Training of supermodels
in the context of weather and climate forecasting

Francine Schevenhoven[1,2], and Alberto Carrassi[3,4]

[1]Geophysical Institute, University of Bergen, Norway
[2]Bjerknes Centre for Climate Research, Bergen, Norway
[3]University of Reading, Reading, United Kingdom
[4]Utrecht University, Utrecht, the Netherlands

francine.schevenhoven@uib.no

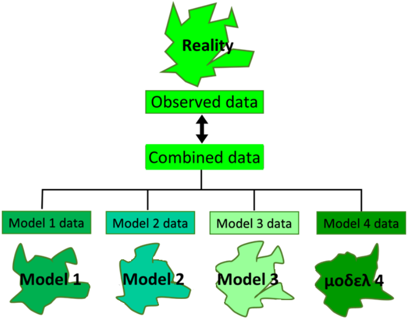
Workshop: Multi-annual to Decadal Climate Predictability in the North Atlantic-Arctic Sector

BJERKNES CENTRE for Climate Research

Standard Multi Model Ensemble (MME)

Standard Multi Model Ensemble (MME): *combining output* ensemble.

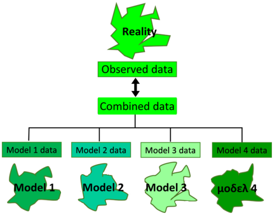
- MME can give *improved statistics*, like mean and variance, but it does not ensure an improved trajectory.
- This since averaging *uncorrelated climate trajectories* leads to *variance reduction* and smoothing.
- We think we can do better!



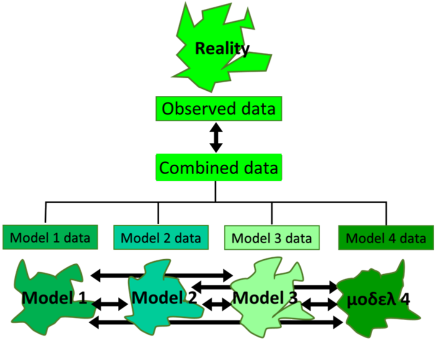
MME

Supermodeling

- To improve predictions we propose a **supermodel**: an optimal *dynamical combination of imperfect models*.
- Within a supermodel new dynamical behavior can be created.
- Errors can be corrected at an earlier stage.
- Models are *synchronized* within a supermodel: no variance reduction or smoothing



MME



Supermodel

A *weighted* supermodel

Consider two imperfect models with parametric error, with s denoting the supermodel solution and $\mathbf{W}$ the weights.

Combine states with frequency Δt:

$$\dot{\mathbf{x}}_1 = \delta_{\text{mod}(t,\Delta T)}\mathbf{f}(\mathbf{x}_s, \mathbf{p}_1) + (1 - \delta_{\text{mod}(t,\Delta T)})\mathbf{f}(\mathbf{x}_1, \mathbf{p}_1) \tag{1a}$$

$$\dot{\mathbf{x}}_2 = \delta_{\text{mod}(t,\Delta T)}\mathbf{f}(\mathbf{x}_s, \mathbf{p}_2) + (1 - \delta_{\text{mod}(t,\Delta T)})\mathbf{f}(\mathbf{x}_2, \mathbf{p}_2) \tag{1b}$$

$$\boxed{\mathbf{x}_s = \mathbf{W}_1\mathbf{x}_1 + \mathbf{W}_2\mathbf{x}_2} \quad \text{if} \quad \delta_{\text{mod}(t,\Delta T)} = 1. \tag{1c}$$

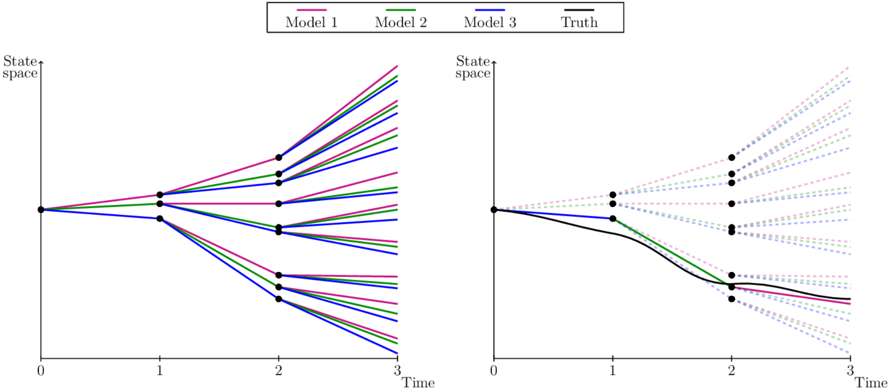
Kronecker $\delta = 1$ if $\text{mod}(t, \Delta T) = 0$, and 0 otherwise.

Training a supermodel implies learning the weights $\mathbf{W}$.

Learning method 1: *Cross Pollination in Time*

The method **Cross Pollination in Time** (**CPT**) proposed by *Smith, (2001)* combines different models by "crossing" their trajectories.

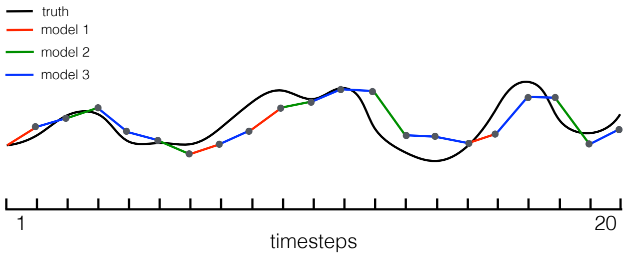
- CPT is designed to increase the amount of possible trajectories: additional areas of the state space are explored.



CPT in general.

CPT in the context of supermodeling.

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



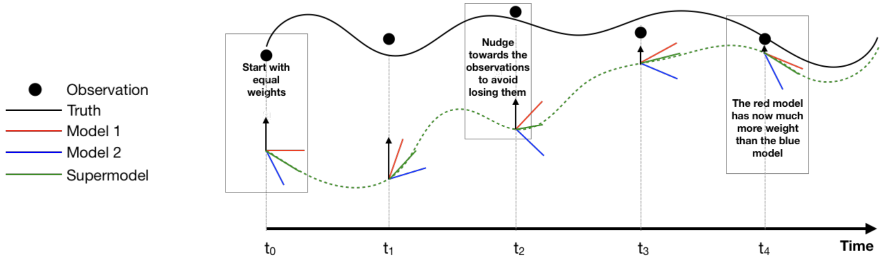
- Training phase gives the frequency at which each individual model prediction was found closest to the observations/truth.
- These frequencies determine the weights $\mathbf{W}_k$ for model k in the supermodel.

To help to follow the observations:

- *Iterative method*: next iteration also the supermodel can be chosen as closest model.
- *Nudging* towards the observations if CPT trajectory diverges too much.

Learning method 2: *synchronization based learning*

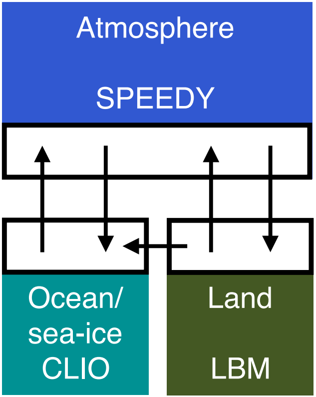
The **synch rule** *(Duane, 2007)* updates the weights such that synchronization errors between truth and supermodel are minimized.



$$\dot{\mathbf{w}}_k = -\delta \mathbf{e}(\mathbf{f}_k - \mathbf{f}_E) \qquad (2)$$

- Update of the weight $\mathbf{w}_k$ the weight for model k depends on covariance between $\mathbf{e}$, the difference between the supermodel and the truth, and $\mathbf{f}_k$, the time derivative of imperfect model k.
- *Stability*: Add equally weighted tendency $\mathbf{f}_E$ such that the total update of the weights for the N imperfect models equals zero.

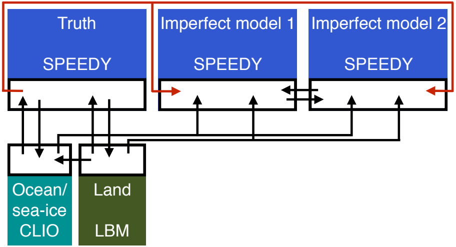
Results with an intermediate complex global climate model



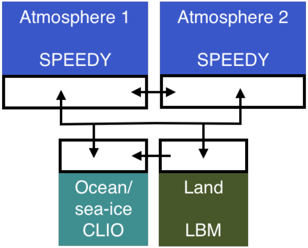
SPEEDO characteristics:

- *Spectral atmosphere model* with over 30,000 degrees of freedom, spatial resolution at the equator around 700 km.
- *Land model* with over 6,000 degrees of freedom
- *Ocean model* with primitive equations with free surface and over 200,000 degrees of freedom

SPEEDO configuration during training and prediction



Training configuration



Supermodel configuration

Aim: build up a weighted supermodel with two different SPEEDO atmospheres. Weights will be learned from both **CPT** and the **synch rule**.

SPEEDO experiment 1: Sparse and noisy observations

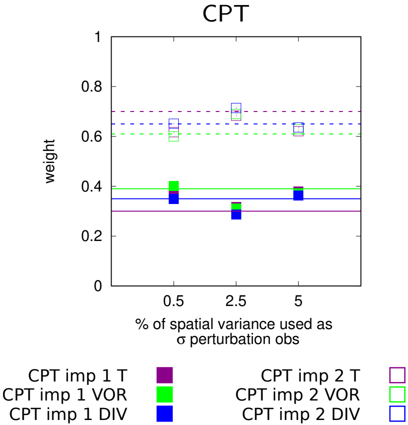
Training setup:

- Training period $T = 1$ year.
- *Global weights*: every grid point of the model obtains the same weights.
- Different weights for *temperature* (T), *vorticity* (VOR) and *divergence* (DIV).
- Model time step of 15 minutes.
- *Sparse observations*: only every 24hr.
- Observations with *Gaussian noise* $\sim N(0, \sigma)$. For T $\sigma \sim 0.15°$C, $0.75°$C and $1.5°$C.

Parameter perturbations create different models.

	Truth	**Model 1**	**Model 2**
relaxation timescale of convection (RtC)	6 hours	4 hours	8 hours
relative humidity threshold (RH)	0.9	0.85	0.95
momentum diffusion timescale (MDt)	24 hours	18 hours	30 hours

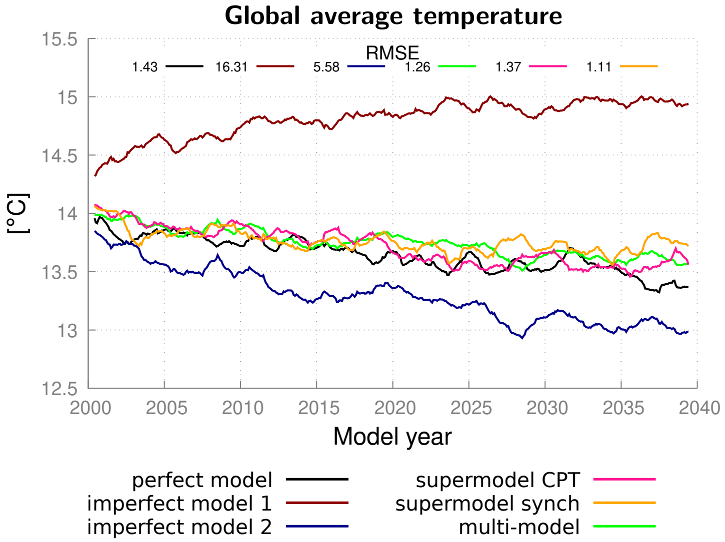
SPEEDO experiment 1: Sparse and noisy observations



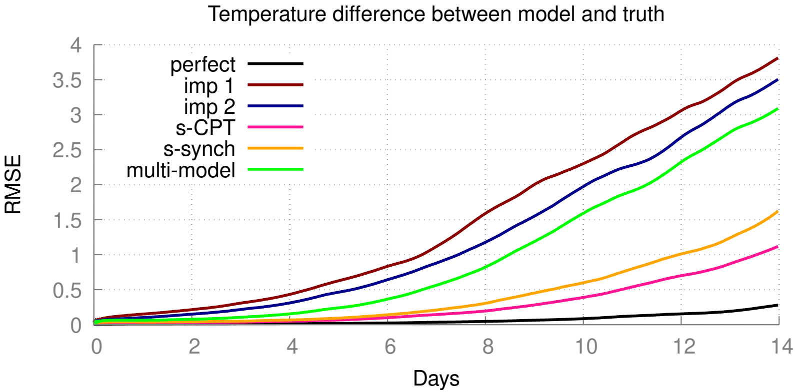
Weights for the supermodel consisting of mod 1 and 2.

- Horizontal lines (continuous mod 1, dashed mod 2) indicate the good weights from perfect observation experiments.
- Similar results for the synch rule.

SPEEDO experiment 1: Long term forecast quality



SPEEDO experiment 1: Short term forecast quality



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

SPEEDO experiment 2: Negative weights with CPT

Models cannot always compensate for each other with positive weights $\implies$ *negative weights!*

- Choose during training as closest model either: $\mathbf{x}_{neg1} = -\mathbf{x}_1 + 2\mathbf{x}_3$ or $\mathbf{x}_{neg2} = 2\mathbf{x}_1 - \mathbf{x}_3$.
- Then all weights of $\mathbf{w}_{1,3} \in [-1, 2]$.

- Weights obtained with *CPT* are very close to the (good) weights straightforwardly obtained by the *synch rule*.

	Truth	**Model 1**	**Model 3**
RtC	6 hours	4 hours	3 hours
RH	0.9	0.85	0.75
MDt	24 hours	18 hours	14 hours

Parameters $\mathbf{p}_{1,3}$ are all smaller than the corresponding values of $\mathbf{p}_{truth}$.

Average error in wind at 200 hPa over 30 years

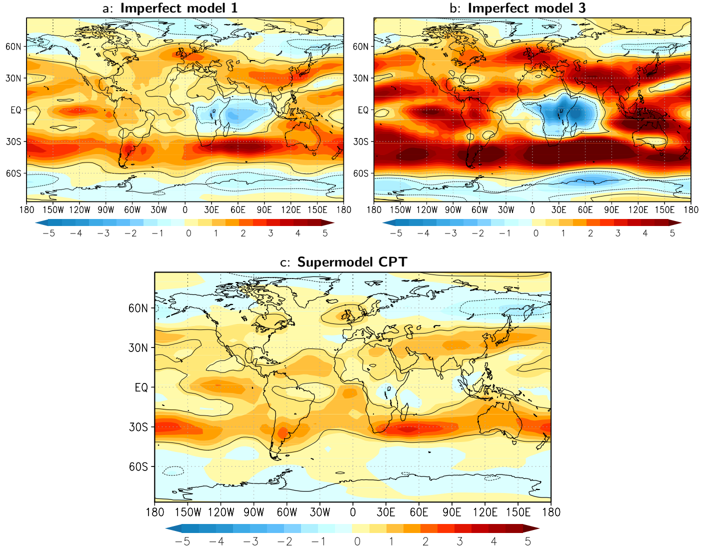


Figure: Average error in wind at 200 hPa over 30 years. Contours denote areas where the difference is larger than the sampling error at 95% confidence.

Summary

- We adapted the *CPT* and *synch rule* training methods such that they are suitable for training supermodels with *sparse and noisy observations*.
- The supermodels *outperformed the individual imperfect models* as well as the *multi-model ensemble* approach in the context of the SPEEDO model.
- Next to the synch rule, we are now able to obtain *negative connections* between the models from CPT as well.

(Results from Schevenhoven and Carrassi (2021), under review at GMD).

Discussion

Can we now make the step towards training complex state-of-the-art models?

- To what extend can we use the methods for training on longer time scales?
  - Creating **larger ensemble** of trajectories during training.
  - Not (only) creating a training trajectory with the smallest RMSE compared to the observations, but also w.r.t **climatological features**.
- Using a **neural network** to obtain non-lineair combinations between the models?

- **Structural difference** between state-of-the-art models, not parametric error only $\Longrightarrow$ define a **common state space**.