizontal resolution and 20 unevenly spaced layers in the vertical. A three-layer thermodynamic-dynamic sea-ice model describes the evolution of sea-ice in the event that ocean temperatures drop below freezing levels. Heat storage in the snow-ice system is accounted for and snow amounts and ice thickness evolve in response to surface and bottom heat fluxes. Sea ice is considered to behave as a viscous-plastic continuum as it moves under the action of winds and ocean currents.

Formally, the SPEEDO equations can be written as

$$\dot{\mathbf{a}} = \mathbf{f}^a(\mathbf{a};\mathbf{p}^a) + \mathbf{g}^a(\mathbf{e}^h,\mathbf{e}^w,\mathbf{e}^m) \tag{4a}$$

$$\dot{\mathbf{o}} = \mathbf{f}^o(\mathbf{o};\mathbf{p}^o) + \mathbf{g}^o(\mathcal{P}^o\mathbf{e}^h,\mathcal{P}^o\mathbf{e}^w,\mathcal{P}^o\mathbf{e}^m,\mathcal{P}^o\mathbf{r}) \tag{4b}$$

$$\dot{\mathbf{l}} = \mathbf{f}^l(\mathbf{l};\mathbf{p}^l) + \mathbf{g}^l(\mathcal{P}^l\mathbf{e}^h,\mathcal{P}^l\mathbf{e}^w,\mathbf{r}), \tag{4c}$$

where $\mathbf{a}$ is the atmospheric state vector, $\mathbf{o}$ the ocean/sea-ice state vector, $\mathbf{l}$ the land state vector, $\mathbf{e}^h$ the heat exchange vector between atmosphere and surface, $\mathbf{e}^w$ the water exchange vector, $\mathbf{e}^m$ the momentum exchange vector and $\mathbf{r}$ the river outflow vector describing the flow of water from land to ocean. The exchange vectors depend on the state of the atmosphere and the surface but this dependency is not made explicit in Eq. 4 to simplify the notation. The projection operators $\mathcal{P}$ represent the regridding operations between the computational grids. These operations are conservative so that the globally integrated heat and water loss of the atmosphere at any time at the surface equals the integrated heat and water gain of the land and ocean. The non-linear functions $\mathbf{f}$ represent the cumulative contribution of the modeled physical processes to the change in the climate state vector and depend on the values of the parameter vectors $\mathbf{p}$. Some of these parameters go through a daily and/or seasonal cycle and/or have a spatial de-

pendence like the reflectivity of the surface. The non-linear functions **g** describe how the exchange of heat, water and momentum between the subsystems affects the change of the climate state vector.

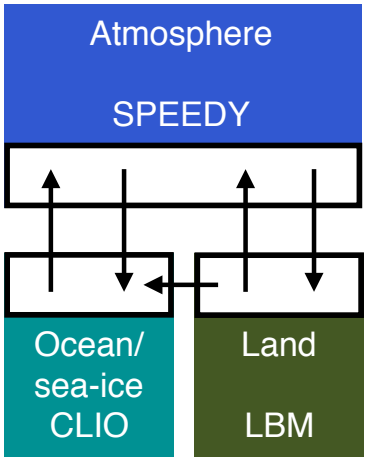


Figure 1: Schematic representation of the SPEEDO climate model. The atmosphere needs surface characteristics (temperature, roughness, reflectivity, soil moisture) in order to calculate the exchange of heat, water and momentum. Coupler software communicates this information between the components and interpolates between the computational grids.

3.1 SPEEDO supermodel

The training experiments of this study are evaluated in a noise-free observation framework, with perfect observations generated by sampling a reference model trajectory. This "perfect model" provides a set of time-ordered observations, called the "truth". We consider the SPEEDO climate model with standard parameter values as truth and create imperfect models by perturbing parameter values in the atmospheric component. A supermodel is formed by combining the imperfect atmosphere models through a weighted superposition of the time derivatives of the imperfect models (Eq. 2) which are each coupled to the same ocean and land model (Fig. 2). All atmosphere models receive the same state information from the ocean and land model but each calculates their own water, heat and momentum exchange. On the other hand, the ocean and land model receive the multi-model weighted average of these atmospheric components; this follows the interactive ensemble approach (*Kirtman and Shukla*, 2002). Following

Eq. 2 , the SPEEDO weighted supermodel equations are given by

$$\dot{\mathbf{a}}_i = \mathbf{f}^a(\mathbf{a}_s;\mathbf{p}_i^a) + \mathbf{g}^a(\mathbf{e}_i^h, \mathbf{e}_i^w, \mathbf{e}_i^m) \tag{5a}$$

$$\dot{\mathbf{o}} = \mathbf{f}^o(\mathbf{o};\mathbf{p}^o) + \mathbf{g}^o(\mathcal{P}^o\overline{\mathbf{e}^h}, \mathcal{P}^o\overline{\mathbf{e}^w}, \mathcal{P}^o\overline{\mathbf{e}^m}, \mathcal{P}^o\mathbf{r}) \tag{5b}$$

$$\dot{\mathbf{l}} = \mathbf{f}^l(\mathbf{l};\mathbf{p}^l) + \mathbf{g}^l(\mathcal{P}^l\overline{\mathbf{e}^h}, \mathcal{P}^l\overline{\mathbf{e}^w}, \mathbf{r}) \tag{5c}$$

$$\dot{\mathbf{a}}_s = \sum_i \mathbf{W}_i \dot{\mathbf{a}}_i, \tag{5d}$$

where $\dot{\mathbf{a}}_s$ denotes the time derivative of the supermodel, $\mathbf{W}_i$ values denote diagonal matrices with weights on the diagonal, i refers to imperfect model i, and the overbar denotes a weighted average over the models.

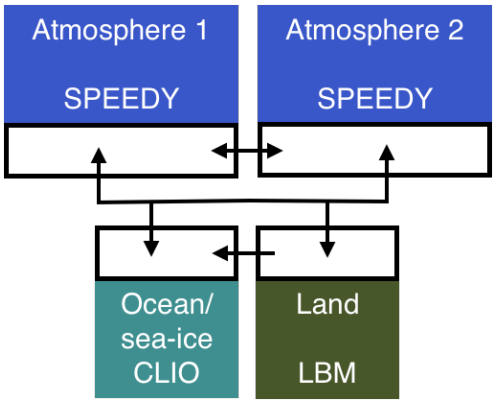


Figure 2: Schematic representation of the SPEEDO climate supermodel based on two imperfect atmosphere models. The two atmosphere models exchange water, heat and momentum with the perfect ocean and land model. The ocean and land model send their state information to both atmosphere models. The atmosphere models exchange state information in order to combine their time derivatives.

4 Learning methods

Two different learning strategies are evaluated in this study in order to train the SPEEDO weighted supermodel: learning based on CPT as developed and applied to low-order dynamical systems in *Schevenhoven and Selten* (2017), and learning based on synchronization as applied to a connected SPEEDO supermodel in *Selten et al.* (2017).

4.1 Cross pollination in time

The Cross Pollination in Time (CPT) learning approach is based on an idea proposed by *Smith* (2001). CPT "crosses" trajectories of different models in order to create a larger solution space. The aim is to generate trajectories that follow the truth more

closely. The training phase of CPT starts from an observed initial condition in state space. For simplicity, assume the model is one-dimensional. From the same initial state, the imperfect models compute one time step each ending in a different state. Next, all models compute another time step from each of these new states. Continuing this process leads to a rapid increase in the number of trajectories with time (Fig. 3a) that will ultimately cover a larger area of the state space. Among the full set of mixed trajectories, the one which is closest to the truth (i.e., to the data) is continued; the others are discarded, resulting in a pruned ensemble, as is depicted in Fig. 3b.

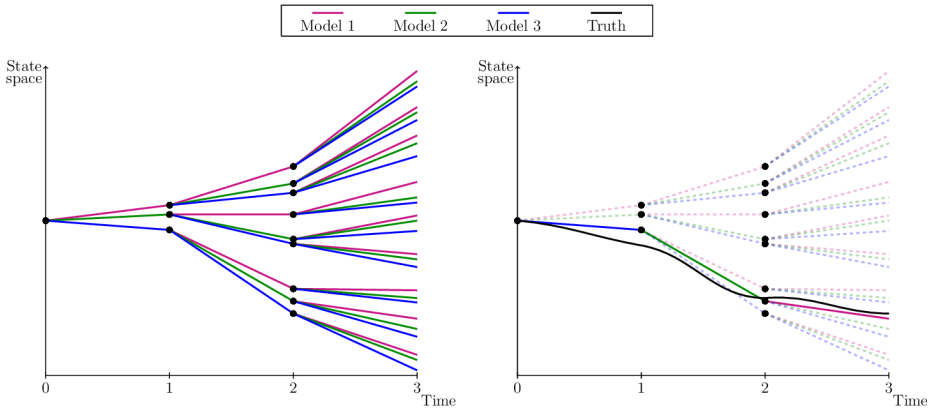


*Figure 3: Adapted from Schevenhoven and Selten (2017). A one-dimensional schematic of CPT for three models, a full ensemble (**a**) and a pruned ensemble (**b**). Note that the "truth" has been drawn here as a continuous line for illustrative purpose. In practice, the truth is only known at discrete times (the observation times) and the distance with respect to model trajectories is computed at those times only.*

In the case of a multi-dimensional model, such as SPEEDO, it is possible that at each time step different models are closest to the truth for different state variables and at different grid locations. In this case, we continue per state variable with the model that is closest. This means that the initial state for the next time step can consist of a combination of models. As the values for the different state variables might not be in agreement with each other, this creates imbalances that can lead to numerical instabilities. A (partial) solution is to decrease the time step, as we shall see in Sect. 5.

The training period is terminated when the CPT trajectory starts to deviate from the truth beyond a given pre-specified threshold. After training, an optimal trajectory is obtained that is produced by a combination of different imperfect models (Fig. 4). Next, we count how often during training each model has produced the best prediction of a particular component of the state vector. This frequency of occurrences is used to

compute weights **W** for the corresponding time derivative of the state vector. This superposition of weighted imperfect models forms a weighted supermodel, as expressed in the example of Eq. 2.

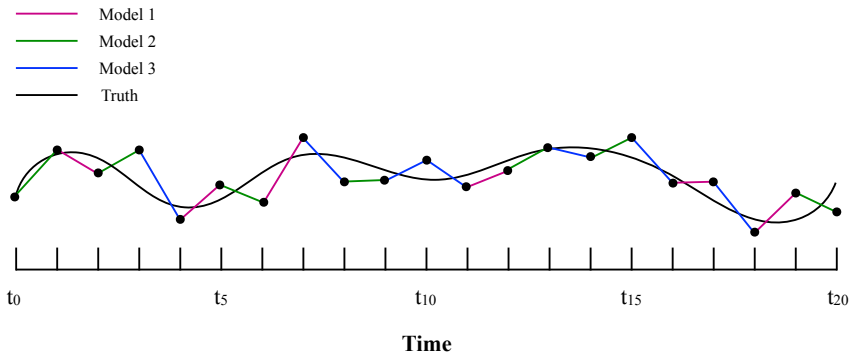


Figure 4: CPT trajectory after a training period of 20 time steps. Model 1 is used for 6 out of 20 time steps; hence, model 1 will get a weight of 0.3.

4.2 Synchronization-based learning

For the training of a supermodel based on synchronization, a learning rule (the synch rule) is used that updates the weights such that synchronization errors between truth and supermodel are minimized. In contrast to CPT learning, initial values for the weights need to be chosen and the weights are updated during training. Under certain conditions, the supermodel will fall into synchronized motion with the truth as the weights are updated and the supermodel is nudged to the truth (black arrows in Fig. 5).

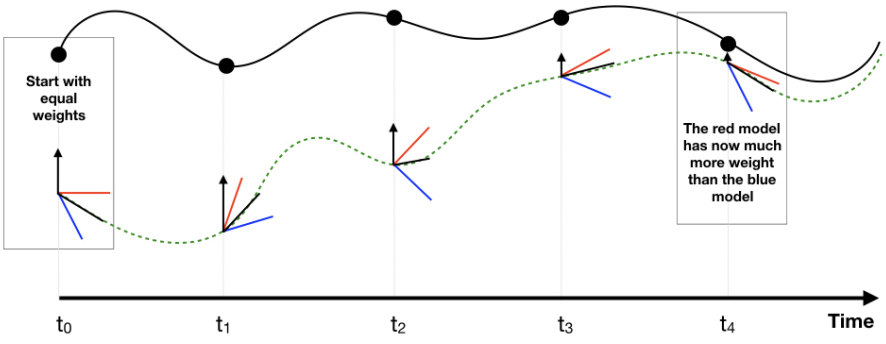


Figure 5: At each observation (dots) of the truth (continuous black line), the weights of the imperfect models (red, blue) are updated which gives a new supermodel solution (green dotted line). The black arrows indicate the nudging to the truth.

The synch rule for the weights is an application of the general synchronization-based parameter estimation approach suggested in (*Duane et al.*, 2007). Recently, the synch rule was applied to train the connections in a connected SPEEDO supermodel (*Selten et al.*, 2017). We follow a similar strategy and implement the synch rule to train the weights of a weighted SPEEDO supermodel.

In the context of two dynamical systems that differ in parameter values only, the general synch rule for parameter estimation is given by

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x};\mathbf{p}) \tag{6a}$$

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y};\mathbf{q}) - \mathbf{K}(\mathbf{y}-\mathbf{x}) \tag{6b}$$

$$\dot{q}_j = -\delta_j \sum_i e_i \frac{\partial f_i(\mathbf{y},\mathbf{q})}{\partial q_j}, \tag{6c}$$

where $\mathbf{p}$ and $\mathbf{q}$ are vectors of parameters. $\mathbf{K}(\mathbf{y}-\mathbf{x})$ is a connecting term between the two systems that nudges $\mathbf{y}$ towards $\mathbf{x}$. $\mathbf{K}$ is a diagonal matrix of nudging coefficients, $\mathbf{K} = \mathrm{diag}(\mathbf{k})$. Suppose the two systems Eq. 6a and Eq. 6b synchronize if $\mathbf{p} = \mathbf{q}$; that is, as $t \to \infty$, $\mathbf{y}(t) \to \mathbf{x}(t)$. We further assume that the parameters appear only linearly in the model equations. Then it can be proven that, using the learning rule Eq. 6c, even if the two systems are not identical, $\mathbf{p} \neq \mathbf{q}$, the systems will still synchronize and the parameters will become equal, $\mathbf{q}(t) \to \mathbf{p}$ as $t \to \infty$. Here, q_j denotes the parameter values, with j indexing the elements of the parameter vector. Furthermore, $e_i = y_i - x_i$ denotes the synchronization error at the current time step with i indexing the elements of the state vector and δ_j an adjustable rate of learning scaling factor. At every time step, the update $\dot{q}_j$ for the weight q_j is calculated.

In training a supermodel, we assume that the truth can be described by a weighted dynamical combination of imperfect models with the weights as adjustable parameters. In this case, the function $\mathbf{f}$ corresponds to the supermodel time derivative, $\mathbf{q}$ corresponds to the weights of the supermodel, $\mathbf{x}$ denotes the truth and $\mathbf{y}$ the supermodel solution. The derivative of $\mathbf{f}$ with respect to a certain weight is the tendency of the imperfect model belonging to that weight (see Eq. 2c). In our SPEEDO case, the truth cannot exactly be described as a weighted superposition of imperfect models since the perturbed parameters do not appear linearly in the equations, yet the approximation is close enough for the learning rule to work well.

Integration of the synch rule implies that as long as the time-series of the synchronization error e_i and the effect of the parameter on the imperfect model evolution $\frac{\partial f_i(\mathbf{y},\mathbf{q})}{\partial q_j}$ are correlated, the parameter will be updated. For instance, when a parameter update systematically enhances warming in the model when the model is colder than the truth and the same holds when the parameter update systematically cools the model when it is

too warm, then the updated parameter will decrease the synchronization error between the model and truth over time. When this correlation vanishes, hence there is no systematic relation anymore between updating the parameter and the state of the model; then systematic updates cease. When perfect synchronization is reached (hence $e_i = 0$), then naturally updates also stop.

5 Training in SPEEDO

In training the SPEEDO supermodel, we regard the atmospheric model with standard parameter values as truth, whereas imperfect atmospheric models are created by perturbing those parameter values. Figure 6 depicts the configuration during training. All atmosphere models are independently coupled to the same ocean and land model. They each calculate their own water, heat and momentum fluxes and receive the information from the ocean and the land model from the truth only.
During training, the truth and imperfect models all share their states. In the case of CPT, this state information is used by each imperfect model to check which model is closest to the truth and continue the integration from that state. In the case of the synch rule, this state information is used to calculate the synchronization error between the supermodel and the truth.

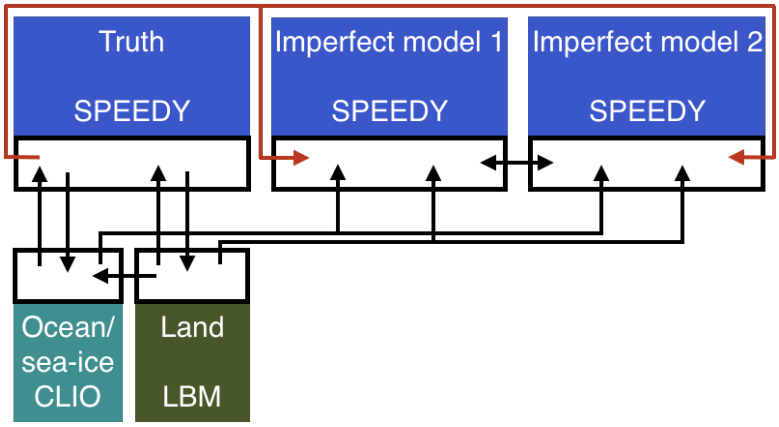


Figure 6: Schematic representation of the SPEEDO system during training (Selten et al., 2017).

Application of the synch rule to a weighted SPEEDO supermodel of two imperfect

models implies integration of the following set of equations:

$$\dot{\mathbf{a}}_0 = \mathbf{f}^a(\mathbf{a}_0;\mathbf{p}_0^a) + \mathbf{g}^a(\mathbf{e}_0^h,\mathbf{e}_0^w,\mathbf{e}_0^m) \tag{7a}$$

$$\dot{\mathbf{a}}_1 = \mathbf{f}^a(\mathbf{a}_s;\mathbf{p}_1^a) + \mathbf{g}^a(\mathbf{e}_1^h,\mathbf{e}_1^w,\mathbf{e}_1^m) - \mathbf{K}(\mathbf{a}_s - \mathbf{a}_0) \tag{7b}$$

$$\dot{\mathbf{a}}_2 = \mathbf{f}^a(\mathbf{a}_s;\mathbf{p}_2^a) + \mathbf{g}^a(\mathbf{e}_2^h,\mathbf{e}_2^w,\mathbf{e}_2^m) - \mathbf{K}(\mathbf{a}_s - \mathbf{a}_0) \tag{7c}$$

$$\dot{\mathbf{o}} = \mathbf{f}^o(\mathbf{o};\mathbf{p}^o) + \mathbf{g}^o(\mathcal{P}^o\mathbf{e}_0^h,\mathcal{P}^o\mathbf{e}_0^w,\mathcal{P}^o\mathbf{e}_0^m,\mathcal{P}^o\mathbf{r}) \tag{7d}$$

$$\dot{\mathbf{l}} = \mathbf{f}^l(\mathbf{l};\mathbf{p}^l) + \mathbf{g}^l(\mathcal{P}^l\mathbf{e}_0^h,\mathcal{P}^l\mathbf{e}_0^w,\mathbf{r}) \tag{7e}$$

$$\dot{\mathbf{a}}_s = \mathbf{W}_1\dot{\mathbf{a}}_1 + \mathbf{W}_2\dot{\mathbf{a}}_2 \tag{7f}$$

$$\dot{W}_{i,j} = -\delta_j(a_{s,j} - a_{0,j})\dot{a}_{i,j}, \tag{7g}$$

where index 0 refers to the truth and $W_{i,j}$ refers to the weight of model i and state vector element j. During training, we choose a uniform nudging strength corresponding to a 24 h time scale, as motivated by *Selten et al.* (2017). They showed that, for this value of $\mathbf{K}$, two connected identical SPEEDO models (perfect model scenario) almost perfectly synchronize with very small synchronization errors in temperatures of the order of 0.01°C. Imperfect models, on the other hand, have synchronization errors with respect to the truth that are usually 10 times larger.

5.1 Construction of imperfect models

In order to be able to compare results of the weighted supermodels of this study to the connected supermodels in *Selten et al.* (2017), we choose the same parameter values for the imperfect models. These parameters are the convection relaxation timescale, the relative humidity threshold and the momentum diffusion timescale. The reason to perturb these parameters is because the uncertainty in climate models mostly lies in the parameterization of clouds and convection, and perturbing these parameters in the SPEEDO model results in a spread in the simulated climate that characterizes this uncertainty. The parameters are listed in Table 1, where model 1 and model 2 correspond to the imperfect models of *Selten et al.* (2017). The impact of the parameter perturbations on the climate (i.e., long-term behavior) of the models is assessed on the basis of 40-year simulations initiated on 1 January of model year 2001 of a long control simulation as in *Selten et al.* (2017). Table 2 shows the global mean average difference between the truth and the imperfect models of Table 1 for different variables. From the table, it appears evident how the imperfect models all drift away from the truth giving raise to biases. For example, the global mean temperature of imperfect model 1 rises about 1.4°C within a couple of decades, whereas model 2 cools around 0.4°C. These global mean temperature biases are comparable to the biases of state-of-the-art global climate models compared to real-world observations (*IPCC*, 2013).

The first supermodel that we will train will consist of a weighted superposition of models 1 and 2. The second supermodel will consist of a weighted superposition of models 1, 3, 4 and 5. The parameter values of these models are chosen such that they form a so-called convex hull around the true parameter values (see *Schevenhoven and Selten*, 2017, for a discussion on the convex hull principle). Note that we use only two perturbed values for each parameter; the imperfect models differ only in the combination of these values, such that in the four-model supermodel, a convex hull is formed. This implies that, provided the model functional dependence on the parameters is linear, the true parameter values can be obtained as a linear combination with positive coefficients/weights of the four parameter values of the imperfect models. While these conditions do not perfectly hold in this case, we expect that we can create a weighted supermodel based on these four models that will be close to the truth. All of these four models overestimate the global mean temperature and precipitation (Table 2). Therefore, simply taking the MME mean with positive weights will not produce a climatology closer to the truth. However, we expect that, based on the convex hull principle, the weighted supermodel will nevertheless be able to produce a climatology that is closer to the truth.

The third supermodel consists of a weighted superposition of models 1 and 6. In this case, both imperfect models have parameter values that are smaller than the corresponding true values. A weighted superposition with positive weights does not correspond to a model with parameter values that are closer to the truth. Note that both models overestimate the average temperature and precipitation (Table 2); hence, taking the MME mean with positive weights also does not produce a climatology closer to the truth. In this case, we will explore whether a weighted supermodel with negative weights can be trained in order to improve the climatology and short-term forecasts.

Table 1: Parameter values of perfect and imperfect models.

model	convection relaxation timescale	relative humidity threshold	momentum diffusion timescale
perfect	6 hours	0.9	24 hours
model 1	4 hours	0.85	18 hours
model 2	8 hours	0.95	30 hours
model 3	4 hours	0.95	30 hours
model 4	8 hours	0.95	18 hours
model 5	8 hours	0.85	30 hours
model 6	3 hours	0.75	14 hours

Table 2: Global mean average difference between the imperfect models and the perfect model, calculated over the last 30 years of the simulation.

model	temperature $[C^\circ]$	precipitation $[mm/day]$	wind at 200 hPa $[m/s]$	wind at 850 hPa $[m/s]$	solar surface radiation $[W/m^2]$	cloudcover $[\%]$
mod 1	1.37	0.11	1.04	0.07	2.06	-1.59
mod 2	-0.38	-0.04	-0.31	-0.03	-1.13	0.87
mod 3	0.99	0.10	1.14	0.06	1.21	-1.03
mod 4	0.45	0.04	-0.04	-0.01	-0.20	0.10
mod 5	0.86	0.08	0.72	-0.01	-0.19	-0.12
mod 6	3.20	0.26	2.25	0.03	3.95	-3.37

5.2 Global weights

For both CPT and the synch rule, we choose to work with global weights, which means that for each meteorological variable we use the same weight at every grid point. In principle, one could allow different weights per each grid point but it could induce dynamic imbalances that pull the model away from its attractor. The model's reaction is then to restore the dynamical balances and return to its own attractor (*Pecora and Carroll*, 1990). In SPEEDO, this leads to the generation of fast gravity waves and fast convective adjustments. An adequately small time step is required in order to prevent numerical instabilities. We choose instead to use global weights in order to limit the computational time.

5.3 Exchange of state information

The SPEEDO model has five prognostic variables: temperature, vorticity, divergence, specific humidity and surface pressure (T, VOR, DIV, TR, PS). Best results were obtained by limiting the weighted averaging of state information to temperature, vorticity and divergence only. We suspect that exchanging specific humidity and surface pressure leads to imbalances and fast spurious adjustments that deteriorate the supermodel solution. We found that a perfect SPEEDY atmosphere only fully synchronizes with the truth when at least temperature, vorticity and divergence are nudged to the truth (not shown). Therefore, in a weighted supermodel, at least these variables need to be exchanged.

5.4 Required time step

We found that smaller time steps were required during CPT training as compared to standard integrations. Gravity waves induced by the state replacement during training require a smaller time step in order to prevent numerical instabilities. We found that a 15 min time step was sufficient with our choice of imperfect models, which is half the

time step of the standard integration.

5.5 Initialization of the weights for the synch rule

In CPT training, the sum of the weights is normalized to 1. In the application of the synch rule, on the other hand, the sum of the weights is not explicitly constrained. One can start from zero weights and let the synch rule find the optimal set of weights. Initializing weights with a sum larger than 1 easily leads to numerical instabilities because the weighted mean state becomes more energetic. Imposing the constraint of the sum of weights being 1 during the training also led to numerical instabilities. We chose to initialize with equal weights that sum to 1.

5.6 Rate of learning in the synch rule

The synch rule contains an adjustable rate of learning scaling factor δ_j, with j the index of the state vector. A large rate of learning is desirable since it leads to faster convergence and shorter training periods. However, the parameters should vary on a slower time-scale than the dynamical variables and this provides an upper bound for the value of δ_j. Furthermore, it turns out that if δ_j is too large, the sum of the weights can become greater than 1 which easily leads to numerical instabilities. The size of δ_j in the synch rule depends on the variable that is being exchanged and was determined by trial and error during the training experiments. The largest values for δ_j that resulted in converged weights were on the order of 10^7 for divergence and vorticity and 10^{-4} for temperature. With these scaling factors, approximately similar rates of learning were achieved for the different variables. This makes sense since the state values for divergence and vorticity are much smaller than for temperature, so the product of δ_j and the state values in the synch rule is of the same order of magnitude.

5.7 Weights for the heat, water and momentum fluxes

In the experimental setup during training, we assume a perfect ocean and land models which receive fluxes from the perfect atmosphere. However, in the supermodel setup, perfect fluxes are not available and we use a weighted combination of the fluxes from both imperfect models instead. In the connected supermodel of *Selten et al.* (2017), the fluxes are averaged using equal weights. In this paper, we further optimize the weights for the fluxes, because we found they have a big influence on the supermodel's performance. In particular, we selected weights given by the average of the three weights for the prognostic variables. To check whether this choice was optimal, we used a least squares minimization method in order to optimize the weights for the fluxes. During one year of training, the fluxes from the perfect and imperfect models were saved at

every time step. The weights were determined by a least squares fit of a weighted sum of the imperfect fluxes to the perfect fluxes. The flux weights obtained from the minimization method did differ slightly per flux (heat, water or momentum flux), but the average weights were close to the average of the weights for the prognostic variables.

6 Results

We describe the learning results and the forecast short- and long-term capabilities of the three supermodel configurations separately.

6.1 Supermodels based on two imperfect models

We first trained a weighted supermodel based on imperfect models 1 and 2 (see Table 1), applying both CPT and the synch rule. As a benchmark, we compare the quality of the weighted supermodel after training with the connected SPEEDO supermodel of *Selten et al.* (2017). This supermodel is based on the same imperfect models and was trained by the synch rule.

Ideally, both CPT and the synch rule should produce converged weights, i.e., weights that remain stable if the training period is extended. The required length of the training period for the convergence of the two methods turns out to be very different. For CPT, a training period as short as a couple of days produces converged weights, whereas for the synch rule it takes about a year. Note that we limit the CPT training period to a week, as the CPT trajectory starts to deviate significantly from the truth after approximately 10 days. The reason that CPT diverges from the truth is because we have a limited ensemble size. With non-linear processes causing rapid error growth, the truth soon falls outside the limited ensemble. The problem is exacerbated by replacing a model state with state variables mixed from different models which introduces imbalances that cause additional error growth.

In order to check the difference between the CPT weights during a year, the CPT method is applied for each week during 1 year. After each week, the values for all prognostic variables are reset to the truth, and the procedure is repeated. Figure 7a shows the values of the weights during training. The weights for both temperature and vorticity remain fairly constant. The weights for divergence vary within 0.04 of a mean value. For the final supermodel weights, we just take the average over the whole year (Table 3).

Using the synch rule, weights for temperature and vorticity converge within the first couple of weeks, whereas for divergence the weights cannot be learned faster than within a year in order to avoid numerical instabilities (see Fig. 7b) . When using the

synch rule, the weights converge to similar values as compared to the CPT training (Table 3). Converged values of both methods are within 0.05. Whether these small differences matter for climate and weather forecasts will be assessed in the next two sections. Although not imposed, the training yields sum of weights equal to 1 as an optimal solution.

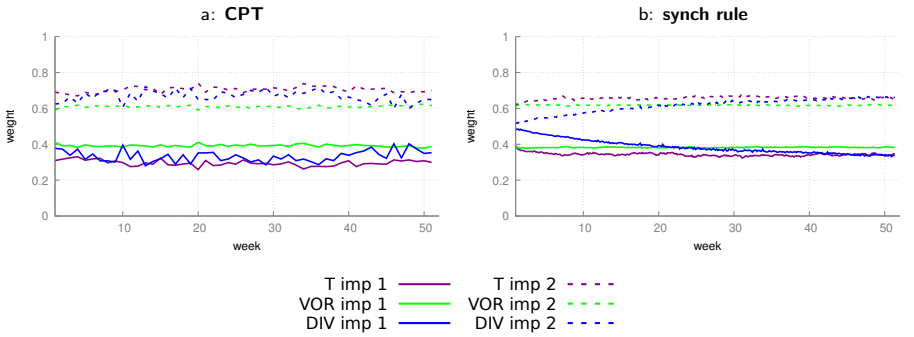


*Figure 7: Calculation of weights for a supermodel constructed from two imperfect models using two different training schemes. (**a**) CPT weights calculated during a training period of 1 week estimated for each week of a year. (**b**) Weights for the synch rule during a training period of 1 year.*

Table 3: Weights for the supermodel trained by CPT and the synch rule. Between brackets, the standard deviation over the year (CPT) or the standard deviation over the last 10 weeks of training (synch rule) is given.

model	method	T	VOR	DIV
model 1	**CPT**	0.30 (0.016)	0.39 (0.007)	0.35 (0.031)
model 2		0.70 (0.016)	0.61 (0.007)	0.65 (0.031)
model 1	**synch rule**	0.35 (0.0043)	0.38 (0.0018)	0.34 (0.0052)
model 2		0.65 (0.0043)	0.62 (0.0018)	0.66 (0.0053)

6.1.1 Climate measures

The imperfect models and the supermodel are integrated for 40 years in time, starting from 1 January of model year 2001. The climatology is defined as the average over years 11-40. The error in the climatology is defined as the root of the global mean squared error (RMSE) between the model and the truth. In addition, the perfect model is integrated for 40 years from a slightly perturbed initial condition, in order to obtain an estimate of the sampling error, i.e., to estimate the representativeness of the errors of the different models. Global mean time-series for surface air temperature, precipitation, surface solar radiation and cloud cover for the different models show that both weighted supermodels behave very similar and remain close to the perfect model (Fig.

8). The errors in the climatologies of the various fields of both supermodels are much reduced as compared to both imperfect models and are indistinguishable from the statistical sampling error of the perfect model. Both training methods succeed in greatly improving the simulation of the climate. Compared to the trained connected supermodels of *Selten et al.* (2017), the weighted supermodels have reduced climatological errors (see Fig. 15). Training of a connected supermodel by the synch rule on the other hand is more efficient as faster learning rates could be used, leading to convergence within 2 weeks of training.

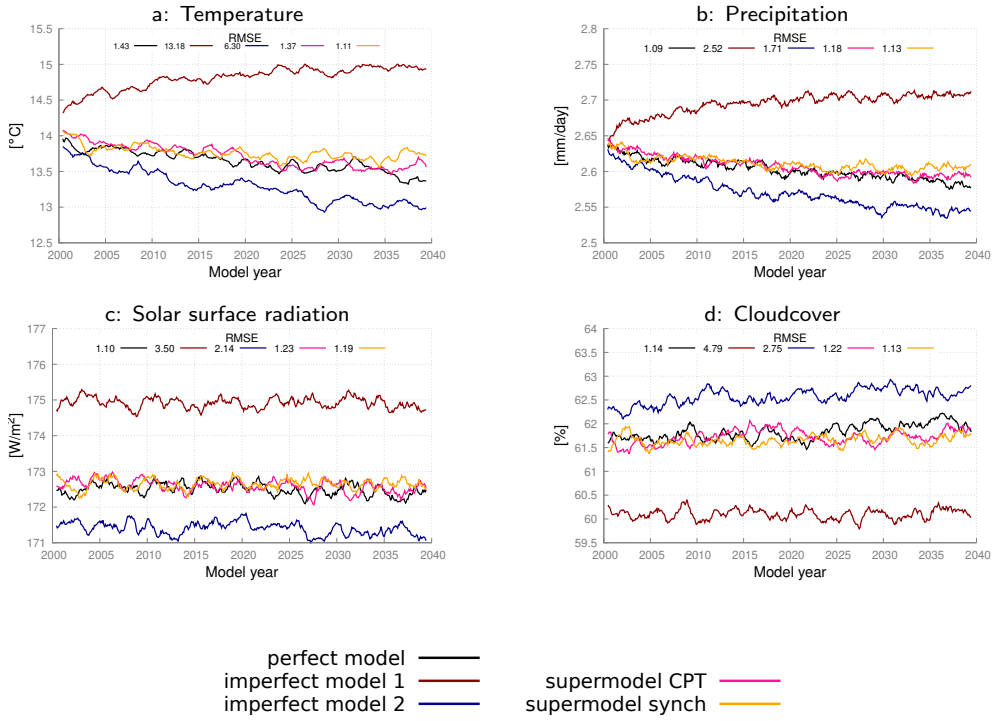


Figure 8: Global mean time-series for the perfect model, the imperfect models and the two supermodels trained by CPT and the synch rule. The normalized root mean squared error (RMSE) in the climatology of model years 2011-2040 with respect to the climatology of the truth is given in each panel. The normalization is such that the expected value of the perfect model error is 1.

A spatial characterization of the performance of the supermodel in simulating the climatology of the zonal wind at 200 hPa is given in Fig. 9. Clearly, both supermodels outperform the imperfect models and their local errors are of similar magnitude as the sampling error of the perfect model. We computed an optimal weighted average of the

climatology of both imperfect models (optimal in the sense that the RMSE in the climatology is minimized) as in *Selten et al.* (2017). This MME mean climatology (Fig. 9f) has errors of the same order of magnitude as both trained weighted supermodels due to fact that the imperfect model errors are near-mirror images of each other.
In the context of simpler models, *Schevenhoven and Selten* (2017) noted that CPT training of a couple of days duration was sufficient to reduce climatological errors substantially, and this result carries over to the complex SPEEDO model used here. This notion that errors in fast processes contribute substantially to errors in the long-term mean state is also supported by other studies, for example, by *Rodwell and Palmer* (2007). Since the climatological errors are reduced, we expect the trained supermodels to produce better short-term forecasts as compared to the imperfect models.

6.1.2 Forecast quality

In order to assess the quality of short-term forecasts, we initialized the various models from slightly perturbed states of the truth and integrated the models for 2 weeks. We selected 25 initial states, 2 weeks apart, starting 1 January, so the forecasts cover almost 1 year. The quality of the forecast is measured by the RMSE in the global surface air temperature forecast, averaged over the 25 forecasts, and is shown in Fig. 10. In these forecasts, the atmosphere models are forced by the ocean and land conditions of the truth; this is to exclude error growth related to the coupled interactions. As expected, the RMSE in surface air temperature of the perfect model is the one growing the slowest, and it is still as small as about 0.3°C at day 14. On the other hand, the forecast errors of both imperfect models is 0.3°C around day 3 and grow to over 3°C at day 14. Both trained weighted supermodels reach 0.3°C around day 8 and over 1°C at day 14. For comparison we computed the forecast error of the weighted mean forecast of both imperfect models using the same weights as those used in Fig. 9 in the calculation of the optimal climatology. This MME mean forecast has smaller forecast errors than the imperfect models, yet both supermodels are clearly superior.

6.2 Supermodels based on four imperfect models forming a convex hull

As explained in Sect. 5.1, the parameter perturbations of models 1, 3, 4 and 5 form a convex hull around the true parameter values (Table 1). We therefore expect to be able to create a weighted supermodel based on these four models that will be close to the truth, despite the fact that all four have a warmer climatology than the truth (see Table 2). The weights are trained using both CPT and the synch rule in the same way as in the previous case with two imperfect models and are shown in Fig. 11; nevertheless, given that the supermodels are now based on four imperfect models, the number of

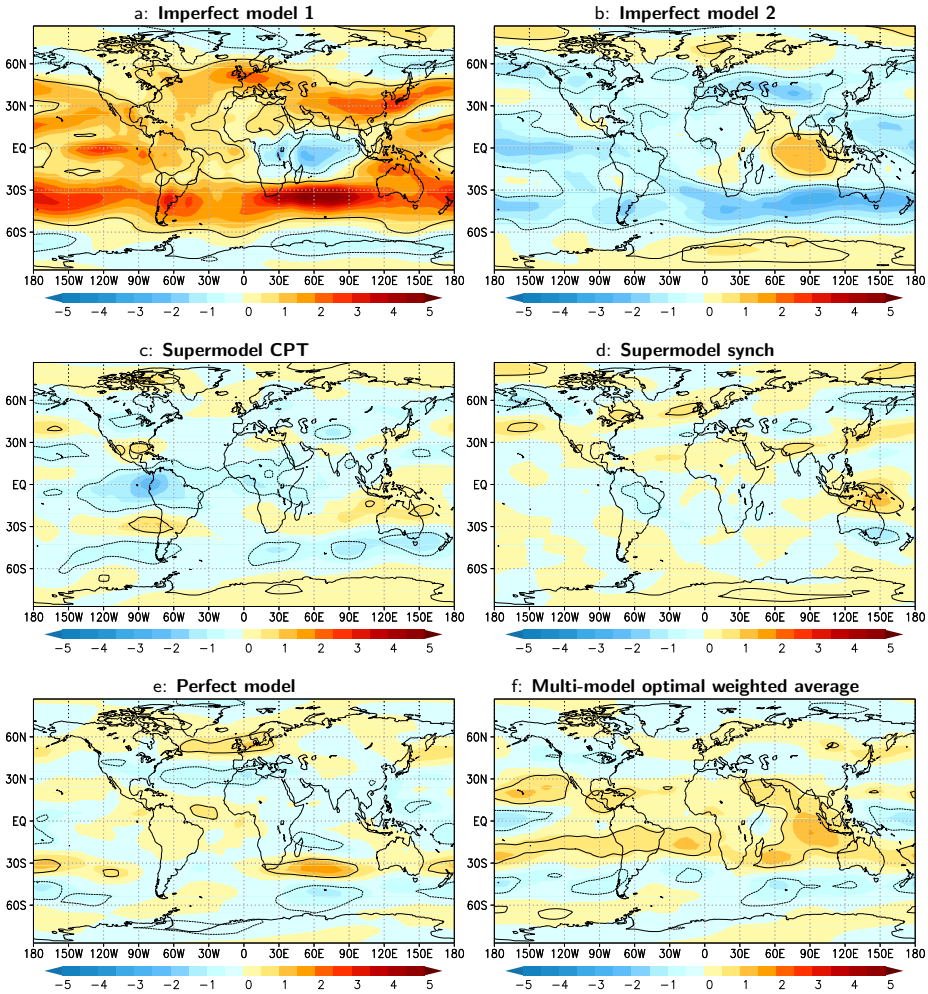


Figure 9: Difference in the zonal wind at 200 hPa averaged over model years 2011-2040 for the various models with respect to the truth. Contours denote areas where the difference is larger than the sampling error at 95% confidence (solid for positive difference, dotted for negative). Positive values imply stronger mean winds blowing eastward. Units: ms^{-1}.

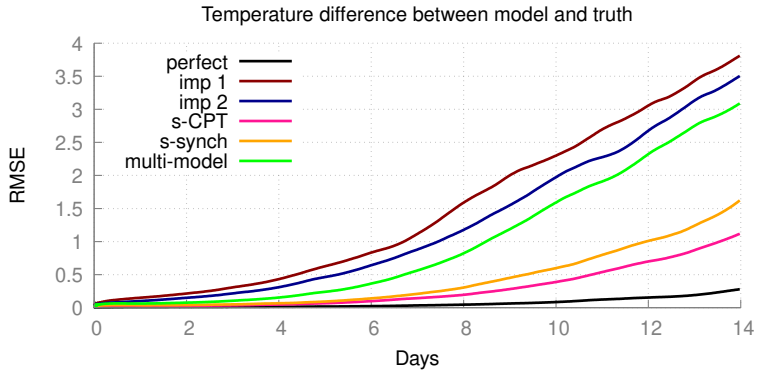


Figure 10: Forecast quality as measured by the root mean squared error (RMSE) of the truth and a model with a perturbed initial condition. The control is the difference between the perfect model and the perfect model with a perturbed initial condition.

weights is doubled. Again the weights during CPT training vary from week to week within 0.05 and converge within a year using the synch rule. Weights for vorticity turn out to be a special case since the change in vorticity as calculated by imperfect models 1 and 3 is equal to, respectively, models 4 and 5. The reason is that only the perturbation in the momentum diffusion timescale affects the vorticity change, and model 1 and 3 have the same diffusion timescale as in models 4 and 5, respectively. Therefore, their weights are equal. Table 4 denotes the final supermodel weights, where for vorticity the weight is equally distributed over models 1 and 4 and model 3 and 5.

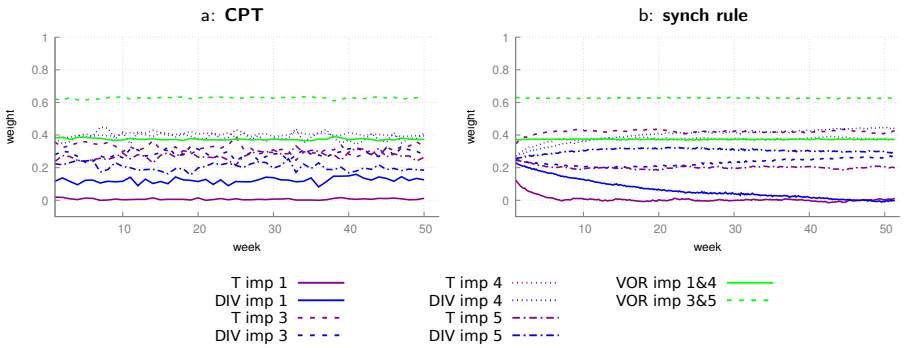


*Figure 11: CPT weights calculated during a training period of 1 week for 1 year (**a**) and the weights for the synch rule for a training period of 1 year (**b**) with four imperfect models.*

The values of the weights for vorticity trained by the synch rule are very close to the values obtained by CPT training. This is not the case for temperature and divergence.

Table 4: Weights for the supermodel trained by CPT and the synch rule. Between brackets, the standard deviation over the year (CPT) or the standard deviation over the last 10 weeks of training (synch rule) is given.

model	method	T	VOR	DIV
model 1	**CPT**	0.01 (0.005)	0.19 (0.022)	0.12 (0.017)
model 3		0.33 (0.030)	0.31 (0.037)	0.28 (0.024)
model 4		0.40 (0.009)	0.19 (0.022)	0.41 (0.028)
model 5		0.26 (0.026)	0.31 (0.040)	0.19 (0.023)
model 1	**synch rule**	0.01 (0.0070)	0.19 (0.0007)	0.00 (0.0064)
model 3		0.42 (0.0072)	0.31 (0.0007)	0.27 (0.0047)
model 4		0.37 (0.0065)	0.19 (0.0007)	0.44 (0.0044)
model 5		0.20 (0.0062)	0.31 (0.0007)	0.29 (0.0029)

For temperature, CPT puts 10% less weight on imperfect model 3 compared to the synch rule and a 10% stronger weight on model 1 for divergence. The synch rule puts (almost) zero weight on imperfect model 1. This is because imperfect model 1 calculates exactly the same vorticity change as imperfect model 4; hence, the synch rule suggests that imperfect model 1 has no added value in the weighted supermodel. Again, the synch rule training yields sum of weights equal to 1 as an optimal solution. Using these weights, we will compare the climatology and forecast skill of both supermodels.

6.2.1 Climate measures and forecast quality

We repeated similar climate integrations as in the case of the supermodels based on two imperfect models and assessed the climatological errors. By comparing the 40-year time-series of global mean values in Fig. 12, both supermodels remain close to the perfect model and are clearly superior to all the imperfect models. Despite all imperfect models becoming too warm and precipitating too much on the global scale, the supermodels balance model deficiencies and produce climate simulations that are close to the truth. Inspection of the RMSE of the 30-year mean fields in the different figure panels indicates that for temperature the supermodel with weights from the CPT training is substantially better than the supermodel with weights from the synch rule. Recall that while imperfect model 1 almost does not contribute to the supermodel with the weights from the synch rule, it does so for the supermodel with the weights from the CPT training. Although imperfect model 1 has larger climatological errors than the other imperfect models (Fig. 12), it nevertheless improves the quality of the CPT supermodel.

This experiment demonstrates the potential of supermodels to mitigate common errors, and thereby clearly outperform the standard MME approach. Since all imperfect mod-

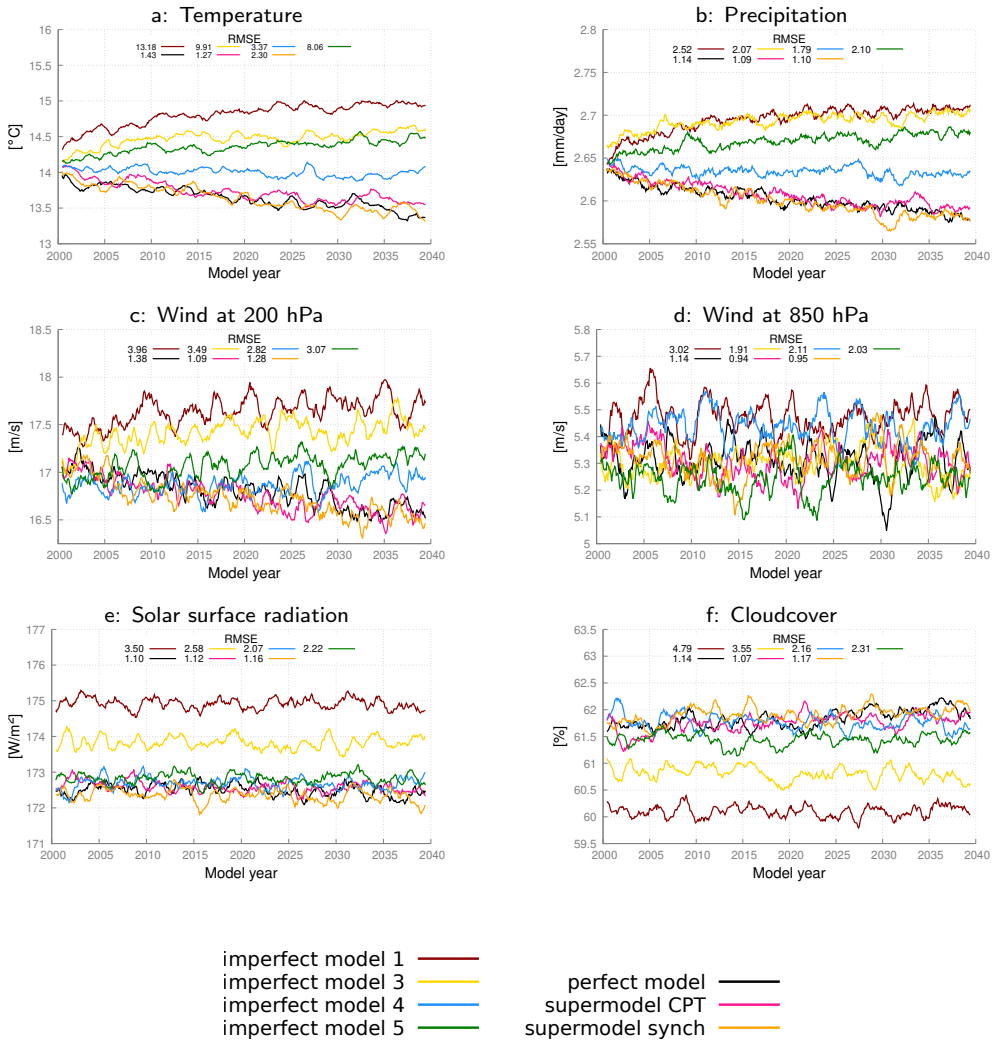


Figure 12: Global mean time series for the truth, the perfect model, the imperfect models and the two supermodels trained by CPT and the synch rule. Included is the RMSE of the model years 2011-2040 with respect to the truth. The normalized RMSE in the climatology of model years 2011-2040 with respect to the climatology of the truth is given in each panel. The normalization is such that the expected value of the perfect model error is 1.

els overestimate the global average temperature and simulate too much precipitation, a standard weighted MME approach results in a climatological forecast worse than the best imperfect model. In the case that the imperfect parameters form a convex hull around the true parameter values, we may expect that a supermodel can be constructed with a climatology much closer to the truth as compared to the best imperfect model. In the case that the imperfect models do not form a convex hull around the true parameter values, allowing negative weights in the weighted supermodel might still improve the climatology and forecast skill. This will be explored in the next section.
We repeated the same forecast experiment as in the case of the supermodel based on two imperfect models. Also, in this case, the supermodels have forecast errors that are substantially reduced as compared to the imperfect models, up to a factor of 3 smaller (not shown). Both supermodels have comparable forecast skill in this measure.

6.3 Negative weights

The CPT training method only produces positive weights, since the weights are defined as being equal to the frequency that the solution of a particular model is closest to the truth during the training period. The synch rule training, on the other hand, does not impose any constraint on the weights. The weights came out positive due to the convex hull principle: the imperfect models considered so far surrounded the truth and with positive weights the effect of the true parameter values can be approximated. But in the event that the imperfect models have parameter values that are all smaller or larger than the truth, only by allowing negative weights one can construct a linear superposition of imperfect models that is closer to the truth. To test if such a supermodel with negative weights indeed shows the desired physical behavior and to test if we can obtain such a model with the synch rule, we construct a weighted supermodel based on two imperfect models (models 1 and 6) with parameter values on the same side of the true parameter values (Table 1).
After a training period of 1 year using the synch rule, stable weights are obtained, which indicates that at least a local minimum is reached. And as expected, the training produces negative weights (Table 5). In contrast to the previous experiments, however, the weights for temperature, divergence and vorticity are quite different. The weights for divergence are positive and do not substantially differ from the weights of the previous experiments. The weights for temperature and vorticity are negative for one of the imperfect models and larger than 1 for the other such that the sum is again close to 1.
Stable climate simulations turn out to be possible with a weighted supermodel using negative weights. The climatology of the supermodel has improved significantly compared to both imperfect models, as displayed in Table. 6. Global mean values of the

Table 5: Weights for the supermodel trained by the synch rule. Between brackets the standard deviation over the last 10 weeks of training is given.

model	T	VOR	DIV
model 1	1.30 (0.016)	2.00 (0.011)	0.40 (0.010)
model 2	-0.30 (0.016)	-1.00 (0.010)	0.60 (0.009)

Table 6: Global mean average difference with the perfect model, calculated over the last 30 years of the simulation.

model	temperature $[C^\circ]$	precipitation $[mm/day]$	solar surface radiation $[W/m^2]$	cloudcover $[\%]$
model 1	1.37	0.11	2.06	-1.59
model 6	3.20	0.26	3.95	-3.37
supermodel	0.64	0.06	1.68	-1.16

various fields are closer to the truth, despite the fact that the global mean climatological errors of both imperfect models have the same sign. Also, local model errors largely have the same sign but are smallest for the supermodel as shown in Fig. 13 for the zonal wind at 200 hPa. Nevertheless, despite the improvement, substantial errors still remain in the supermodel solution.

The forecast errors are evaluated in a similar fashion as in the previous cases and shown in Fig. 14. Although there is a significant improvement in quality for the supermodel as compared to the imperfect models, the forecast error is still quite large. Closer correspondence to the truth can only be expected if all prognostic variables are exchanged, hence also specific humidity and surface pressure, and if the perturbed parameters appear linearly in the equations. Both conditions are not fulfilled in this case.

6.4 Summary of supermodel climate errors

We conclude this section with a summary of the climatological errors of the weighted supermodels of this study and the connected supermodel of *Selten et al.* (2017) in Fig. 15. The climatological errors of the weighted supermodels of this study based on two imperfect models are of the order of the sampling error of the perfect model, whereas the connected supermodel based on the same two imperfect models has substantially larger errors. Also, the weighted supermodel based on the four imperfect models trained by CPT is indistinguishable from the truth with respect to its climatological errors, whereas the synch-rule-trained weighted supermodel has substantially larger errors. These results suggest that CPT training might yield more robust results. The largest climatological errors remain for the supermodel with negative weights.

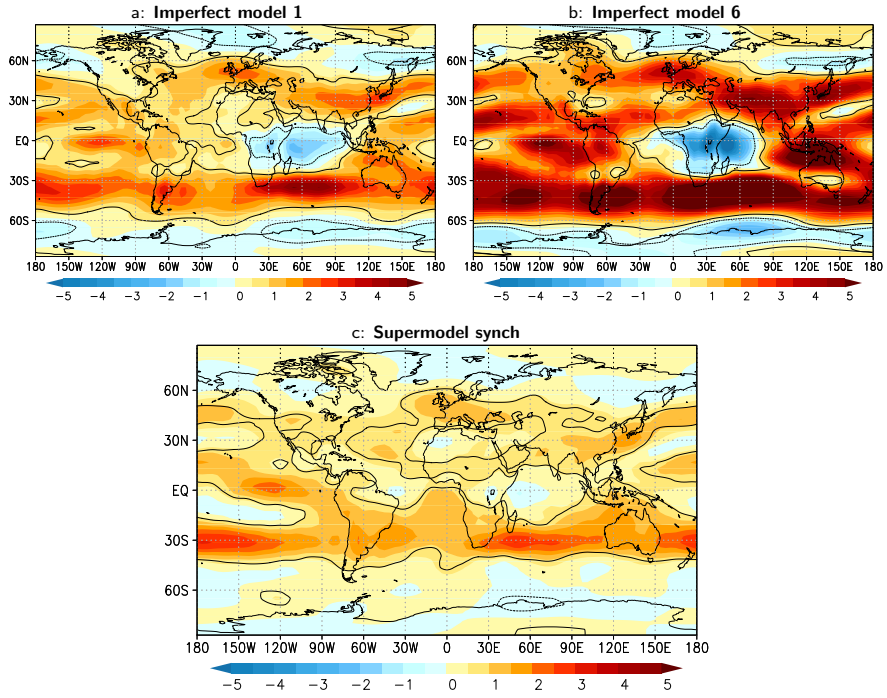


Figure 13: Difference in the east-west component of the wind at the 200 hPa pressure level averaged over model years 2011-2040 for the various models with respect to the truth. Contours denote areas where the difference is larger than the sampling error at 95% confidence (solid for positive difference, dotted for negative). Positive values imply stronger mean winds blowing eastward. Units: ms^{-1}.

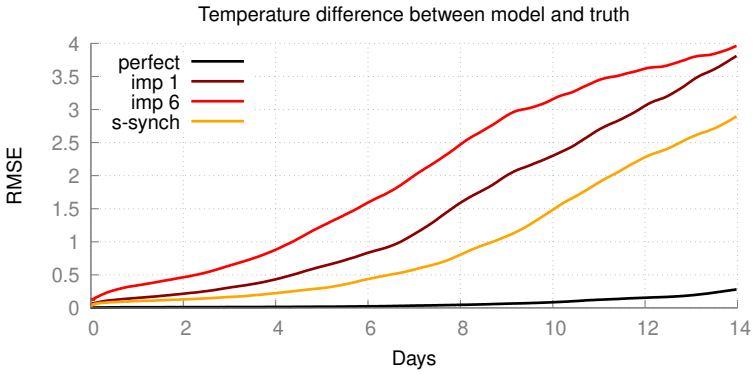


Figure 14: Forecast quality as measured by the RMSE of the truth and a model with a perturbed initial condition. The control is the difference between the perfect model and the perfect model with a perturbed initial condition.

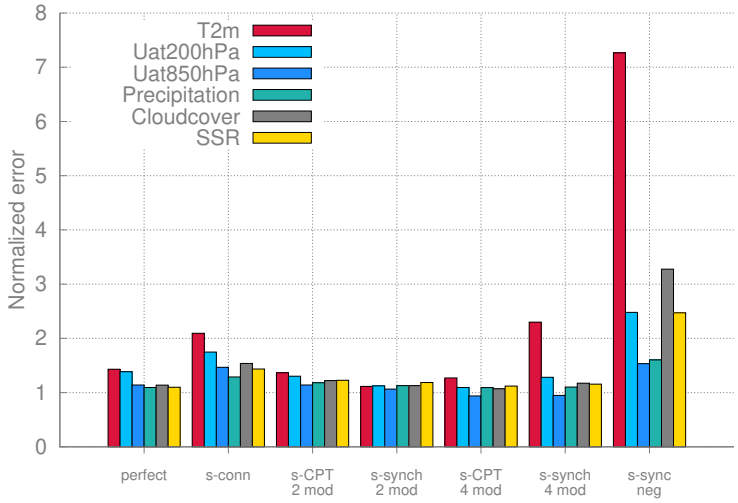


Figure 15: Overview the RMSE of the different supermodels (the connected supermodel of Selten et al., 2017, and the weighted supermodels from the experiments of this paper) over the model years 2011-2040 with respect to the truth.

7 Discussion and conclusions

We have demonstrated the potential of weighted supermodeling to improve weather and climate predictions using the global coupled atmosphere-ocean-land model SPEEDO in the presence of parametric error. Weighted supermodels are constructed based on SPEEDO with perturbed parameters. The perturbations are chosen such that the spread in imperfect models reflects the uncertainty in climate models realistically. The weights are trained using data from the perfect model (i.e., our reference simulated truth) using two different training schemes having low computational cost. The first method is based on CPT, where different model trajectories are "crossed" in order to create a larger ensemble of possible trajectories. The second method is a synchronization-based learning rule (synch rule), which adapts the weights of the different imperfect models during training such that the supermodel synchronizes with the perfect model.

Both training methods yield supermodels that outperform the individual imperfect models, in short-term forecasts as well as in long-term climate simulations. CPT training required shorter training periods (1 week as opposed to a year for the synch rule), but both are much more efficient than cost-function-based approaches that are known to require many climate simulations in an iterative process to reach convergence on optimal weights (*van den Berge et al.*, 2011; *Shen et al.*, 2016). An advantage of the synch

rule is that it allows for negative weights that can potentially improve the weighted supermodel in case model errors do not compensate for positive weights. In addition, CPT requires fairly good models such that mixed trajectories are able to track an observed trajectory for some time. During the synch rule training on the other hand the nudging terms keep the supermodel in the neighborhood of the observed trajectory and is therefore more robust (i.e. less sensitive) with respect to the quality of the imperfect models.

In the application of CPT in this study, we encountered numerical issues due to the partial state replacement. A possible solution is the use of data assimilation techniques to combine state information from different models in a dynamical consistent manner (*Asch et al.*, 2016; *Carrassi et al.*, 2018). One straightforward solution along this line could be based on the idea of *Du and Smith* (2017), in which pseudo-orbit data assimilation is used instead of replacement of the entire state. *Du and Smith* (2017) have already used this approach successfully for low-order dynamical systems. These data assimilation techniques would also allow application of CPT in the event that the different models differ in state representation, for instance, different numerical grids.

The weighted supermodels of this study have smaller climatological errors as compared to the connected supermodel based on the same two imperfect models in *Selten et al.* (2017). Also, in the four-model experiment, the CPT supermodel has substantially better climatology than the supermodel trained by the synch rule. This suggests that synchronization with the truth can be difficult to obtain, especially when the imperfect models that form the supermodel are not fully synchronized in the case of a connected supermodel or when the weighted supermodel consists of several imperfect models. Although it is a common result in synchronization theory that identical systems will synchronize if the nudging strength is strong enough and if there are enough observations from the truth, in practice, this can be a challenge. The issues with synchronized-based learning can be easily demonstrated using a low-order dimensional system (not shown).

In the second supermodel experiment of this paper, the parameter perturbations of four imperfect models were chosen such that they formed a so-called convex hull around the true parameter values. This implies that a linear combination with positive weights of these four models is able to reproduce the model equations with the true parameter values, provided that the parameters appear only linear in the equations. This is not exactly true in this case, but the trained weighted supermodel based on these four models turned out to have a climatology close to the truth. As all four imperfect models have a warmer and wetter climatology than the truth, simply taking the MME mean with positive weights thus does not improve the climatology. This experiment is a clear example

of the potential benefit of the supermodeling approach to ameliorate common model errors. This benefit arises due to the fact that model errors are compensated at an early stage, in the time derivative, and not a posteriori, as in the MME approach where model errors have propagated spatially across the globe, across scales and across the different meteorological fields and other components of the climate system.

In the final supermodel experiment, we have explored the use of negative weights in order to improve predictions in the case that model errors do not compensate; i.e., both imperfect models have parameter perturbations and climatological errors of the same sign. A supermodel trained using the synch rule yielded negative weights. With these weights, stable and credible simulations turn out to be possible and forecast errors as well as climatological errors are reduced with respect to the imperfect models. Substantial errors remain as not all prognostic equations are combined (only temperature, vorticity and divergence, not humidity and surface pressure) and the parameters do not appear linearly in the equations.

Although the synch rule training does not impose that the weights sum to 1, the training inevitably yielded sum of weights equal to 1. An example based on the Lorenz 1963 equations (*Lorenz*, 1963) serves to illustrate why this might be the case. The Lorenz 1963 equations are:

$$\dot{x} = \sigma(y - x) \tag{8a}$$

$$\dot{y} = x(\rho - z) - y \tag{8b}$$

$$\dot{z} = xy - \beta z, \tag{8c}$$

where the standard parameter values are $\sigma = 10, \rho = 28$ and $\beta = \frac{8}{3}$. Assume we have two imperfect models with imperfect parameters ρ_1 and ρ_2. Then $\dot{y}_s = w_1(x(\rho_1 - z) - y) + w_2(x(\rho_2 - z) - y)$, with "s" denoting the supermodel solution. We can rewrite this as: $\dot{y}_s = x(w_1\rho_1 + w_2\rho_2) - (w_1 + w_2)(xz + y)$. To reproduce the standard parameter model solution, two conditions must be satisfied: $(w_1\rho_1 + w_2\rho_2) = \rho$ and $w_1 + w_2 = 1$. Not only should the linear combination of imperfect parameter values match the true parameter value, but also the weights have to sum to 1.

The ultimate goal of our research is to apply supermodeling to realistic climate models. But, will it work? Based on the current results, we believe that this is possible, although the application is not as straightforward as for SPEEDO. First, state-of-the-art models are far bigger and more complex, making their numerical computation a substantial burden. This makes numerical efficiency a key aspect to consider. Second, the real world is not simply a perturbed parameter version of these complex models. In this paper, we have worked under the hypothesis that model error only originates by error in the model parameters in the atmosphere. The imperfect atmosphere models were cou-

pled to the same ocean and land model, which constrains the variability on longer time scales. So far we have demonstrated that the long-term behaviour of the supermodel improves while training only short-term prediction errors. It remains to be seen how much the long-term evolution will improve in the presence of imperfections in the slow components of the climate system. Furthermore, it is essential to extend the approach to other sources of model error towards the application with real climate models. In that case, on top of parametric error, model error can arise from the presence of unresolved scale, numerical discretization or incorrect physics.
Together with the realisms of the models (and of the related model error), those of the observations are also of central importance. In all previous studies with supermodeling, including the current, observations were assumed to be perfect, i.e., to be complete and noise free. To use real data, it will thus be necessary to study the robustness of the supermodeling approach to noisy and unevenly distributed observations and to extend the methods to account for the observational noise. This latter problem is the subject of ongoing research of scientists which are making use of ideas and techniques from data assimilation. Data-assimilation-based supermodeling is also envisioned to account for generic source of model error in the construction of the supermodel, and it will be the subject of future research.

Data availability

No data sets were used in this article.

Author contributions

FSc conceived the study, carried out the research and led the writing of the manuscript. FSe provided codes and technical advice and provided with AC and NK input for the interpretation of the results and the writing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Alberto Carrassi has been funded by the Trond Mohn Foundation under the project no. BFS2018TMT0.

Financial support

This research has been supported by the H2020 European Research Council (grant no. STERCP (648982)).

Review statement

This paper was edited by Andrey Gritsun and reviewed by two anonymous referees.

References

Asch, M., M. Bocquet, and M. Nodet (2016), *Data Assimilation: Methods, Algorithms, and Applications*, Fundamentals of Algorithms, Society for Industrial and Applied Mathematics. 7

Bauer, P., A. Thorpe, and G. Brunet (2015), The quiet revolution of numerical weather prediction, *Nature*, *525*, 47 EP –. 1.1

Carrassi, A., M. Bocquet, L. Bertino, and G. Evensen (2018), Data assimilation in the geosciences: An overview of methods, issues, and perspectives, *Wiley Interdisciplinary Reviews: Climate Change*, *9*(5), e535, doi:10.1002/wcc.535. 1.1, 7

Challinor, A., and T. Wheeler (2008), Crop yield reduction in the tropics under climate change: processes and uncertainties, *Agricultural and Forest Meteorology*, *148*(3), 343 – 356, (c) 2007 Elsevier B.V. All rights reserved. This is an author produced version of a paper published in Agricultural and Forest Meteorology. Uploaded in accordance with the publisher's self-archiving policy. 1.2

Collins, M., and M. R. Allen (2002), Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability, *Journal of Climate*, *15*(21), 3104–3109, doi:10.1175/1520-0442(2002)015<3104:ATRROI>2.0.CO;2. 1.1

Du, H., and L. A. Smith (2017), Multi-model cross-pollination in time, *Physica D: Nonlinear Phenomena*, *353-354*, 31 – 38, doi:https://doi.org/10.1016/j.physd.2017.06.001. 2, 7

Duane, G. S., D. Yu, and L. Kocarev (2007), Identical synchronization, with translation invariance, implies parameter estimation, *Physics Letters A*, *371*(5–6), 416 – 420, doi:http://dx.doi.org/10.1016/j.physleta.2007.06.059. 1.2, 4.2

Goosse, H., and T. Fichefet (1999), Importance of ice-ocean interactions for the global ocean circulation: A model study, *Journal of Geophysical Research: Oceans*, *104*(C10), 23,337–23,355, doi:10.1029/1999JC900215. 3

Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108, doi:10.1175/2009BAMS2607.1. 1.1

IPCC (2013), *Climate Change 2013: The Physical Science Basis*, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel

on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.) Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp]. 1.1, 5.1

Kirtman, B. P., and J. Shukla (2002), Interactive coupled ensemble: A new coupling strategy for cgcms, *Geophysical Research Letters*, *29*(10), 5–1–5–4, doi:10.1029/2002GL014834. 3.1

Lorenz, E. (1963), Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, *20*, 130–140. 7

Mirchev, M., G. S. Duane, W. K. Tang, and L. Kocarev (2012), Improved modeling by coupling imperfect models, *Communications in Nonlinear Science and Numerical Simulation*, *17*(7), 2741 – 2751, doi:http://dx.doi.org/10.1016/j.cnsns.2011.11.003. 1.2

Pecora, L. M., and T. L. Carroll (1990), Synchronization in chaotic systems, *Phys. Rev. Lett.*, *64*, 821–824, doi:10.1103/PhysRevLett.64.821. 2.2, 5.2

Rodwell, M. J., and T. Jung (2008), Understanding the local and global impacts of model physics changes: an aerosol example, *Quarterly Journal of the Royal Meteorological Society*, *134*(635), 1479–1497, doi:10.1002/qj.298. 1.1

Rodwell, M. J., and T. N. Palmer (2007), Using numerical weather prediction to assess climate models, *Q.J.R. Meteorol. Soc.*, (133), 129–146, doi:10.1002/qj.23. 6.1.1

Schevenhoven, F. J., and F. M. Selten (2017), An efficient training scheme for supermodels, *Earth System Dynamics*, *8*(2), 429–438, doi:10.5194/esd-8-429-2017. 1.2, 4, 3, 5.1, 6.1.1

Selten, F. M., F. J. Schevenhoven, and G. S. Duane (2017), Simulating climate with a synchronization-based supermodel, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126,903, doi:10.1063/1.4990721. 1.2, 2.2, 3, 4, 4.2, 6, 5, 5.1, 5.7, 6.1, 6.1.1, 6.1.1, 6.4, 15, 7

Severijns, C. A., and W. Hazeleger (2010), The efficient global primitive equation climate model speedo v2.0, *Geoscientific Model Development*, *3*(1), 105–122, doi:10.5194/gmd-3-105-2010. 1.2, 3

Shen, M.-L., N. Keenlyside, F. Selten, W. Wiegerinck, and G. S. Duane (2016), Dynamically combining climate models to "supermodel" the tropical pacific, *Geophysical Research Letters*, *43*(1), 359–366, doi:10.1002/2015GL066562, 2015GL066562. 1.2, 7

Smith, L. A. (2001), *Nonlinear Dynamics and Statistics*, chap. Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems, pp. 31–64, Mees, Alistair I., Birkhäuser Boston, Boston, MA, doi:10.1007/978-1-4612-0177-9_2. 1.2, 4.1

Sterl, A., H. van den Brink, H. de Vries, R. Haarsma, and E. van Meijgaard (2009), An ensemble study of extreme storm surge related water levels in the north sea in a changing climate, *Ocean Science*, *5*(3), 369–378, doi:10.5194/os-5-369-2009. 1.2

van den Berge, L. A., F. M. Selten, W. Wiegerinck, and G. S. Duane (2011), A multi-model ensemble method that combines imperfect models through learning, *Earth System Dynamics*, *2*(1), 161–177, doi:10.5194/esd-2-161-2011. 1.2, 7

van der Wiel, K., N. Wanders, F. M. Selten, and M. F. P. Bierkens (2019), Added value of large ensemble simulations for assessing extreme river discharge in a 2 ° c warmer world, *Geophysical Research Letters*, *0*(ja), doi:10.1029/2019GL081967. 1.2

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Quarterly Journal of the Royal Meteorological Society*, *134*, 241–260, doi:10.1002/qj.210. 1.1

Wiegerinck, W., and F. M. Selten (2017), Attractor learning in synchronized chaotic systems in the presence of unresolved scales, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126,901, doi:10.1063/1.4990660. 1.2

Wiegerinck, W., M. Mirchev, W. Burgers, and F. Selten (2013), Consensus and synchronization in complex networks, chap. Supermodeling Dynamics and Learning Mechanisms, pp. 227–255, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-642-33359-0_9. 2.2

Bibliography

Afraimovich, V. S., N. N. Verichev, and M. I. Rabinovich (1986), Stochastic synchronization of oscillation in dissipative systems, *Radiophysics and Quantum Electronics*, *29*(9), 795–803, doi:10.1007/BF01034476. 2.2.2

Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard (2014), Enso representation in climate models: from cmip3 to cmip5, *Climate Dynamics*, *42*(7), 1999–2018, doi:10.1007/s00382-013-1783-z. 2.1.2

Carrassi, A., and S. Vannitsem (2016), *Deterministic Treatment of Model Error in Geophysical Data Assimilation*, vol. 15, pp. 175–213, doi:10.1007/978-3-319-39092-5_9. 2.1.2

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A. Weaver, and M. Wehner (2013), *Long-term Climate Change: Projections, Commitments and Irreversibility*, book section 12, p. 1029–1136, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.024. 1.2, 1.1

Du, H., and L. A. Smith (2017), Multi-model cross-pollination in time, *Physica D: Nonlinear Phenomena*, *353-354*, 31 – 38, doi:https://doi.org/10.1016/j.physd.2017.06.001. 3, 4, 5

Duane, G. S., J. J. Tribbia, and J. B. Weiss (2006), Synchronicity in predictive modelling: a new view of data assimilation, *Nonlinear Processes in Geophysics*, *13*(6), 601–612, doi:10.5194/npg-13-601-2006. 2.4.2

Duane, G. S., D. Yu, and L. Kocarev (2007), Identical synchronization, with translation invariance, implies parameter estimation, *Physics Letters A*, *371*(5), 416 – 420, doi:https://doi.org/10.1016/j.physleta.2007.06.059. 2.2.2

Fletcher, R., and C. M. Reeves (1964), Function minimization by conjugate gradients, *The Computer Journal*, *7*(2), 149–154, doi:10.1093/comjnl/7.2.149. 2.5.1

Hiemstra, P. H., N. Fujiwara, F. M. Selten, and J. Kurths (2012), Complete synchronization of chaotic atmospheric models by connecting only a subset of state space, *Nonlinear Processes in Geophysics*, *19*(6), 611–621, doi:10.5194/npg-19-611-2012. 2.4.2

Jungclaus, J. H., N. Keenlyside, M. Botzet, H. Haak, J.-J. Luo, M. Latif, J. Marotzke, U. Mikolajewicz, and E. Roeckner (2006), Ocean circulation and tropical variability in the coupled model echam5/mpi-om, *Journal of Climate*, *19*(16), 3952–3972, doi:10.1175/JCLI3827.1. 2.1.2

Kirtman, B. P., and J. Shukla (2002), Interactive coupled ensemble: A new coupling strategy for cgcms, *Geophysical Research Letters*, *29*(10), 5–1–5–4, doi:10.1029/2002GL014834. 2.1.1, 2.1, 2.1.2, 2.1.3, 2.2.2

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl (2010), Challenges in combining projections from multiple climate models, *Journal of Climate*, *23*(10), 2739–2758, doi:10.1175/2009JCLI3361.1. 1.2

Krishnamurti, T., and J. Sanjay (2003), A new approach to the cumulus parameterization issue, *Tellus A: Dynamic Meteorology and Oceanography*, *55*(4), 275–300, doi:10.3402/tellusa.v55i4.12099. 1.2

Krishnamurti, T. N., V. Kumar, A. Simon, A. Bhardwaj, T. Ghosh, and R. Ross (2016), A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes, *Reviews of Geophysics*, *54*(2), 336–377, doi:10.1002/2015RG000513. 1.2

Liu, J., and E. Kalnay (2005), 3.4 assimilating specific humidity observations with local ensemble transform kalman filter. 2.4.1

Lloyd, J., E. Guilyardi, and H. Weller (2011), The role of atmosphere feedbacks during enso in the cmip3 models. part ii: using amip runs to understand the heat flux feedback mechanisms, *Climate Dynamics*, *37*(7), 1271–1292, doi:10.1007/s00382-010-0895-y. 2.1.2

Lorenz, E. (1963), Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, *20*, 130–140. 2.2.2

Marshall, J., and F. Molteni (1993), Toward a dynamical understanding of planetary-scale flow regimes, *Journal of the Atmospheric Sciences*, *50*(12), 1792–1818. 2.4.2, 2.5.2

Mirchev, M., G. S. Duane, W. K. Tang, and L. Kocarev (2012), Improved modeling by coupling imperfect models, *Communications in Nonlinear Science and Numerical Simulation*, *17*(7), 2741 – 2751, doi:http://dx.doi.org/10.1016/j.cnsns.2011.11.003. 2.5.1

Nelder, J. A., and R. Mead (1965), A Simplex Method for Function Minimization, *The Computer Journal*, *7*(4), 308–313, doi:10.1093/comjnl/7.4.308. 2.1.2, 2.5.1

Nordeng, T.-E. (1994), Extended versions of the convective parametrization scheme at ecmwf and their impact on the mean and transient activity of the model in the tropics, (206), 41, doi:10.21957/e34xwhysw. 2.1.2

Pecora, L. M., and T. L. Carroll (1990), Synchronization in chaotic systems, *Phys. Rev. Lett.*, *64*, 821–824, doi:10.1103/PhysRevLett.64.821. 2.3.1

Pecora, L. M., T. L. Carroll, G. A. Johnson, D. J. Mar, and J. F. Heagy (1997), Fundamentals of synchronization in chaotic systems, concepts, and applications., *Chaos*, *7 4*, 520–543. 2.2.2

Richter, I., T. Doi, S. K. Behera, and N. Keenlyside (2018), On the link between mean state biases and prediction skill in the tropics: an atmospheric perspective, *CLIMATE DYNAMICS*, *50*(9-10), 3355–3374, doi:10.1007/s00382-017-3809-4. 2.1.1

Rodwell, M. J., and T. N. Palmer (2007), Using numerical weather prediction to assess climate models, *Q.J.R. Meteorol. Soc.*, (133), 129–146, doi:10.1002/qj.23. 5

Scaife, A. A., and D. Smith (2018), A signal-to-noise paradox in climate science, *npj Climate and Atmospheric Science*, *1*(1), 28, doi:10.1038/s41612-018-0038-4. 2.1.1

Severijns, C. A., and W. Hazeleger (2010), The efficient global primitive equation climate model speedo v2.0, *Geoscientific Model Development*, *3*(1), 105–122, doi:10.5194/gmd-3-105-2010. 2.4.1

Shen, M.-L., N. Keenlyside, F. Selten, W. Wiegerinck, and G. S. Duane (2016), Dynamically combining climate models to "supermodel" the tropical pacific, *Geophysical Research Letters*, *43*(1), 359–366, doi:10.1002/2015GL066562, 2015GL066562. 2.1.2, 2.1.3, 2.2.1, 2.3, 2.2.2, 2.5, 2.5.1

Shen, M.-L., N. Keenlyside, B. C. Bhatt, and G. S. Duane (2017), Role of atmosphere-ocean interactions in supermodeling the tropical pacific climate, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126,704, doi:10.1063/1.4990713. 2.2, 2.1.2

Smith, L. A. (2001), *Nonlinear Dynamics and Statistics*, chap. Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems, pp. 31–64, Mees, Alistair I., Birkhäuser Boston, Boston, MA, doi:10.1007/978-1-4612-0177-9_2. 3

Szendro, I. G., M. A. Rodríguez, and J. M. López (2009), On the problem of data assimilation by means of synchronization, *Journal of Geophysical Research: Atmospheres*, *114*(D20), doi:10.1029/2009JD012411. 2.2.2, 2.3.1

Tiedtke, M. (1989), A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Monthly Weather Review*, *117*(8), 1779–1800, doi:10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2. 2.1.2

Toth, Z., and E. Kalnay (1997), Ensemble forecasting at ncep and the breeding method, *Monthly Weather Review*, *125*(12), 3297–3319, doi:10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2. 2.4.2

van den Berge, L. A., F. M. Selten, W. Wiegerinck, and G. S. Duane (2011), A multi-model ensemble method that combines imperfect models through learning, *Earth System Dynamics*, *2*(1), 161–177, doi:10.5194/esd-2-161-2011. 2.5.1

Vannitsem, S. (2017), Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(3), 032,101, doi:10.1063/1.4979042. 2.2.2

Weber, R. J. T., A. Carrassi, and F. J. Doblas-Reyes (2015), Linking the Anomaly Initialization Approach to the Mapping Paradigm: A Proof-of-Concept Study, *Monthly Weather Review*, *143*(11), 4695–4713, doi:10.1175/MWR-D-14-00398.1. 2.2.2

Weigel, A. P., M. A. Liniger, and C. Appenzeller (2008), Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Quarterly Journal of the Royal Meteorological Society*, *134*(630), 241–260, doi:10.1002/qj.210. 1.2

Wiegerinck, W., and L. Basnarkov (2013), Attractor learning in interactive ensembles, *The Computer Journal*, *15*, 14,085. 2.5.2

Wiegerinck, W., and F. M. Selten (2017), Attractor learning in synchronized chaotic systems in the presence of unresolved scales, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *27*(12), 126,901, doi:10.1063/1.4990660. 2.5.2

Wiegerinck, W., W. Burgers, and F. Selten (2013), On the limit of large couplings and weighted averaged dynamics, *Understanding Complex Systems*, pp. 257–275, doi:10.1007/978-3-642-33359-0-10. 2.3.3, 2.3.4

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230852743 (print)
9788230852422 (PDF)