Supermodeling: Combining Imperfect Models through Learning

W. Wiegerinck
SNN, DCN, Radboud University
Nijmegen, The Netherlands
w.wiegerinck@science.ru.nl

F. M. Selten
Royal Netherlands Meteorological Institute
De Bilt, The Netherlands
selten@knmi.nl

Abstract

At a dozen or so institutes around the world, comprehensive climate models are being developed and improved. Each model provides reasonable simulations of the observed climate, each with its own strengths and weaknesses. In the current multi-model ensemble approach model simulations are combined a posteriori. As an alternative, [1] proposed to dynamically connect the models into one supermodel. The connection coefficients are learned from historical observations. The method is developed and tested on small chaotic dynamical systems, like the Lorenz 63 system. In this paper, we further analyze this supermodeling approach and its results. This leads us to a proposal for a more scalable variant of supermodeling, and we demonstrate this variant on a T5 quasi-geostrophic baroclinic model with 30 degrees of freedom.

1 Introduction

There is a broad scientific consensus that our climate is warming due to anthropogenic emissions of greenhouse gasses [2]. Due to the large impacts of climate change on society there is a growing need to widely sample and assess the possible climate change related to the plausible scenarios for future emissions. At about a dozen climate institutes around the world complex climate models have been developed over the past decades. Despite the improvements in the quality of the model simulations, the models are still far from perfect. For instance a temperature bias of several degrees in annual mean temperatures in large regions of the globe is not uncommon in the simulations of the present climate [2].

Nevertheless these models are used to simulate the response of the climate system to future emission scenarios of greenhouse gases. It turns out that the models differ substantially in their simulation of the response: the global mean temperature rise varies by as much as a factor of 2 and on regional scales the response can be reversed, e.g. decreased precipitation instead of an increase. It is not clear how to combine these outcomes to obtain the most realistic response. The standard approach is to take some form of a weighted average of the individual outcomes [3].

In [1] it has been proposed to combine the strength of the individual models by letting them exchange information during the simulation instead of combining the results afterwards. This proposal is motivated from a study [4] in which two different atmospheric models were coupled to one ocean model. It turned out that the most realistic simulation in terms of the annual mean, annual cycle and interannual variability of sea surface temperatures over the tropical pacific was obtained by coupling the momentum fluxes from one model and the heat and fresh water fluxes from the other to the ocean model [4].

It should be noted that in this context we are interested in climate models. Assuming that the weather system is deterministic chaotic, the climate can be defined as the attractor of the weather system, whereas weather prediction is short term prediction on the basis of current observations. The goal of the climate model is to find statistical properties of the attractor, rather than day-to-day predictions.

Another assumption is that the climate system is so complex (e.g. current climate models have $\mathcal{O}(10^5)$ state variables) that learning a free form model from 'scratch' is infeasible. The idea is to rather use the 'state-of-the-art' (but still imperfect) climate models acting as a kind of building blocks for a supermodel. The parameters, which determine how the models are dynamically combined into a supermodel, are then to be tuned on the basis of data.

In this paper we will further build on the work and ideas in [1]. We first provide a review in section 2. Our first contribution is in the section 3 where we analyze the findings in [1], and explain its empirical success and propose a simplification for supermodeling. In section 4, we apply the method to a T5 spectral truncated quasi-geostrophic baroclinic model with 30 degrees of freedom. Then we end with a short discussion.

2 Supermodeling by coupling connections

In this section we provide a review of the findings in [1]. Assume that we have M good, but imperfect models labeled by μ that describe the dynamics of the model state vector $\boldsymbol{x}_\mu$ according to

$$\dot{x}^i_\mu = f^i_\mu(\boldsymbol{x}_\mu) \tag{1}$$

in which i labels the vector components, and dot-notation is used for time derivatives. Then, the proposal is to combine the individual models μ into one supermodel by prescribing connections between the model equations,

$$\dot{x}^i_\mu = f^i_\mu(\boldsymbol{x}_\mu) + \sum_\nu C^i_{\mu\nu}(x^i_\nu - x^i_\mu)\,. \tag{2}$$

The argument is that with sufficient connectivity, the individual models will synchronize and get a kind of consensus with each other [5]. The connection coefficients $C^i_{\mu\nu}$ are to be inferred from a training set of historical observations. The solution of the supermodel is defined to be the average of the coupled imperfect models, $\boldsymbol{x}(t) \equiv \frac{1}{M}\sum_\mu \boldsymbol{x}_\mu(t)$. The goal of the supermodel is to do climate simulation, i.e. the supermodel should converge an attractor which reflects the climate attractor. However, direct attractor learning is difficult. Therefore a proxy cost function has been proposed, which basically consists of a sum of short term prediction errors (see e.g. [6] for a discussion attractor learning). The cost function is then to be optimized using standard numerical optimization tools. To demonstrate the supermodeling approach, a number of simple chaotic systems has been studied such as the Lorenz 63 model [7]. The equations for the Lorenz 63 model are

$$\dot{x} = \sigma(y - x)\,, \quad \dot{y} = x(\rho - z) - y\,, \quad \dot{z} = xy - \beta z\ . \tag{3}$$

This model is often used as a metaphore for the atmosphere, because of its regime changes and unstable nature. The model with standard parameter values ($\sigma = 10$, $\rho = 28$, $\beta = 8/3$) is used as ground truth. Imperfect models are assumed to be in the same model class with perturbed parameter values. These imperfect models are connected and combined into a supermodel. Training data is generated from the assumed ground truth and used to tune the connections in the supermodel. By inspecting plots of the attractor, as well as by considering means, variances, covariances and autocorrelations, it is concluded that the supermodeling approach is a promising modeling approach in the case of complex modeling where good, but still imperfect models are available and a machine learning method starting from scratch is infeasible.

3 Supermodeling by convex combinations of models

To understand why this approach worked so well, we now further analyze the dynamics of the supermodel. We assume that the couplings $C^i_{\mu\nu}$ are large so that the attractive forces between models dominate the nonlinear vector fields $\boldsymbol{f}_\mu$. It can then be argued that the supermodel will be in strong synchronization,

$$x^i_\mu(t) \approx x^i_\nu(t) \tag{4}$$

since any substantial deviation from synchronization will be rapidly equilibrated. To analyze this further, we consider the limit $C \to \infty$. To proceed we rewrite the supermodel equations a bit,

$$\dot{x}^i_\mu = f^i_\mu(\boldsymbol{x}_\mu) + \sum_\nu W^i_{\mu\nu} x^i_\nu \quad \text{where} \quad W^i_{\mu\nu} = C^i_{\mu\nu} - \delta_{\mu\nu} \sum_\kappa C^i_{\mu\kappa} \ . \tag{5}$$

Now it can be seen that the matrix W^i is a "transition rate matrix", known from stochastic processes. That is to say, with such a W^i the equation $\dot{P}_\nu = \sum_\mu P_\mu W^i_{\mu\nu}$ is a continuous time Markov process. If W^i is mixing then P converges to an equilibrium distribution, which is the normalized left eigenvector $\vec{q}^{\,i}$ of W^i with eigenvalue 0. Normalization means $\sum_\mu q^i_\mu = 1$. Its right eigenvector is the vector with all components equal, which can be interpreted as a fully synchronized state. Now if W^i is mixing then the other eigenvalues have a negative real part, which means that the other modes will vanish and states of the different models will synchronize into a joint state, $x^i_\mu(t) = x^i_\nu(t) = x^i(t)$. By multiplying the supermodel (5) from the left by $\vec{q}^{\,i}$, we obtain the synchronized state dynamics

$$\dot{x}^i = \sum_\mu q^i_\mu f^i_\mu(\boldsymbol{x}) \ . \tag{6}$$

This equation states that the supermodel dynamics is basically described by a vector field of which the components are convex combinations of the imperfect model components.

From this result, one can directly predict when the coupled models are able to reproduce the ground truth. In [1] both the ground truth system and the perturbed imperfect systems are linear in the parameters (i.e. σ, ρ, and β in (3), but also the other models described in [1]). So, if we take the convex combination of the imperfect model equations, we obtain a model that is again in that model class, with parameters that are convex combinations of the perturbed parameters. For the C's that have been obtained for the Lorenz 93 models as reported in [1], we noted that some of its components are quite large. We computed the left-eigenvectors and found that the resulting convex combinations of the perturbed parameters were close to the ones of the assumed ground truth. Also the observation that the supermodel followed the ground truth when one of the parameters was doubled in both the ground truth and imperfect models can be understood from this analysis.

Since supermodels with large couplings behave like supermodels of convex combinations, why not short-cut this and model supermodels directly as convex combinations. Such a supermodel is actually an instance of an ensemble of models as in e.g. [8]. So methods for ensemble methods can be directly applied to optimize the parameters, e.g. the following simple and scalable optimization scheme: do for each imperfect model one run of short term predictions starting at different points in the training set to obtain a set of error-vectors. The sum-of-squared errors can then be optimized by quadratic programming.

4 T5 quasi-geostrophic baroclinic model

An important instability mechanism that leads to the growth and decay of weather structures (depressions, low-pressure systems, storms etc.) at mid-latitudes is the baroclinic instability mechanism, a process in which the available potential energy stored in the equator to pole temperature gradient is converted into the kinetic energy of the storms. These storms in turn transport heat polewards, thereby reducing the temperature gradient and the cause for their existence. The temperature gradient is continuously restored by the differential heating of the earth by the incoming solar radiation and the outgoing thermal radiation. To model this instability mechanism [9] developed a two-level, hemispheric, quasi-geostrophic spectral model on the sphere, triangularly truncated at wave number five. The resulting nonlinear ODE with 30 degrees of freedom exhibits chaotic behavior. To give an impression of the complexity of the model, we plotted a state represented as two-level fields in Fig 1.

We took this model as the ground truth and assumed two imperfect models by perturbing parameters. Supermodeling is by convex combination and optimization is by quadratic programming. The optimization of supermodel with coupling coefficients C via the cost function as in [1] would require orders of magnitudes more simulations and has not been pursued. Results are presented in Fig 1. The supermodel mimics the ground truth much better than the imperfect models.

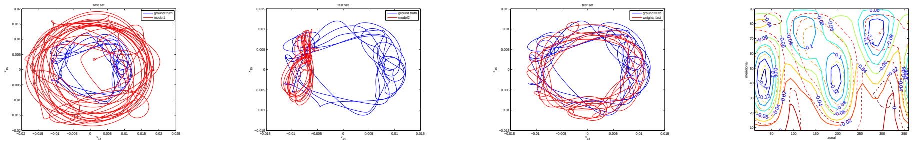

Figure 1: Three left plots: results of model integration projected on (x_{14}, x_{15}) plane. Blue: assumed ground truth. Red: assumed imperfect models (two left ones) and super model (third left one). Right plot: state of the T5 quasi-geostrophic barotropic model, represented through a contour plot of the two-level fields.

5 Discussion

We considered supermodeling as a potential tool for climate research as proposed in [1]. Its aim is to do modeling using data by combining existing good, but imperfect models to a so-called supermodel, since learning a free-form model from scratch seems infeasible for this domain. The coupling method proposed in [1] relies on a synchronization between the models. We argued that in the limit of large coupling and perfect synchronization, the resulting supermodel reduces to a convex combination of imperfect model components. This motivated us to use convex combinations of models from the start, since in practice, the use of convex combinations of models seems to have several advantages, most importantly a scalable learning scheme, which is already apparent in simulations of a 30 dimensional quasi-geostrophic baroclinic model. Other advantages are interpretability and performance guarantees (see e.g. ensemble methods in [8]).

The convex combination approach can be interpreted as a "hard coupling" ($C \to \infty$) limit of the coupling approach in [1]. A disadvantage could be that a hard coupled supermodel has less flexibility than a softly coupled supermodel, which may hinder transitions between regimes in the attractor. This is to be investigated further. Another issue is if supermodeling does not overfit in systems of high dimension. To investigate this, [1] proposed to study supermodeling performances in systems of increasing complexity.

Acknowledgments This work has been partially supported by FP7 FET Open Grant # 266722 (SUMO project). See `www.sumoproject.eu`

References

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

[2] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, KB Averyt, M. Tignor, and HL Miller. Ipcc, 2007: Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change, 2007.

[3] C. Tebaldi and R. Knutti. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857):2053, 2007.

[4] B. Kirtman, D. Min, P. Schopf, and Schneider E. A new approach for coupled gcm sensitivity studies. Technical Report 154, COLA, 2003.

[5] L.M. Pecora and T.L. Carroll. Synchronization in chaotic systems. *Physical review letters*, 64(8):821–824, 1990.

[6] R. Bakker, J.C. Schouten, C.L. Giles, F. Takens, and C.M. Bleek. Learning chaotic attractors by neural networks. *Neural Computation*, 12(10):2355–2383, 2000.

[7] E.N. Lorenz. Deterministic nonperiodic flow. *Atmos J Sci*, 20:130–141, 1963.

[8] C. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[9] P.L. Houtekamer. Variation of the predictability in a low-order spectral model of the atmospheric circulation. *Tellus A*, 43(3):177–190, 1991.