Efficient training schemes for supermodels

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

[1]Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, [2]Royal Netherlands Meteorological Institute, [3]Nansen Environmental and Remote Sensing Center

Although weather and climate predictions improve over time, models remain imperfect. Predictions might be improved if models are combined dynamically to produce solutions that are unique to the combined system. The supermodel is potentially closer to observations than the standard Multi Model Ensemble (MME) method. Errors can be corrected in an earlier stage and since the models in a supermodel are synchronized we do not suffer from variance reduction and smoothing.

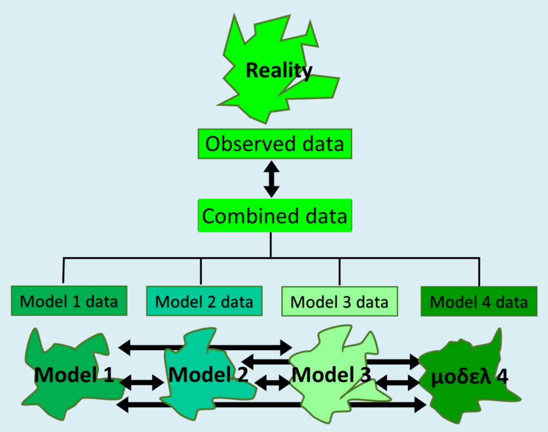


Figure 1: an interacting ensemble of imperfect models

Weighted supermodeling

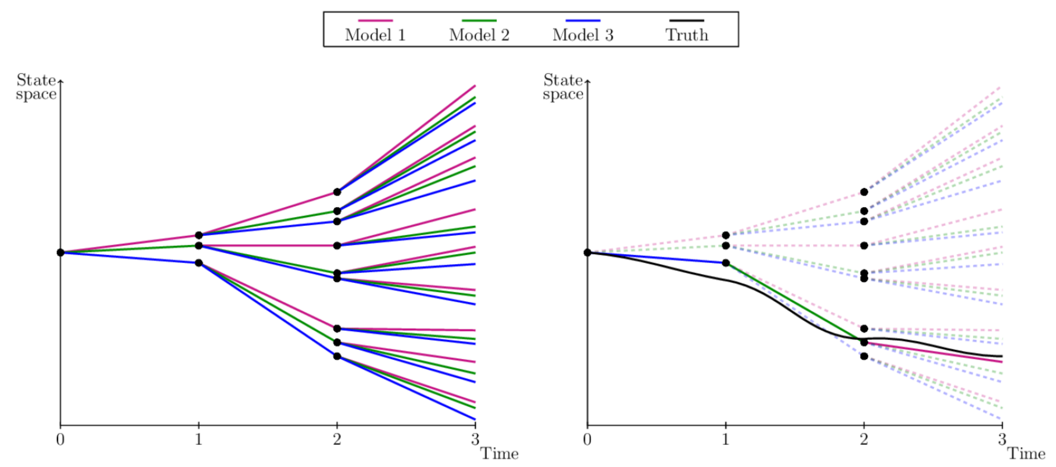
Consider two imperfect models with parametric error, with s denoting the supermodel solution.

$$\begin{aligned}\dot{x}_1 &= \mathbf{f}(\mathbf{x}_1, \mathbf{p}_1)\\ \dot{x}_2 &= \mathbf{f}(\mathbf{x}_2, \mathbf{p}_2)\\ \dot{x}_s &= \mathbf{W}_1\mathbf{f}(\mathbf{x}_s, \mathbf{p}_1) + \mathbf{W}_2\mathbf{f}(\mathbf{x}_s, \mathbf{p}_2),\end{aligned}$$

Aim in this study: learning the optimal weights $\mathbf{W}$.

Learning method 1: *Cross Pollination in Time (CPT)*

It is assumed 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 a certain period each ending in a different state. From these endpoints all models run again. For training a supermodel only those predictions that remain closest to the truth are continued, the others are discarded.



CPT in general. CPT in the context of supermodeling.

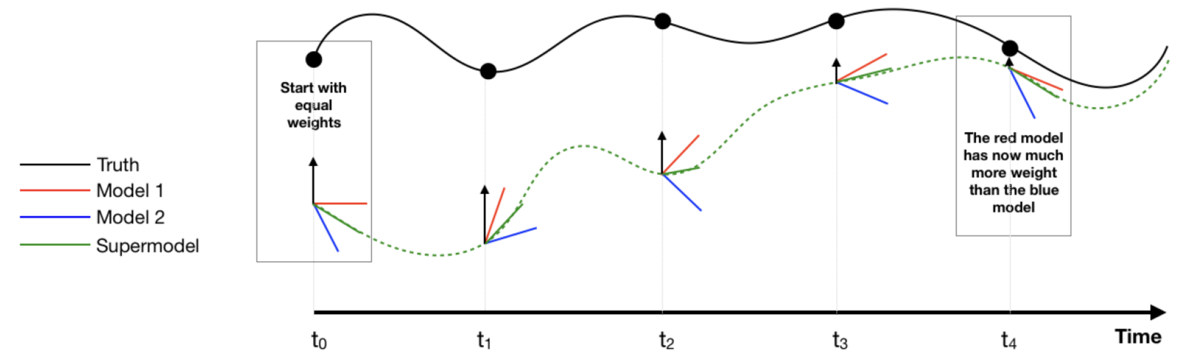
A training phase of a sufficient amount of time steps gives the frequency for each model that its prediction for a variable remains closest to the truth. These frequencies determine the weights in the supermodel.

Learning method 2: *synchronization based learning*

The synch rule updates the weights such that synchronization errors between truth and supermodel are minimized. In contrast to CPT learning, initial values for the supermodel are not needed, in contrast to CPT learning. In contrast to CPT learning, initial values for the weights need to be chosen and weights are updated during training. The synch rule is an application of a general synchronization based parameter estimation method.

- $\mathbf{w}_{kj}$ the weight for model k corresponding to variable j
- $\mathbf{e}$ is the difference between the supermodel and the truth
- $\mathbf{f}_k$ is the time derivative of imperfect model k.

$$\dot{\mathbf{w}}_{kj} = -\delta_j \mathbf{e}\mathbf{f}_{kj}$$



Results for the SPEEDO model

SPEEDO characteristics:

- Spectral atmosphere model with over 30000 degrees of freedom
- Land model with over 6000 degrees of freedom
- Ocean model with primitive equations with free surface and over 200.000 degrees of freedom

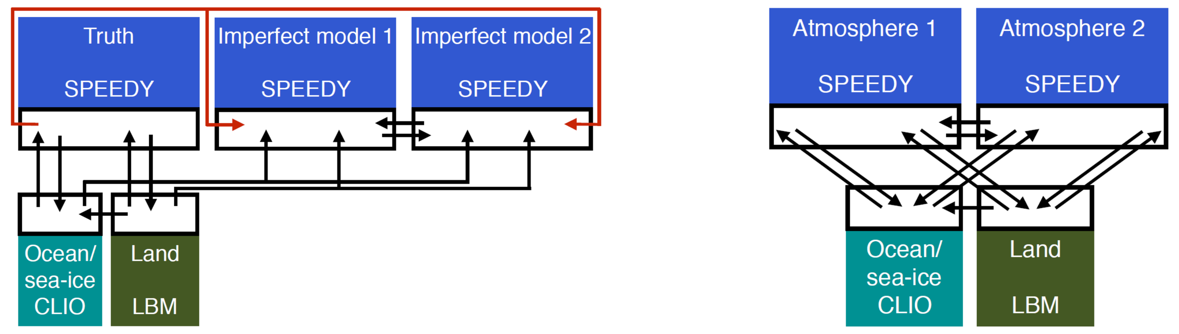


Figure 4: Interactive SPEEDO during training (left) and the SPEEDO supermodel (right)

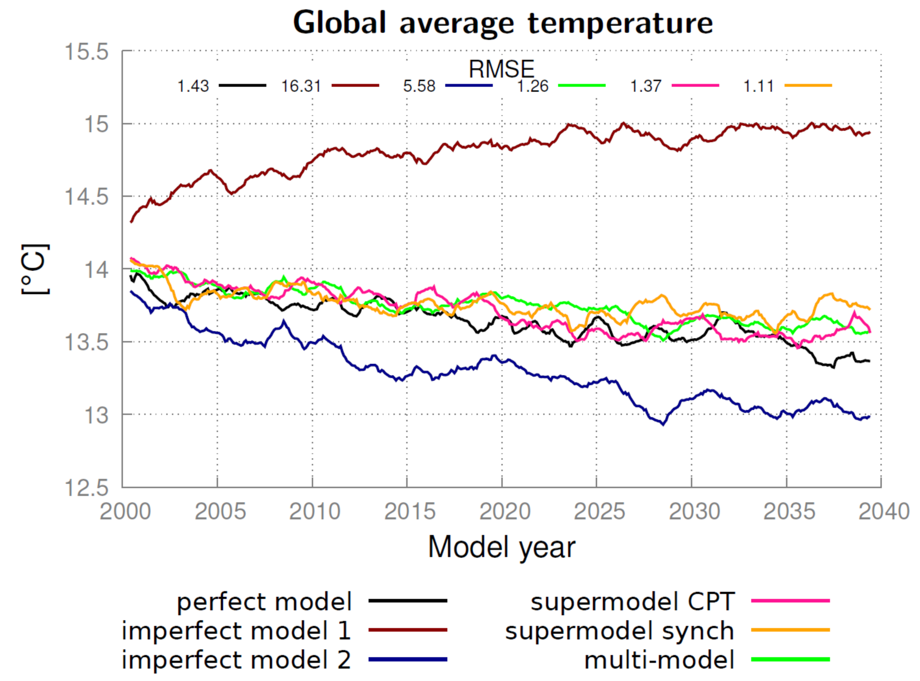
Parameter perturbations create different models.

SPEEDO experiment 1

	Truth	Model 1	Model 2
Relaxation timesale of convection (RTC)	6 hours	4 hours	8 hours
Relative humidity threshold (RHT)	0.9	0.85	0.95
Momentum diffusion timescale (MDT)	24 hours	18 hours	30 hours

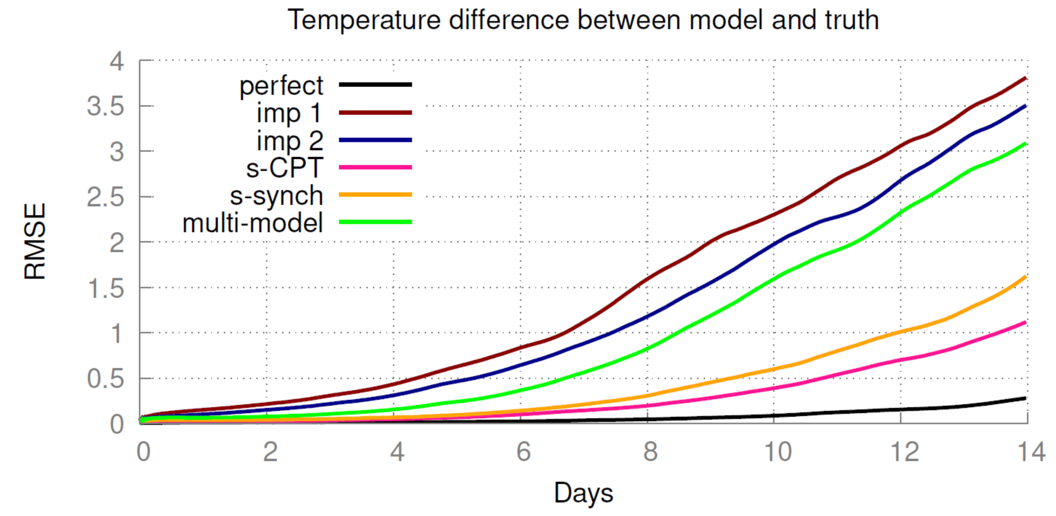
- Training period T=1 week x 52 for CPT and T = 1 year for the synch rule.
- Global weights: every grid point of the model obtains the same weights.
- Different weights are calculated for temperature, vorticity and divergence.

Long term forecast quality



- The supermodels outperform the individual models, but the multi-model mean is as good as the supermodels.

Short term forecast quality



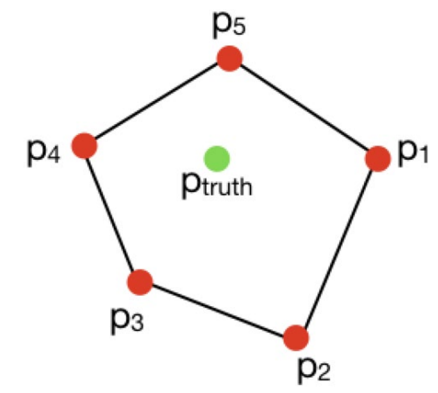
- Both supermodels are better than the individual imperfect models but the multi-model trajectory is worse than the supermodels.

Results for the SPEEDO model

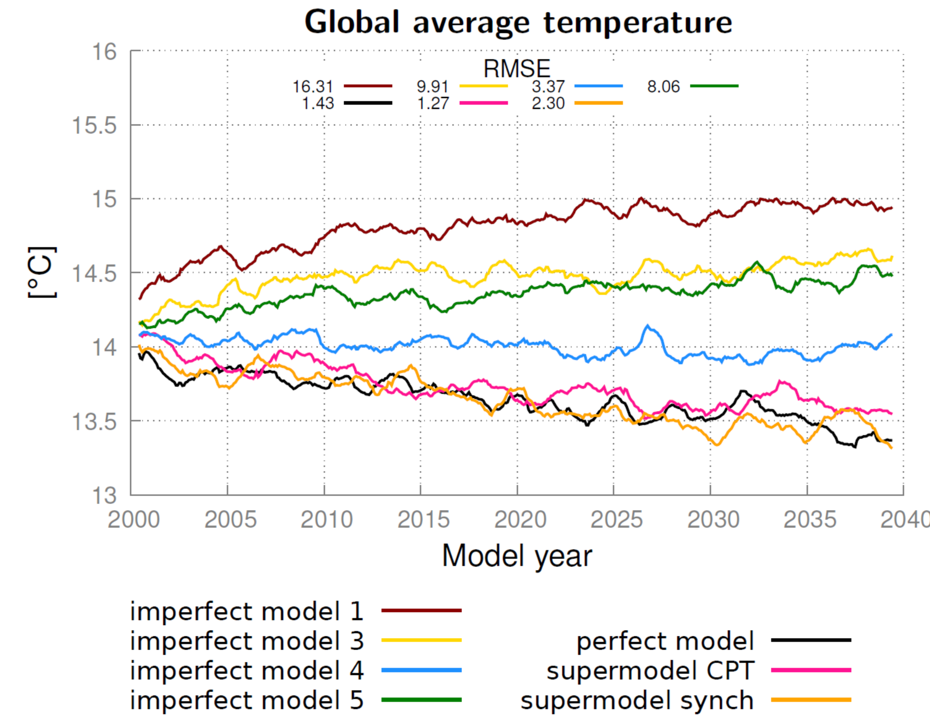
SPEEDO experiment 2: Convex hull approach

	Truth	Model 1	Model 3	Model 4	Model 5
RTC	6 hours	4 hours	4 hours	8 hours	8 hours
RHT	0.9	0.85	0.95	0.95	0.85
MDT	24 hours	18 hours	30 hours	18 hours	30 hours

Convex hull: if the parameters appear linear in the equations, a perfect linear combination with positive weights of the imperfect parameter values $\mathbf{p}_i$ (red dots) can be found that is equal to the true parameter value (green dot).



Long term forecast quality



In this experiment all imperfect models overestimate the global average temperature. Hence a multi-model mean with positive weights can never outperform the best imperfect model. The supermodels however are almost indistinguishable from the truth.

Discussion and conclusion

To combine the strengths of imperfect weather and climate models a new learning approach of cross pollinating trajectories is explored as well as a synchronization based learning rule. Both methods create supermodels that are trained to outperform the individual models as well as the multi-model mean, in short-term forecasts as well as in long-term climate simulations.

The ultimate goal of our research is to apply supermodeling to realistic climate models. But, will it work? In principle yes, but:

- No perfect data ⟹ use of data assimilation.
- Not parametric error only, state-of-the-art models can also differ structurally ⟹ define a common state space.

Acknowledgements

This work is partially supported by the EU project 648982 – STERCP.

References

Smith, L. A. (2001). Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems. In A. I. Mees (ed.), *Nonlinear Dynamics and Statistics* (pp. 31-64). Birkhäuser Boston.

Selten, F.M., Schevenhoven, F.J., and Duane, G.S: Simulating climate with a synchronization-based supermodel, Chaos: An interdisciplinary Journal of Nonlinear Science, 27, 126903, 2017.