WPI: Model Classes, Metrics, Data and Prior Knowledge

Jürgen Kurths and Naoya Fujiwara
Potsdam Institute for Climate Impact Research (PIK), Germany
Objectives of WP1

• To develop a general theory of a supermodel apart from the learning aspect.

• Specify how models with different structure should be connected.

• Specify conditions under which connecting variables among different models will lead to superior skill, and the form of those connections.

• Determine limitations of the super modeling strategy.

• In addition to machine learning we develop a strategy how to define connections based on insight, mathematical arguments or whatever that lead to a useful supermodel because the automatic learning might be too complex, or lead to suboptimal solutions and we spread the risk this way.

• Serve those results as input for WP 2-5.
Tasks of WP I

• Task 1.1 Setting of the scope (models, metrics, data). What scope of general model fusion is a useful playground for the task of fusing different climate models.

• Task 1.1.1 Define Model classes: Determination of the model classes that we will study. We will concentrate on models that mimic to some extent the behavior of realistic climate models. We will also include model classes that have dynamics at different time scales, to anticipate on the differences in time scales of atmosphere and ocean in realistic climate models. We intend to study to what extend different qualitative behavior of the subsystems can be combined in a supermodel and how to treat cases when there are coexisting (multiple) attractors for some parameter configurations. In such a case special combination strategies have to be designed.

• Task 1.1.2 Metrics: Determination of metrics and appropriate test criteria that define how well a model (or a supermodel) performs with regard to the truth. Which are the important statistics to consider? How to deal with variables that have different physical meaning. Metrics are needed both to optimize the supermodel and to evaluate the performance of the supermodel compared to the state of the art.

• Task 1.1.3 Data and prior knowledge: Determination of the data and prior knowledge that one may assume to be available for the purpose of super modeling.
Highlight of Task 1.1.1: Model Classes

Different supermodel classes and different approaches to optimize the supermodel parameters are designed.

- **WP1 & WP2: simpler models in the hierarchy**
  - 1. Chaotic systems (3 degrees of freedom (dof))
  - 2. Atmosphere-ocean toy model (5 dof)
  - 4. Barotropic channel model with topographic forcing (6 dof)
  - 5. Barotropic model on the sphere with topographic forcing with 231 dof
  - 6. Baroclinic model with quasi-geostrophic dynamics and two vertical levels on the sphere with 30 dof:

- **WP3: 3. Kuramoto-Sivashinsky equation** (Duane, Basnarkov)
WP4: intermediate complexity models (Selten, Hiemstra)

7. Baroclinic model with quasi-geostrophic dynamics and three vertical levels on the sphere with 1449 dof

8. Same model, but with moisture and simple parameterizations for convection and radiation and land processes with in total 6313 dof).

9. Same model, but with dynamical ocean-sea-ice model instead of prescribed sea surface temperatures with in total about 210,000 dof.

10. Baroclinic model based on the primitive equations with 8 vertical levels on the sphere with moisture and simple parameterizations and land processes with in total 31680 dof

11. Same model but with dynamical ocean-sea-ice model instead of prescribed sea surface temperatures with in total about 250,000 dof

WP5: most complex state-of-the-art climate models
Simple models
Atmosphere-ocean model (van Veen, 2001)

- atmospheric circulation at midlatitudes (Lorenz 84) + temperature and salinity difference between the equatorial and polar box \((T, S)\): Stommel model

- Study the effect of the slow oceanic time-scale on the ability of the super model to learn on the basis of short integrations.

- Is it possible to obtain a super model solution that is able to describe the true system well with imperfections present in the oceanic subsystem?

\[
\begin{align*}
\dot{x} &= -y^2 - z^2 - ax + a(F_0 + F_1 T), \\
\dot{y} &= xy - b x z - y + G_0 + G_1 (T_{av} - T), \\
\dot{z} &= b x y + x z - z.
\end{align*}
\]

\[
\begin{align*}
\dot{T} &= k_w (\gamma x - T) - |f(T, S)| T - k_w T, \\
\dot{S} &= \delta_0 + \delta_1 (y^2 + z^2) - |f(T, S)| S - k_w S
\end{align*}
\]

\[
\begin{array}{cccc}
 a & 1/4 & \delta_0 & 7.8 \cdot 10^{-7} \\
b & 4 & k_w & 1.8 \cdot 10^{-5} \\
F_0 & 8 & k_a & 1.8 \cdot 10^{-4} \\
G_0 & 1 & \chi & 1.1 \cdot 10^{-3} \\
\gamma & 30 & \omega & 1.3 \cdot 10^{-4} \\
\delta_1 & 9.6 \cdot 10^{-8} & T_{av} & 30
\end{array}
\]
Low order barotropic model

• Barotropic channel model with topographic forcing with 6 dof: an atmospheric system with multiple equilibria

conservation of absolute vorticity (relative vorticity $\zeta$ + planetary vorticity $f=2\Omega \sin \Phi$, $\Omega$: angular velocity of earth, $\Phi$: latitude), $\Psi$: stream function

$$\dot{\zeta} = -J(\psi, \zeta + f + \gamma h) - C \nabla^2 (\psi - \psi^*)$$
$$J(A, B) = (\partial A/\partial \lambda)(\partial B/\partial \mu) - (\partial A/\partial \mu)(\partial B/\partial \lambda).$$

projected onto eigenfunctions of the Laplace operator

$$\phi_{0m} = \sqrt{2} \cos (my/b)$$
$$\phi_{nm} = \sqrt{2} \sin (my/b)e^{inx},$$

Crommelin et al (2004): a six-dimensional model

$$\dot{x}_1 = \gamma_1 x_3 - C(x_1 - x_1^*)$$
$$\dot{x}_2 = - (\alpha_1 x_1 - \beta_1)x_3 - Cx_2 - \delta_1 x_4 x_6$$
$$\dot{x}_3 = (\alpha_1 x_1 - \beta_1)x_2 - \gamma_1 x_1 - Cx_3 + \delta_1 x_4 x_5$$
$$\dot{x}_4 = \gamma_2 x_6 - C(x_4 - x_4^*) + \epsilon(x_2 x_6 - x_3 x_5)$$
$$\dot{x}_5 = -(\alpha_2 x_1 - \beta_2)x_6 - Cx_5 - \delta_2 x_4 x_3$$
$$\dot{x}_6 = (\alpha_2 x_1 - \beta_2)x_5 - \gamma_2 x_4 - Cx_6 + \delta_2 x_4 x_2.$$
Barotropic model with realistic atmospheric flow (Selten, 1995)

- Barotropic model on the sphere with topographic forcing with 231 dof: 
  **address questions regarding the density of connections**

  \[ \dot{\zeta} = -\mathcal{J}(\psi, \zeta + f + h) - k_1 \zeta + k_2 \Delta^3 \zeta + \zeta^* \]

  - \( k_1 \): Ekman damping
  - \( k_2 \): scale selective damping
  - \( h \): non-dimensional orography

  Spherical harmonics expansion

  \[ \psi(\lambda, \mu, t) = \sum_{n=1}^{2l} \sum_{m=-n}^{+n} \psi_{m,n}(t) Y_{m,n}(\lambda, \mu) \]

  \[ Y_{m,n}(\lambda, \mu) = P_{m,n}(\mu)e^{im\lambda} \]
• show movie of Bartropic model created by Frank
Task 1.1.3: Data and prior knowledge

• When applying the super modeling approach to the real climate system, one is constrained by the amount and quality of observational data.

• reanalysis data sets of order 50 years from different weather forecasting centers (NCEP/NCAR, ECMWF)

• For SUMO we propose to use the reanalysis data sets for the learning phase.

• data-assimilation process has taken care of the assessment of the quality of the observations, poor quality observations have been discarded and all available observations have been taken into account to obtain a best estimate of the state of the atmosphere at every temporal and spatial point of the atmosphere model used in the data-assimilation process.
Short Summary

• Classes of climate models are introduced with different aspects to clarify the feasibility of Super modeling

  • Atmosphere-Ocean model: effect of different time scales
  • temperature and velocity as state variables
  • the density of connections

• We use the reanalysis data sets for the learning phase
Task 1.1.2 Metrics to measure supermodel performance
Issues related to the choice of metric

• Quantify difference of the invariant measures of truth and super model

• Capture dynamics of the systems: consider state variables and their velocity

• Is the choice of the metrics of influence on the ability of the learned supermodel to predict the climate response to a perturbation?
Kullback-Leibler divergence

\[
KL(P \| Q) = - \int P(x) \log \frac{Q(x)}{P(x)} \, dx
\]

\[
KL(P \| Q) \geq 0 \quad KL(P \| Q) = 0 \iff P = Q
\]

P: truth
Q: super model output

measure difference between two probability distributions (attractors)
Static metric

Compare static distributions of the ground truth $P(x)$ and the approximation $Q(x)$ with the introduction of the smoothing parameter $h$

$$P(x) \approx \frac{\alpha}{N} \sum_{n=1}^{N} \exp\left(-\frac{\|x - x(t_n)\|^2}{2h^2}\right)$$

$$Q(x) \approx \frac{\alpha}{N} \sum_{n=1}^{N} \exp\left(-\frac{\|x - y(t_n)\|^2}{2h^2}\right)$$

to prevent that KL diverges.
Static metric

$\text{KL}(P \| Q) \approx \dfrac{-1}{N} \sum_{n=1}^{N} \log \left( \dfrac{\sum_{k=1}^{N} \exp \left( -\dfrac{\|x(t_n) - y(t_k)\|^2}{2h^2} \right)}{\sum_{k=1}^{N} \exp \left( -\dfrac{\|x(t_n) - x(t_k)\|^2}{2h^2} \right)} \right)$

$\equiv D_{\text{static}}(\{x(t_n)\}_{n=1}^{N} \| \{y(t_n)\}_{n=1}^{N}; h)$
Dynamic metric

If two dynamical systems have the same attractor but their speed is different, the static metric does not see a difference. Therefore we define speeds

\[
\begin{align*}
u(t_n) &\approx \frac{x(t_{n+1}) - x(t_{n-1})}{2\Delta t} \\
v(t_n) &\approx \frac{y(t_{n+1}) - y(t_{n-1})}{2\Delta t}
\end{align*}
\]
Dynamic metric

And we define the dynamic divergence as

\[
\begin{align*}
\text{KL}(P||Q) & \approx -\frac{1}{N-2} \sum_{n=2}^{N-1} \log \left( \sum_{k=2}^{N-1} \exp \left( - \frac{\|x(t_n) - y(t_k)\|^2}{2h_x^2} - \frac{\|u(t_n) - v(t_k)\|^2}{2h_v^2} \right) \right) \\
& \equiv D_{\text{dynamic}}(\{x(t_n)\}_{n=1}^{N} || \{y(t_n)\}_{n=1}^{N}; h_x, h_v)
\end{align*}
\]
We calculate the KL metric from different sets of time series.

• Three time series $X_1, X_2, X_3$ from the standard Lorenz model with different initial conditions
• One time series $Y$ from the Lorenz model with different parameters
• One time series $V$ from the standard Lorenz model but moving twice as fast
Static metric

<table>
<thead>
<tr>
<th>$h_x$</th>
<th>$S$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$Y$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>3.673</td>
<td>3.966</td>
<td>Inf</td>
<td>6.706</td>
<td></td>
</tr>
<tr>
<td>1.000</td>
<td>0.089</td>
<td>0.030</td>
<td>2.087</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>10.000</td>
<td>0.017</td>
<td>0.008</td>
<td>0.247</td>
<td>0.013</td>
<td></td>
</tr>
</tbody>
</table>

- For small smoothing, divergence between $X_1$ and $X_2$ ($X_3$) can be quite big, due to the spiky character of the distributions, and is smaller when there is more smoothing.

- Value between $X_1$ and $Y$ is larger than that between $X_1$ and $X_2$ ($X_3$).

- For double speed generated series ($V$) the divergence is almost comparable with the divergences between the series of the identical system.

- Static divergence is insensitive to the dynamics.
Static metric

<table>
<thead>
<tr>
<th>$h_x$</th>
<th>$S$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$Y$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td></td>
<td>3.673</td>
<td>3.966</td>
<td>Inf</td>
<td>6.706</td>
</tr>
<tr>
<td>1.000</td>
<td></td>
<td>0.089</td>
<td>0.030</td>
<td>2.087</td>
<td>0.082</td>
</tr>
<tr>
<td>10.000</td>
<td></td>
<td>0.017</td>
<td>0.008</td>
<td>0.247</td>
<td>0.013</td>
</tr>
</tbody>
</table>

- For small smoothing, divergence between $X_1$ and $X_2$ ($X_3$) can be quite big, due to the spiky character of the distributions, and is smaller when there is more smoothing.

- Value between $X_1$ and $Y$ is larger than that between $X_1$ and $X_2$ ($X_3$).

- For double speed generated series ($V$) the divergence is almost comparable with the divergences between the series of the identical system.

- Static divergence is insensitive to the dynamics.

More smoothing
• For small smoothing, divergence between $X_1$ and $X_2$ ($X_3$) can be quite big, due to the spiky character of the distributions, and is smaller when there is more smoothing.

• Value between $X_1$ and $Y$ is larger than that between $X_1$ and $X_2$ ($X_3$).

• For double speed generated series ($V$) the divergence is almost comparable with the divergences between the series of the identical system.

• Static divergence is insensitive to the dynamics.

<table>
<thead>
<tr>
<th>$h_x$</th>
<th>$S$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$Y$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td></td>
<td>3.673</td>
<td>3.966</td>
<td>Inf</td>
<td>6.706</td>
</tr>
<tr>
<td>1.000</td>
<td></td>
<td>0.089</td>
<td>0.030</td>
<td>2.087</td>
<td>0.082</td>
</tr>
<tr>
<td>10.000</td>
<td></td>
<td>0.017</td>
<td>0.008</td>
<td>0.247</td>
<td>0.013</td>
</tr>
</tbody>
</table>

“ground truth”: $X_1$
## Dynamic metric

<table>
<thead>
<tr>
<th></th>
<th>$h_x$</th>
<th>$0.100$</th>
<th>$1.000$</th>
<th>$10.000$</th>
<th>$100.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_v$</td>
<td>1.000</td>
<td>9.236</td>
<td>3.909</td>
<td>3.732</td>
<td>3.673</td>
</tr>
<tr>
<td></td>
<td>5.000</td>
<td>5.204</td>
<td>0.291</td>
<td>0.140</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>10.000</td>
<td>5.114</td>
<td>0.231</td>
<td>0.078</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>100.000</td>
<td>1.000</td>
<td>5.000</td>
<td>10.000</td>
<td>100.000</td>
</tr>
</tbody>
</table>

- **X1 vs X2**: Increasing smoothing of speeds
- **X1 vs X3**: Increasing smoothing of states

### Data and prior knowledge

A prerequisite for the super modeling approach is that enough high quality data of the true system is available for the learning. In the perfect model setting we can generate in principle as much data of the true system as is needed for the learning with the super model. This data is free of observational errors.

When applying the super modeling approach to the real climate system, one is constrained by the amount and quality of observational data. For climate analysis and climate modeling purposes, reanalysis datasets have been compiled by different weather forecasting centers over the recent years (INCEP/NCAR, ECMWF). What are reanalysis datasets? In the practice of weather forecasting, available observational data over the past say 12 hours from meteorological stations, balloons, aircraft, satellites, ships, are combined with an atmospheric model to obtain a best estimate of the state of the atmosphere at the computational grid of the atmosphere model at a particular moment in time. This process is called data assimilation and the estimate is called an analysis. A weather forecast is produced by an integration into the future of the atmospheric model starting from this initial state. Data assimilation
Dynamic metric

<table>
<thead>
<tr>
<th>$h_x$</th>
<th>$h_v$</th>
<th>1.000</th>
<th>5.000</th>
<th>10.000</th>
<th>100.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>Inf</td>
<td>Inf</td>
<td>Inf</td>
<td>Inf</td>
<td>Inf</td>
</tr>
<tr>
<td>1.000</td>
<td>Inf</td>
<td>16.408</td>
<td>7.775</td>
<td>2.189</td>
<td></td>
</tr>
<tr>
<td>10.000</td>
<td>Inf</td>
<td>4.303</td>
<td>1.857</td>
<td>0.345</td>
<td></td>
</tr>
</tbody>
</table>

- the double speed model clearly gives a higher divergence
- sensitive to dynamics.
Highlights of Task 1.1.2

• The proposed metrics (static, dynamic) based on Kullback-Leibler divergence performs well with the Lorenz 63 and T5 models.

• The static metric measures the geometric difference of two attractors, but difference in dynamical property is not taken into account. The dynamic metric detects difference of dynamics.
WP 1 Second year objectives

• Study synchronization between structurally different models (e.g. barotropic model with different truncation) and explore the limits for synchronization to occur

• Study efficient connections (e.g. EOF) for model classes with lower dof (with WP4)

• Apply the KL based metric to the supermodeling problem (with other WPs)