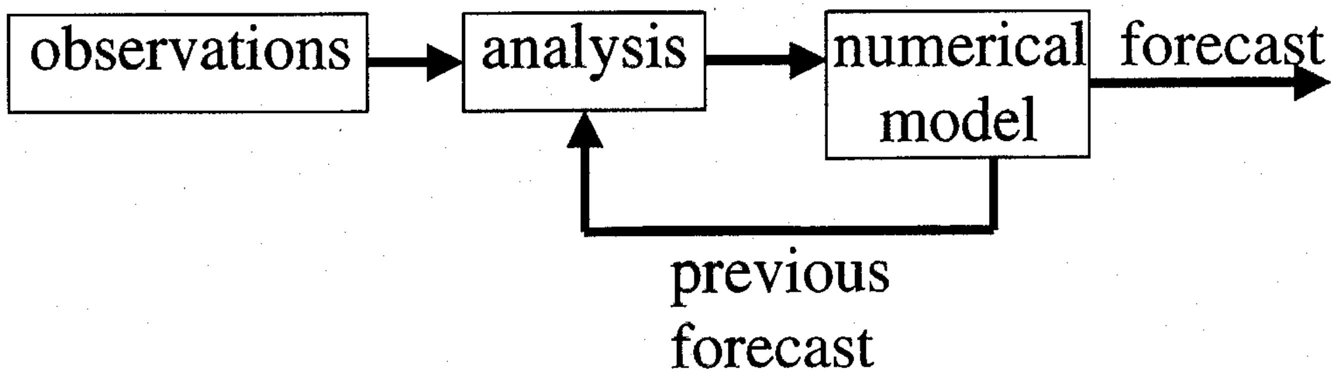
Consensus By Synchronization of Alternative Models

Greg Duane
MANU

-with contributions from
Ljupco Kocarev
Frank Selten
Wim Wiegerinck
Lasko Basnarkov
Leonie van den Berge
Paul Hiemstra

Supported by:
DOE Grant# DE-SC0005238
and
ERC Grant# 266722

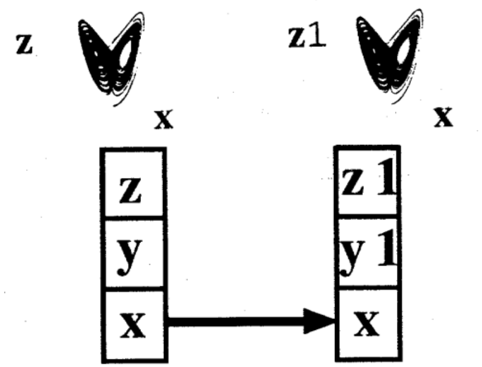
DATA ASSIMILATION

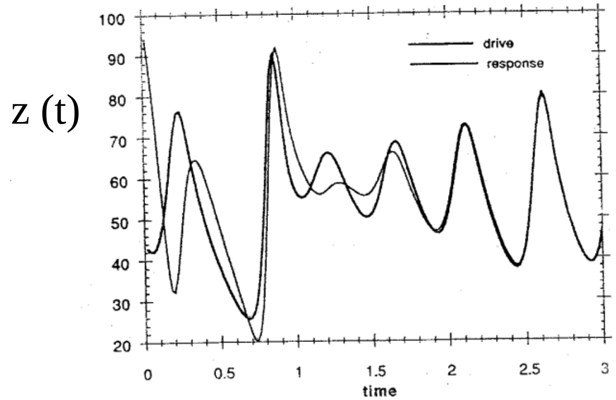


- the model is a semi-autonomous dynamical system influenced by observations

SUPPOSE THE WORLD IS A LORENZ SYSTEM AND ONLY *X* IS OBSERVED

- two coupled chaotic systems can fall into synchronized motion along their strange attractors when linked through only one variable



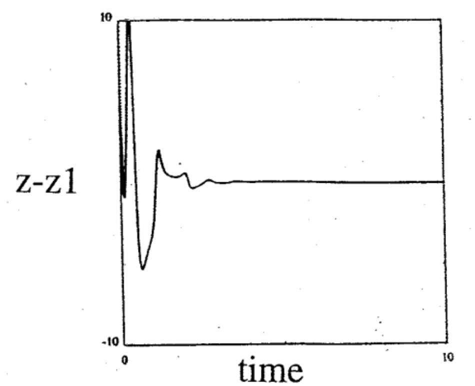


$$x' = \sigma(y-x)$$

$$y' = \rho x - y - xz \qquad y_1' = \rho x - y_1 - x(z_1)$$

$$z' = -\beta z + xy \qquad z_1' = -\beta(z_1) + x(y_1)$$

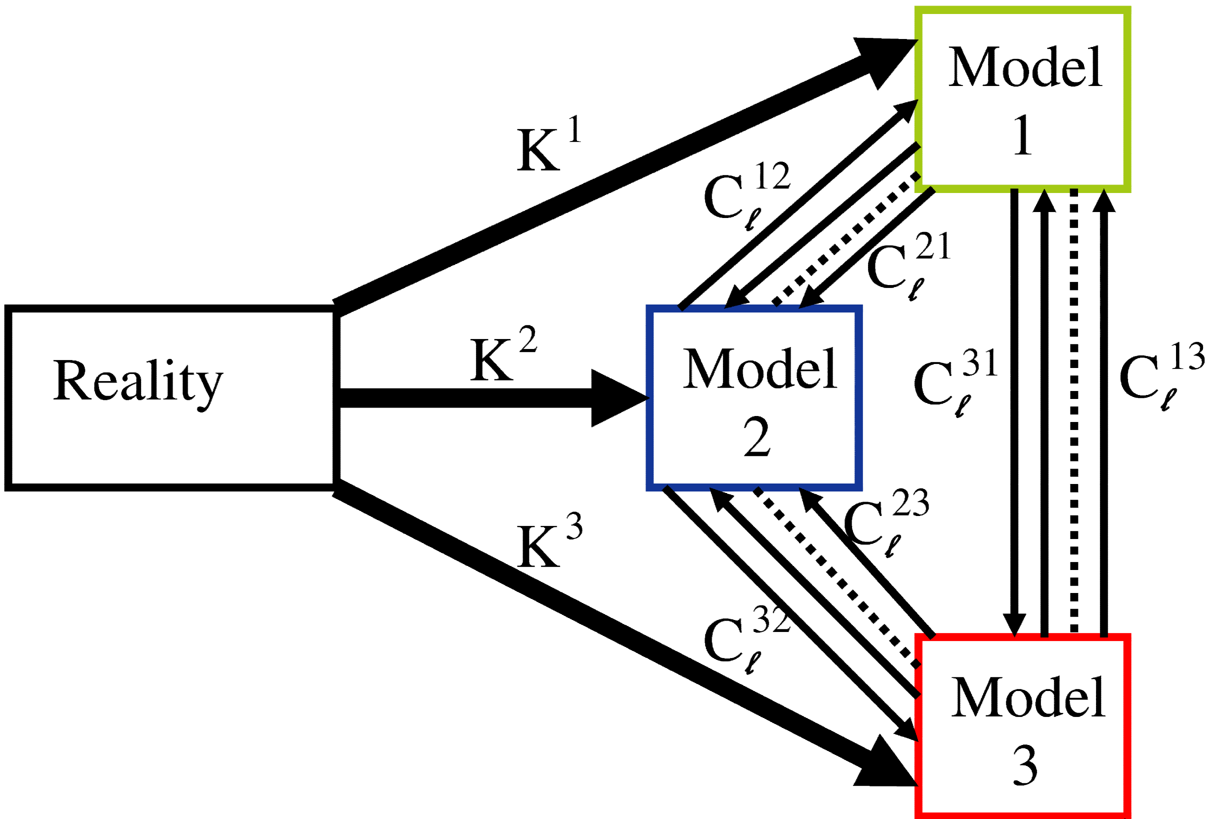
(also works for y-coupling, but not for z-coupling)



(*Pecora and Carroll '90*)

- SYNCHRONIZATION ⟶ DATA ASSIMILATION

LET A COLLECTION OF MODELS ASSIMILATE DATA FROM (SYNCHRONIZE WITH) ONE ANOTHER; ADAPT THE COUPLING COEFFICIENTS



K^i constant:
data assimilation

adapt C_l^{ij}:
learning

CONSENSUS

-couple corresponding "model elements" l

SUMMARY

• Problem: IPCC-class climate models give widely divergent predictions in regard to:

a) magnitude of long-term climate change

b) detailed regional predictions

c) short-term climate change

Can we do better than averaging model outputs?

• Potential Solution: Take the synchronization view of data assimilation, and allow models to form a consensus (synchronize) by assimilating data from one another.

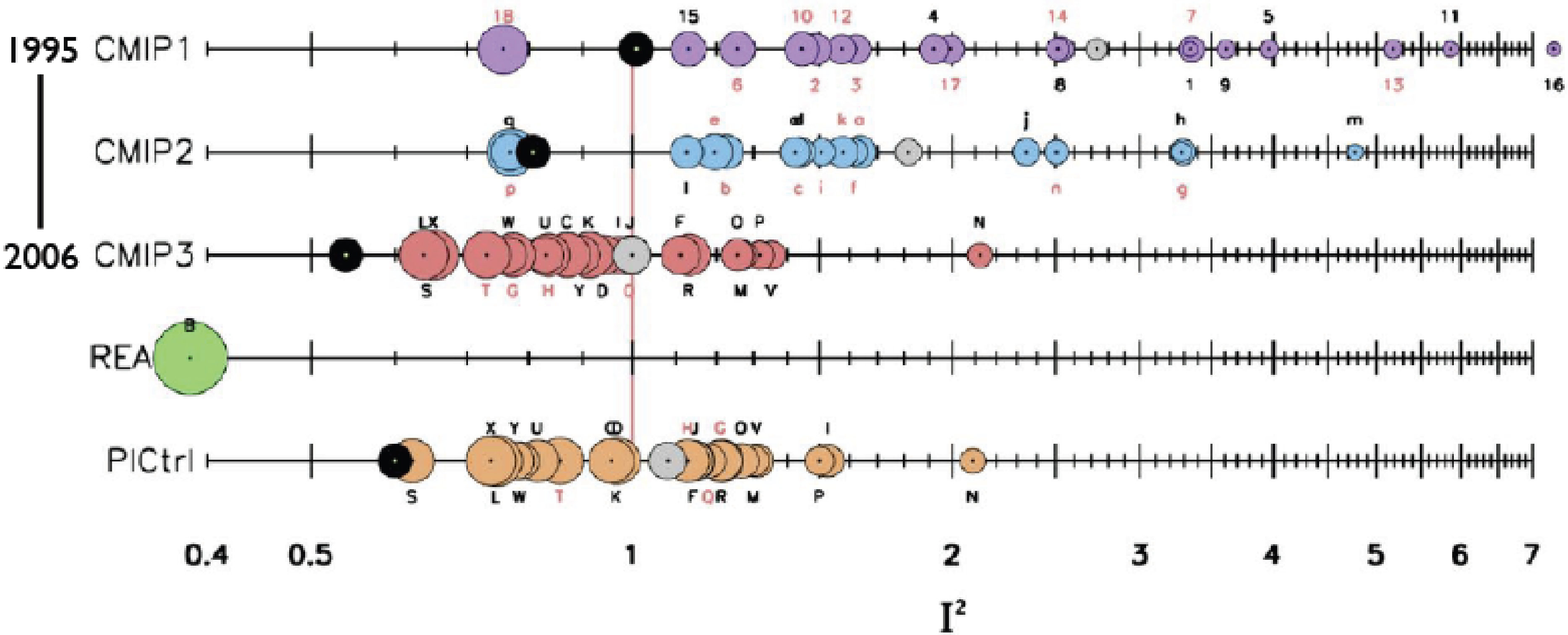
- Sync extends the “nudging” approach to assimilation.
- Parameters can be nudged as well as states *without ensembles*.
- Choose the adaptable parameters to be connection coefficients linking corresponding variables in different models; adapt them using historical data.

Coupled Model Intercomparison Project

Reichler, T., and J. Kim (2008): How Well do Coupled Models Simulate Today's Climate? *Bull. Amer. Meteor. Soc.*, **89**, 303-311.

Performance metric

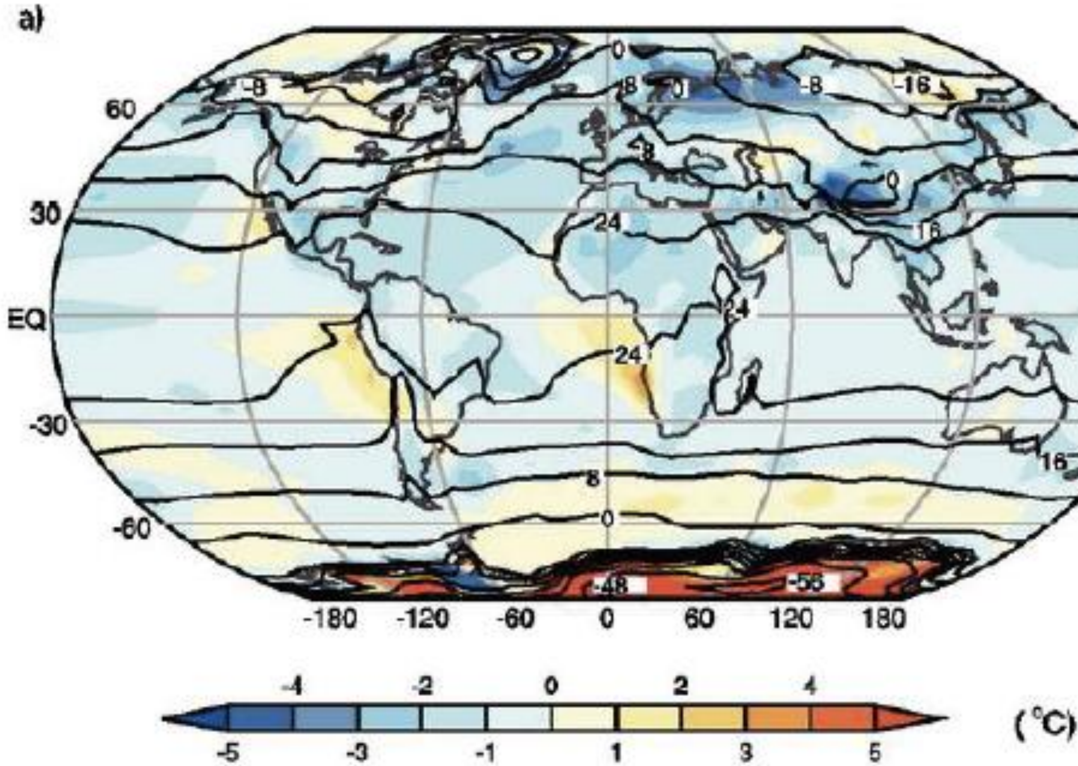
Based on mean squared errors in time mean global temperatures, winds, precipitation, ….



= index value based on multi model mean fields: outperforms individual models: why ?

Error in annual mean surface air temperatures

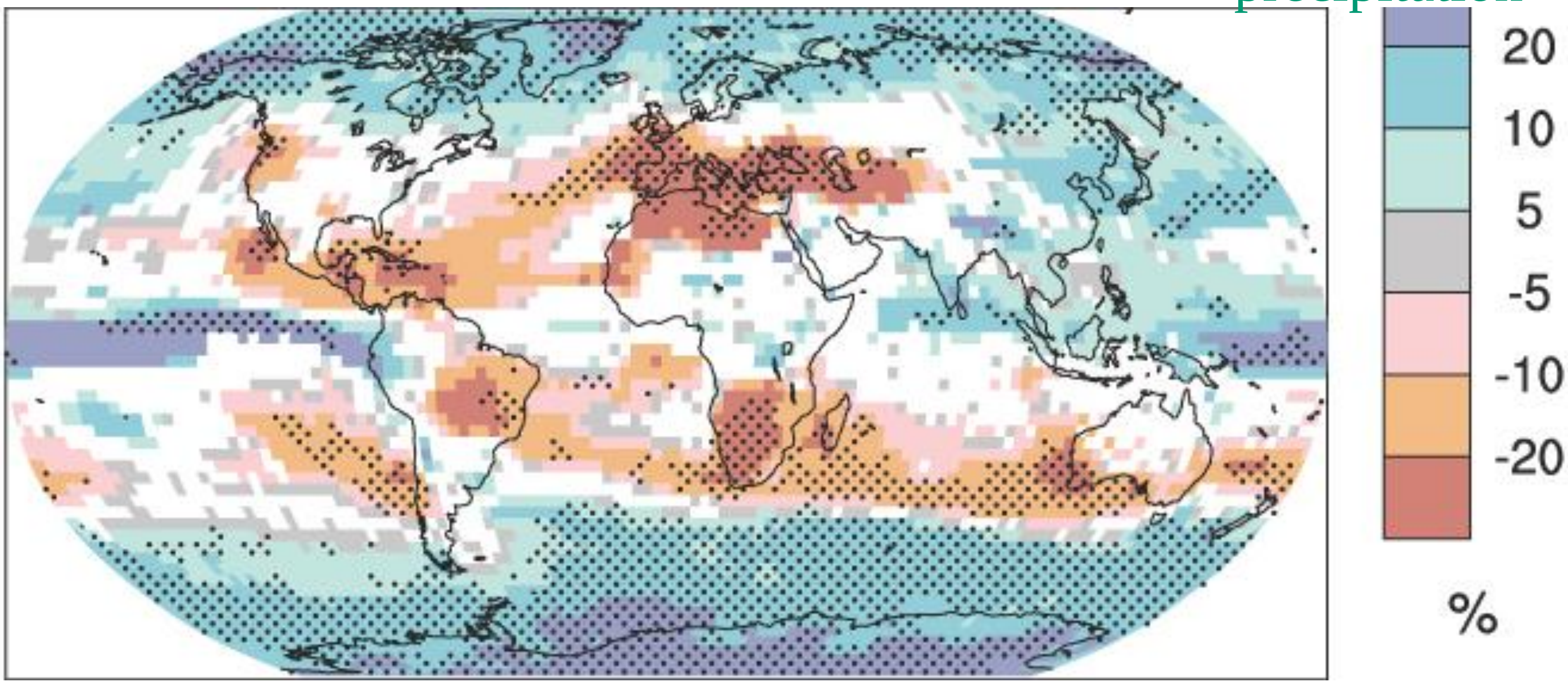
multi model mean over all CMIP3 simulations



IPCC 2007

EXAMPLE: DIVERGENT MODEL PROJECTIONS OF REGIONAL PRECIPITATION CHANGE

increased or decreased precipitation

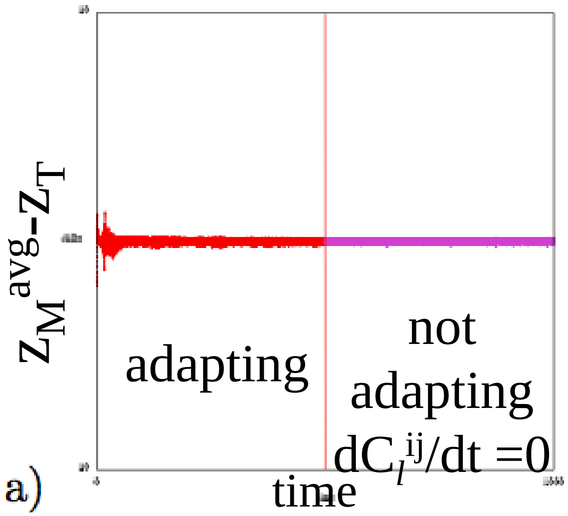


White areas: less than 2/3 of models agree on the sign of precipitation change

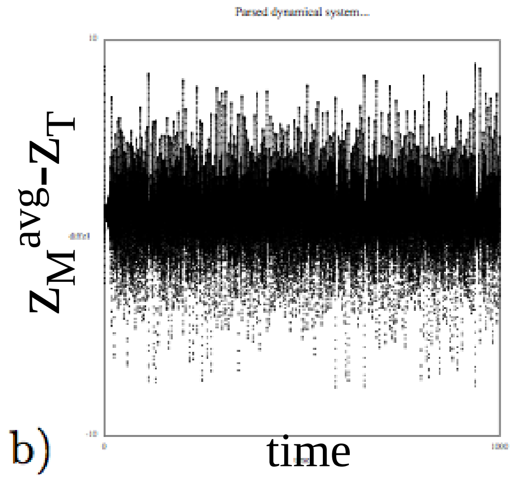
Stippled areas: more than 90% of models agree on the sign

Test Case: Fusing 3 Lorenz Systems With Different Parameters

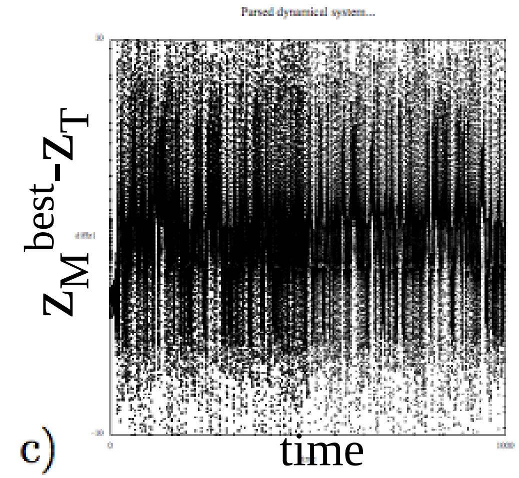
Fused Models



Average Output of Models (Unfused)



z from Model With Best z Eqn



$$
\begin{aligned}
\dot{x} &= \sigma(y-z) \\
\dot{y} &= \rho x - y - xz \\
\dot{z} &= -\beta z + xy \\
\dot{x}_i &= \sigma_i(y_i - z_i) + \sum_{j\neq i} C^x_{ij}(x_j - x_i) + K_x(x - x_i) \\
\dot{y}_i &= \rho x_i - y_i - x_i z_i + \mu_i + \sum_{j\neq i} C^y_{ij}(y_j - y_i) + K_y(y - y_i) \qquad i = 1,2,3 \\
\dot{z}_i &= -\beta_i z_i + x_i y_i + \sum_{j\neq i} C^z_{ij}(z_j - z_i) + K_z(z - z_i)
\end{aligned}
$$

$dC_x^{ij}/dt = a(x_j - x_i)(x - \tfrac{1}{3}\sum x_k)$

$dC_y^{ij}/dt = \ldots\ldots$

$dC_z^{ij}/dt = \ldots\ldots$

- Model fusion is superior to any weighted averaging of outputs

Supermodeling Relies on 3-Way Synchronization of Truth and Alternative Models

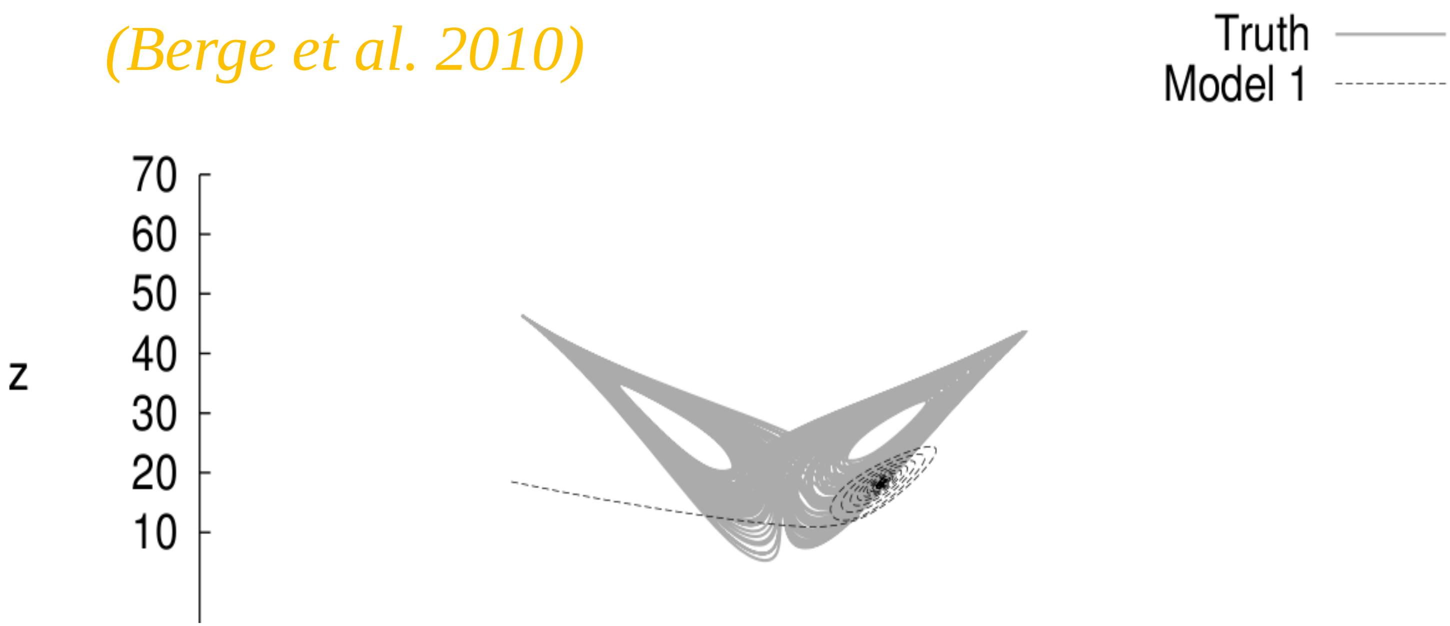
Hebbian learning: "cells that fire together wire together"

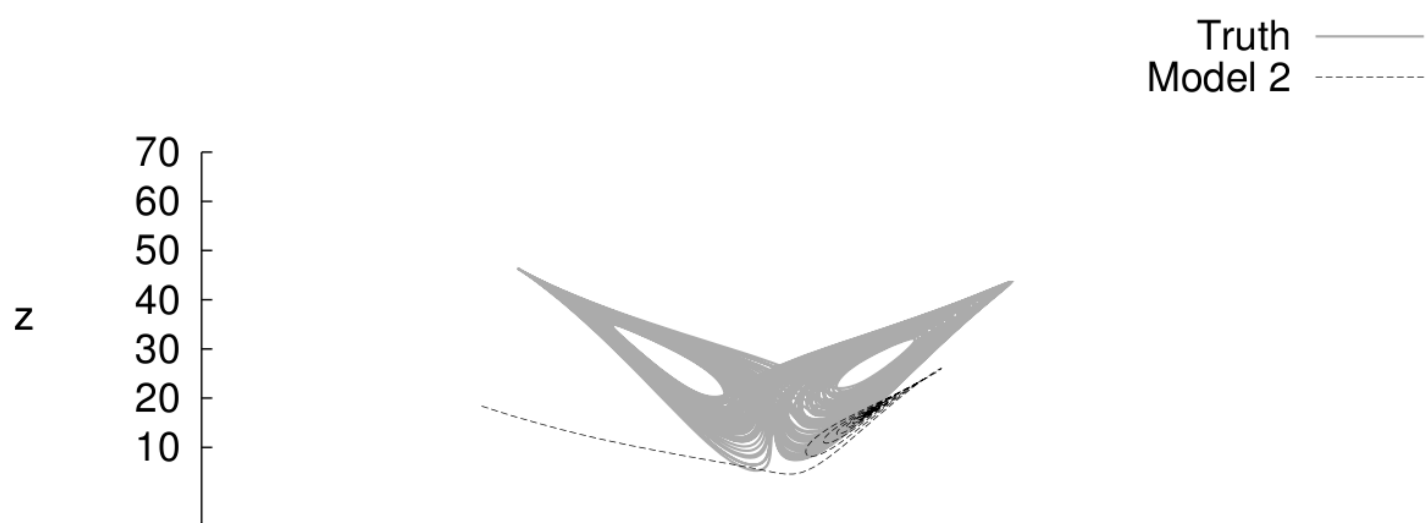
$$\Downarrow$$

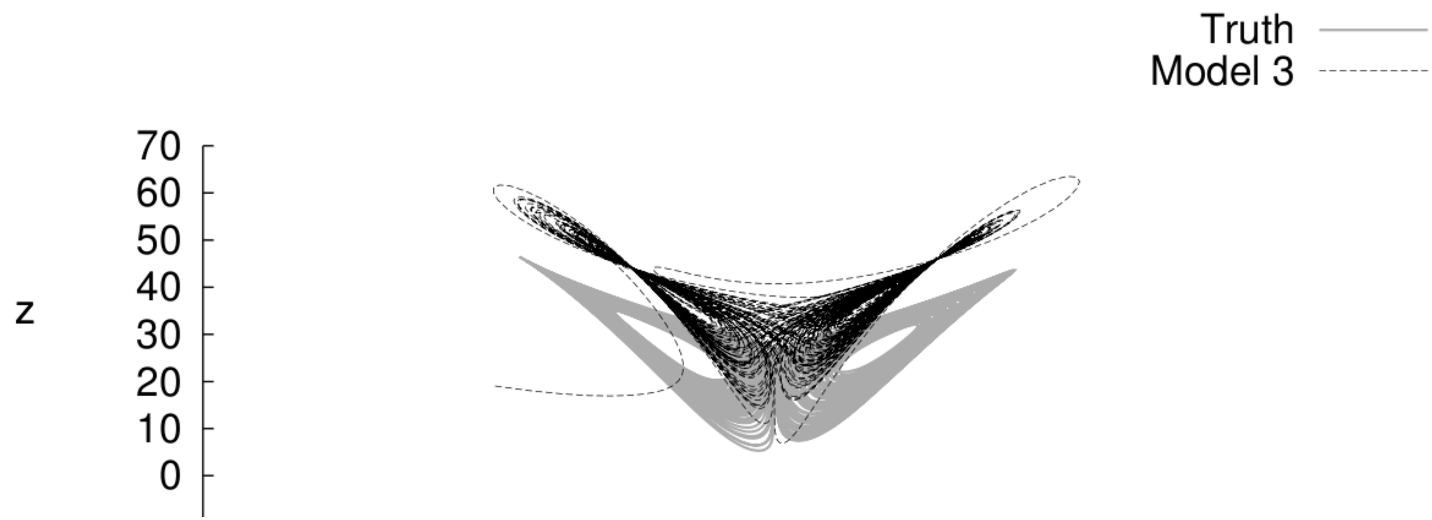
Supermodel learning: model elements "wire" together in such a way that they "fire" in synchrony with reality

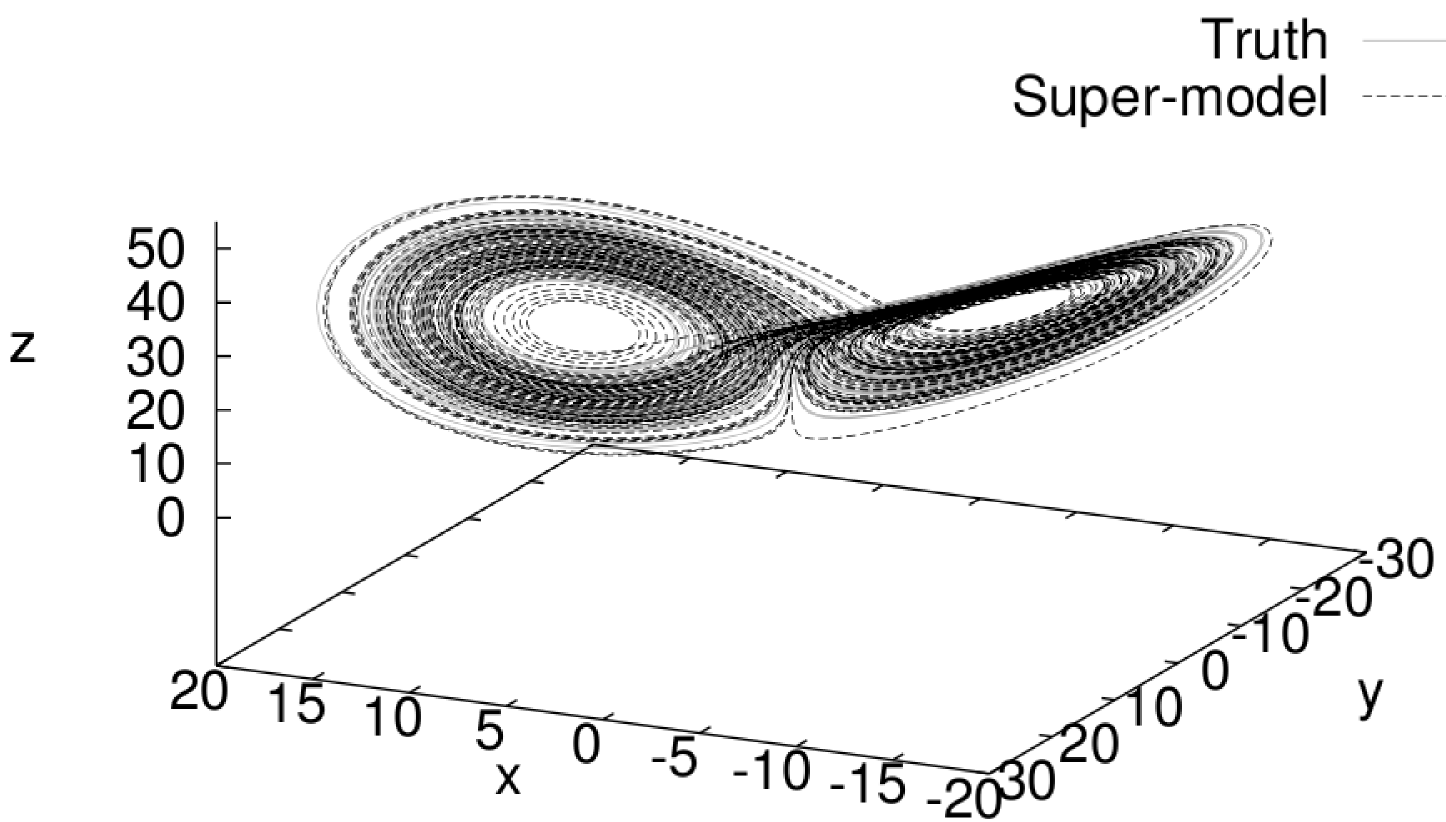
…..OR CAN USE STANDARD MACHINE LEARNING METHODS TO ADAPT INTER-MODEL CONNECTIONS

(Berge et al. 2010)



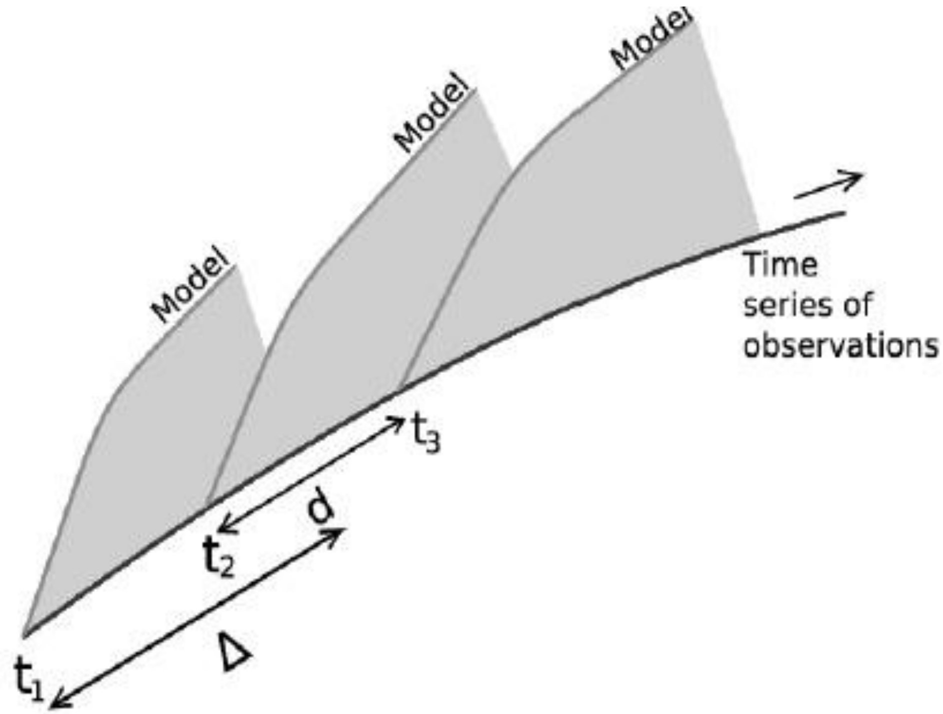



Truth
Model 3
70
60
50
40
30
20
10
0
z


Truth
Super-model
50
40
30
20
10
0
z
20
15
10
5
0
-5
-10
-15
-20
x
-30
-20
-10
0
10
20
30
y

Learning Algorithm

$$
\begin{aligned}
x_s &= \frac{1}{3}(x_1 + x_2 + x_3) \\
y_s &= \frac{1}{3}(y_1 + y_2 + y_3) \\
z_s &= \frac{1}{3}(z_1 + z_2 + z_3)
\end{aligned}
$$



Minimize Cost function:

$$F(\boldsymbol{C}) = \frac{1}{K\Delta}\sum_{i=1}^{K}\int_{t_i}^{t_i+\Delta} |\boldsymbol{x}_s(\boldsymbol{C},t) - \boldsymbol{x}_o(t)|^2 \gamma^t \, dt$$

The cost function is normalized by $\frac{1}{K\Delta}$, so that it represents the time averaged mean squared error. The factor γ^t with $0 < \gamma \leq 1$ is introduced to give stronger weight to the errors close to the initial condition

Supermodeling Works With Multi-time-scale Models

Lorenz '84 coupled to ocean box model:

$$x' = -(y^2) - (z^2) - a\,x + a\,(F_0 + F_1\,T) \qquad f = \omega\,T - \xi\,S$$

$$y' = x\,y - b\,x\,z - y + G_0 + G_1\,(T_{av} - T)$$

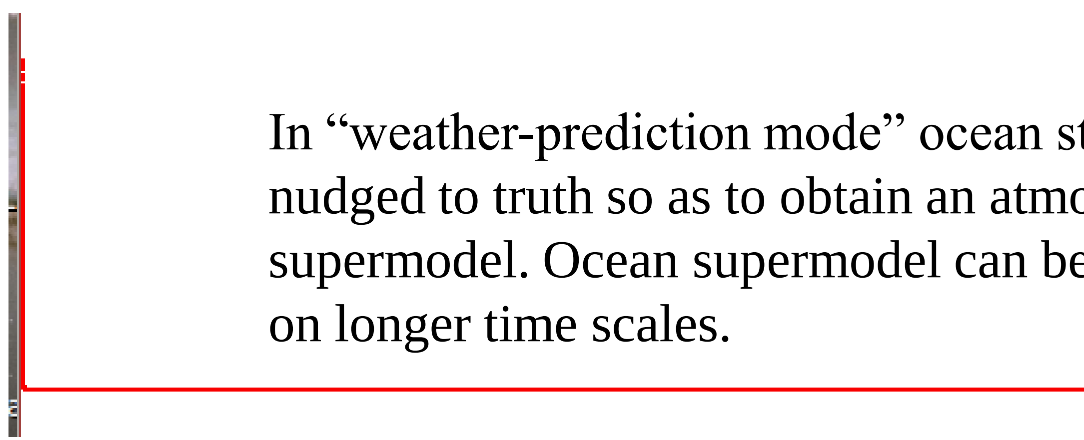
$$z' = b\,x\,y + x\,z - z$$

$$T' = k_a\,(\gamma\,x - T) - |f|\,T - k_w\,T$$

$$S' = \delta_0 + \delta_1\,(y^2 + z^2) - |f|\,S - k_w\,S$$

$X_{supermodel} - X_{truth}$

$T_{supermodel} - T_{truth}$



In "weather-prediction mode" ocean strongly nudged to truth so as to obtain an atmospheric supermodel. Ocean supermodel can be trained on longer time scales.

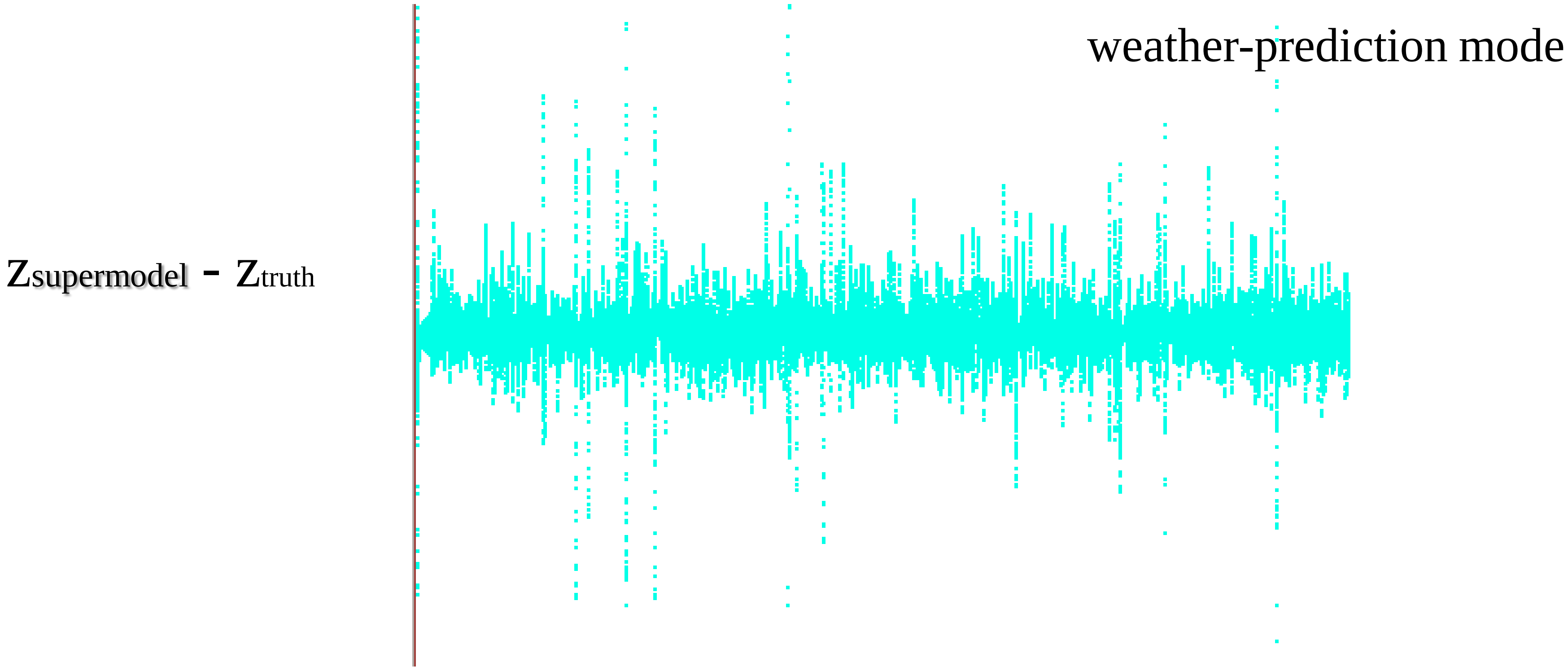
What if all models err in the same way?

- Analogous to group of experts who make the same mistake, but to different degrees
- Use the advice of the expert who is consistently “least wrong” and extrapolate to values outside the range of the group
- A very risky procedure!

What if all models are biased in same direction?

Lorenz supermodel with $\sigma_{truth} < \sigma_1, \sigma_2, \sigma_3$

$\Rightarrow$ Some connections become negative



Not as effective as positive connections, but better than averaging.

What if parameters shift between training and testing?

Train with Lorenz $\rho=28$ and then reset ρ in “reality” and in 3 “models”

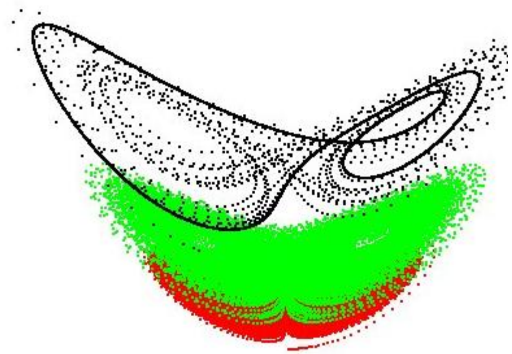
<u>Fusion</u>

adaptation

$\rho=28$ $\rho=50$ $\rho=100$

<u>Average of outputs (unfused)</u>

<u>Attractors</u>

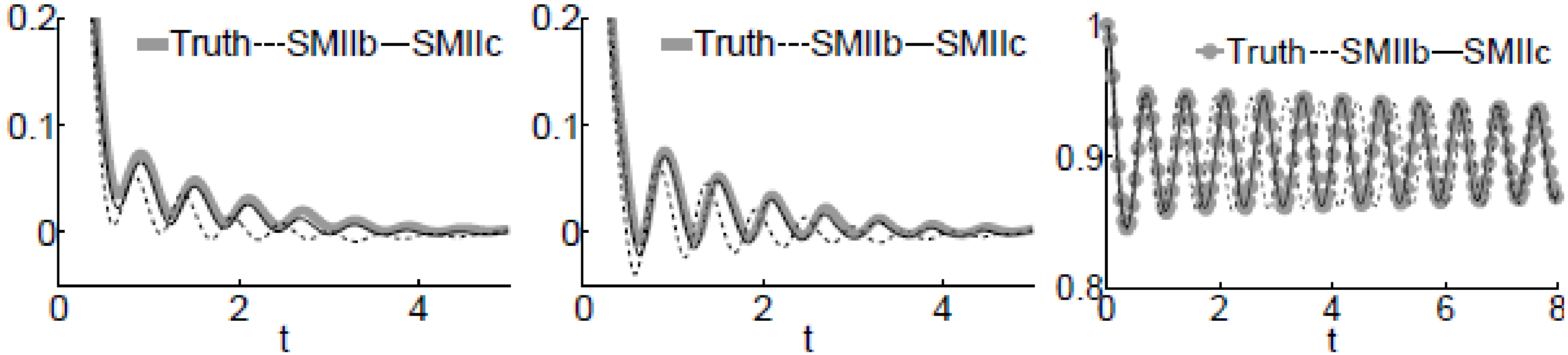


-fusion still better than averaging even when training and test systems differ by a large number of dynamical bifurcations

What if the connection scheme obtained by cost-minimization is only locally optimal?

- stochastic learning methods can help optimize supermodel

Autocorrelations for Truth and Two Supermodels



(a) x (b) y (c) z

SMIIb is formed using a deterministic learning method

SMIIc is formed using a stochastic learning method

Extension to PDE's: What is the required spatial density of inter-model coupling?

Synchronization of two 1D Kuramoto-Sivishinsky systems:

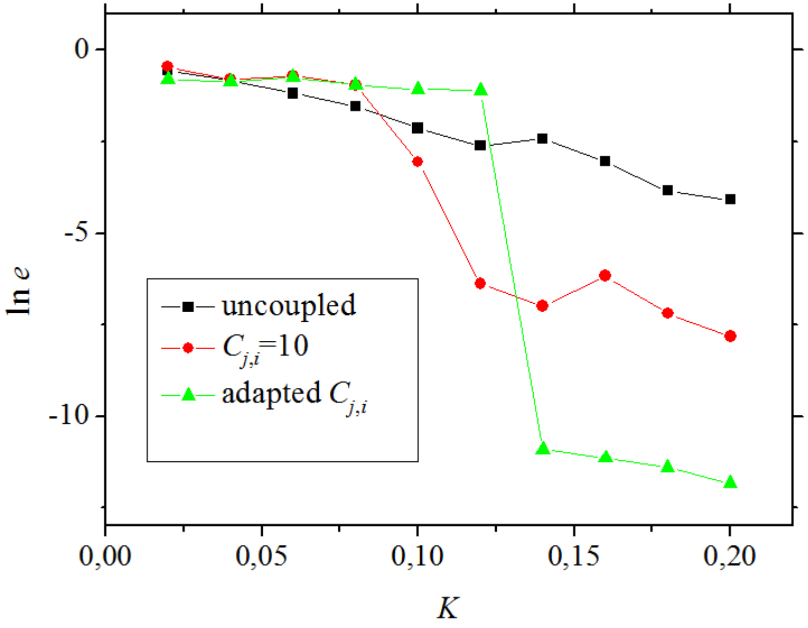
$$u_t = -u_{xxxx} - \alpha_u u_{xxx} - u_{xx} - 2uu_x$$

$$v_t = -v_{xxxx} - \alpha_v v_{xxx} - v_{xx} - 2vv_x + K[u(x) - v(x)]\, f(x)$$

$f(x)$ non-vanishing only at discrete points

Maximum coupling distance is length scale of coherent structures:

Can form supermodel from 3 KS's:



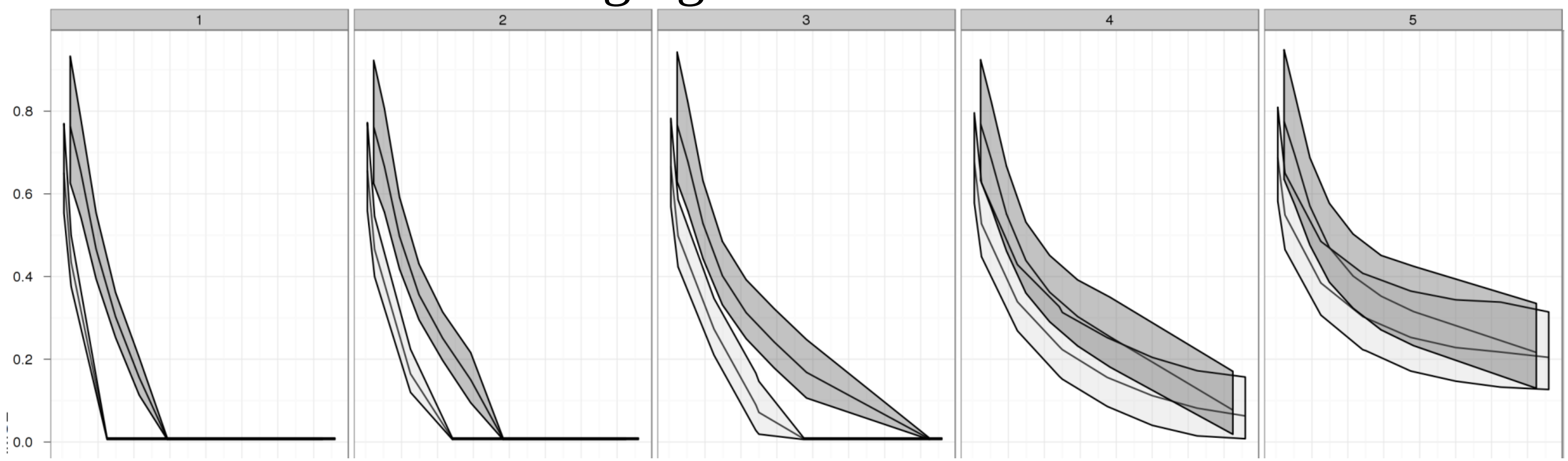
What variables should be coupled?

Consider 3-layer QG model on sphere with realistic topography and a forcing chosen to reproduce the observed winter mean state.

Compare coupling in a basis of spherical harmonics to a basis of EOF's:

nudging time scale

sync error



Number of components that are coupled

dark grey: spherical harmonics

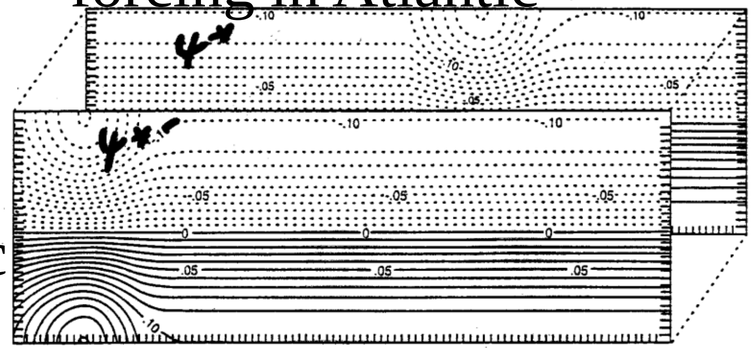
light grey: EOF's

Proposed Adaptive Fusion of Two QG Channel Models

forcing in Atlantic

$Dq/Dt = F + D$

forcing in Pacific



$Dq'/Dt = F' + D'$

$F = f_o\,(q - q^*) + c\,J(\psi, q - q')$

$F' = f_o\,(q' - q^{*\prime}) + c\,J(\psi', q' - q)$

c=1/2 $\Downarrow$

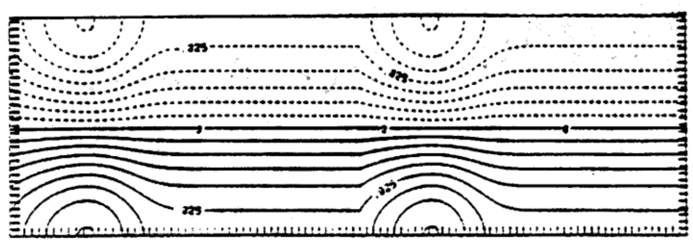
(k-dependence suppressed)

$F_o = f_o(q - q^*)$

$F_o' = f_o(q' - q^{*\prime})$

$D/Dt\,(q + q')/2 = (F_o + F_o')/2 + (D + D')/2$

- If the parallel channels synchronize, their common solution also solves the single-channel model with the average forcing



$\dfrac{\Psi^* + \Psi^{*\prime}}{2}$

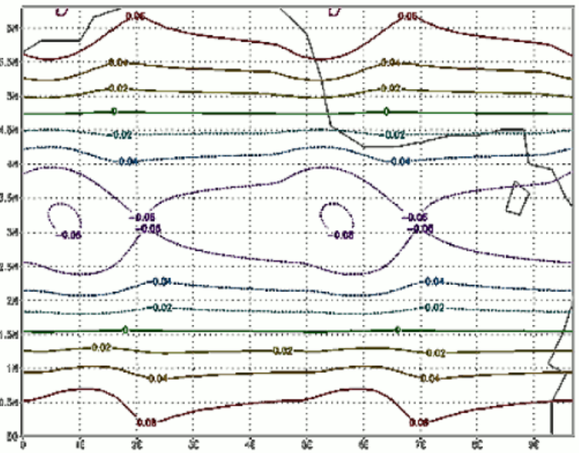
To find c adaptively:

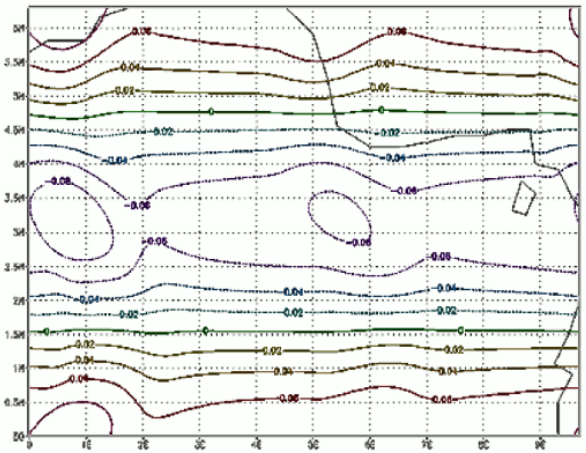
$dc/dt = \int d^2\mathbf{x}\ J(\psi, q' - q)(q - q_{obs}) + \int d^2\mathbf{x}\ J(\psi', q - q')(q' - q_{obs})$

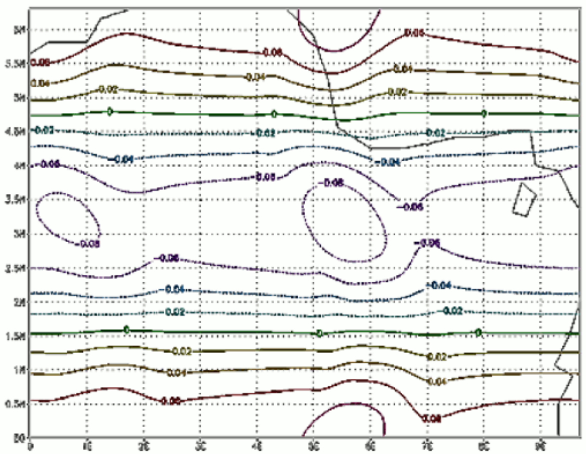
Models Synchronize With Each Other and With "Truth"

"truth"	model with Atlantic forcing	model with Pacific forcing

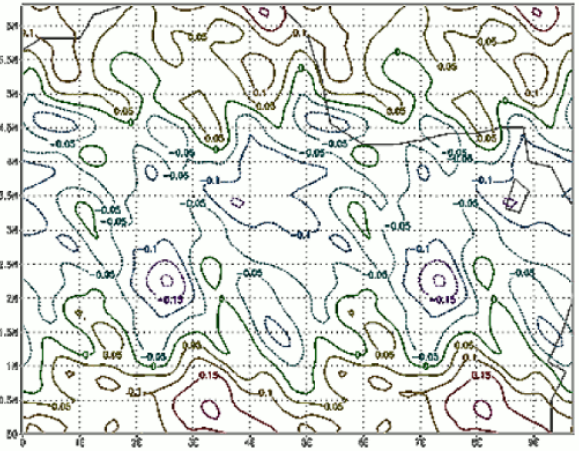
n = 1000:

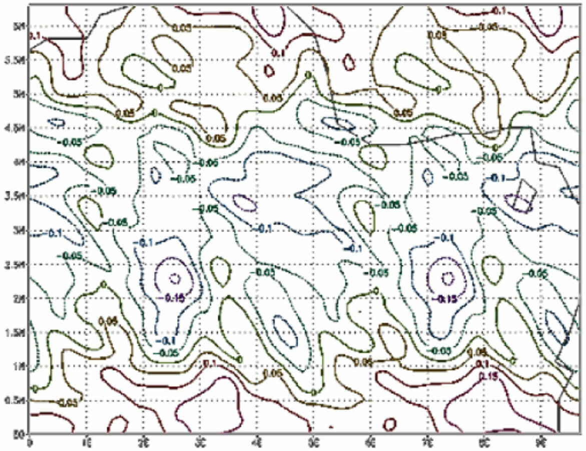


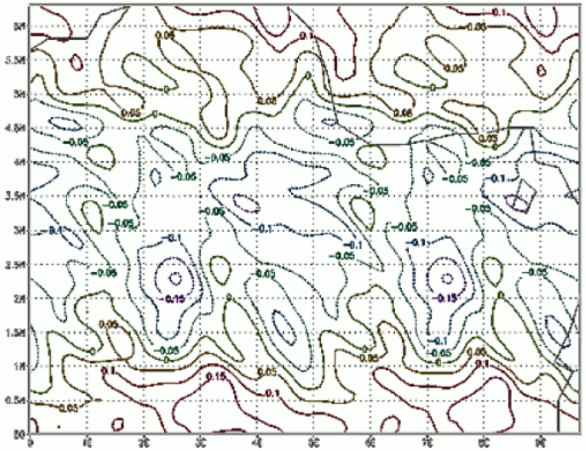




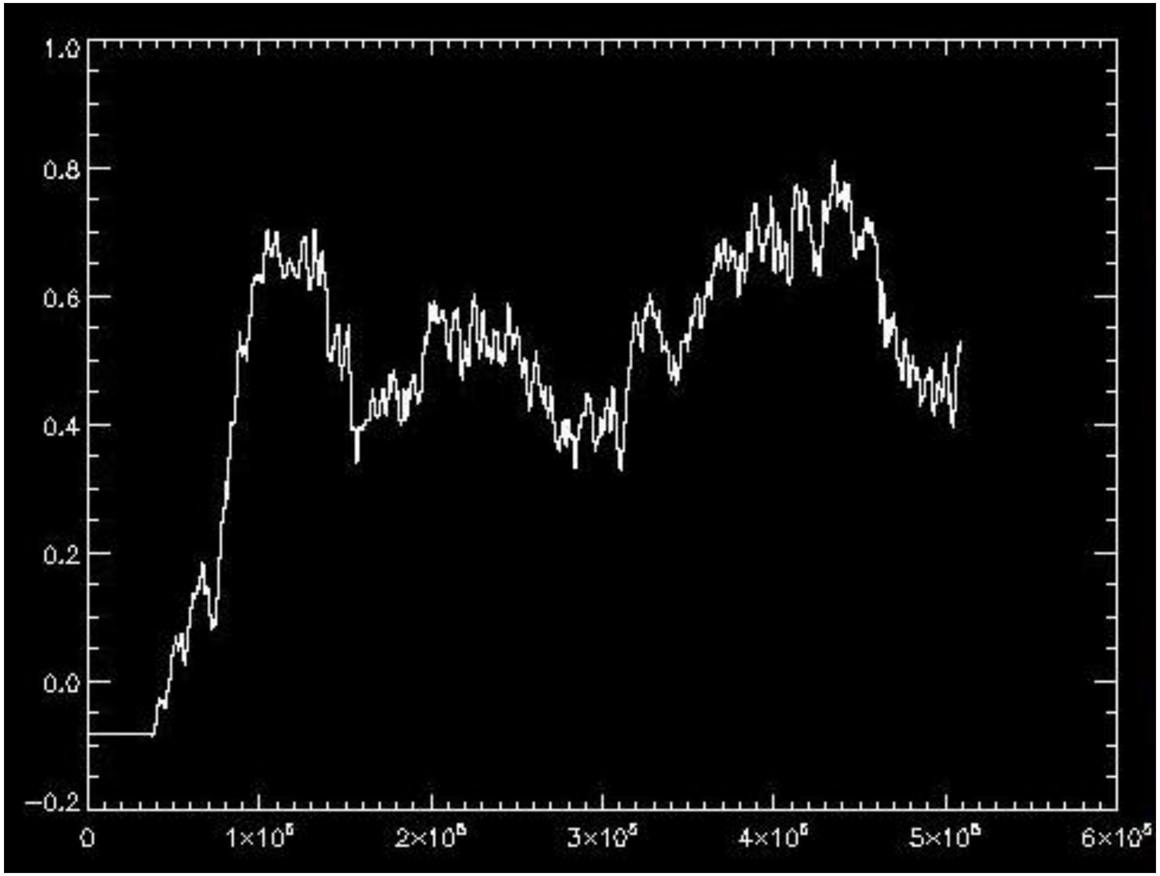
n = 30000:





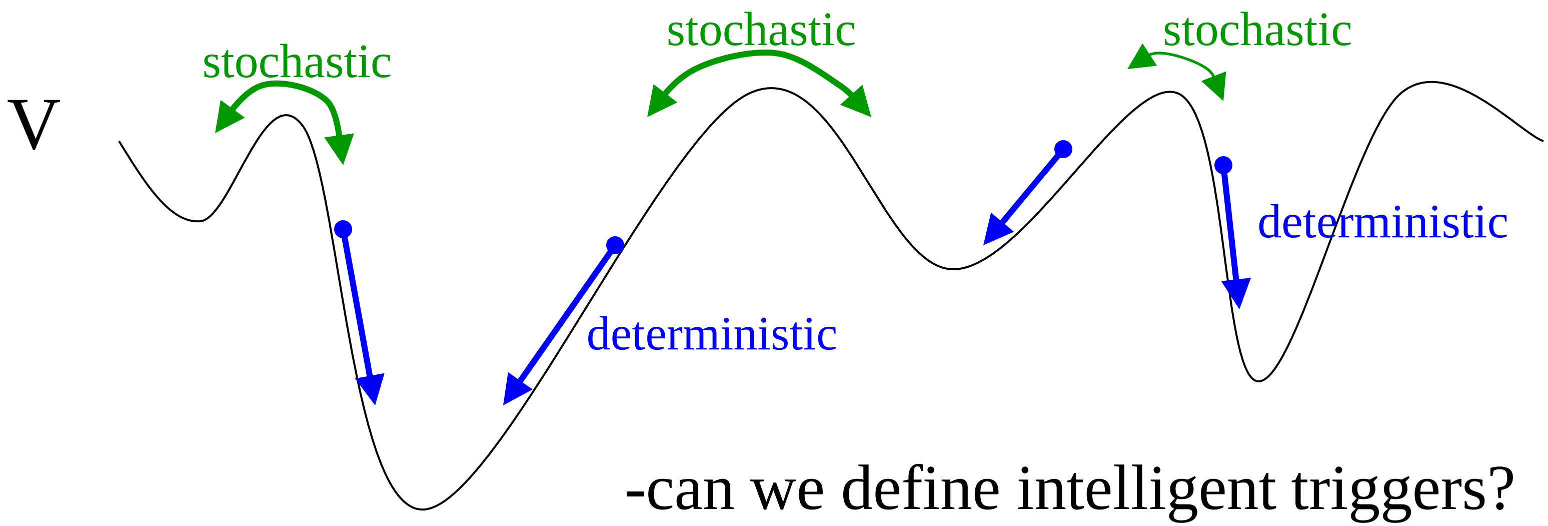


…..As the Adaptation Procedure Estimates the Intermodel Connection Coefficient $c \rightarrow 1/2$



Limits of Supermodeling

- reliance on stochasticity to escape local optima



-can we define intelligent triggers?

- Can we do better than negative coefficients when individual models err in similar ways?

- How can we restrict the number of separately trained connections ?

$\Rightarrow$ Need expert-system-like generalization of supermodeling