The impact of learning strategies for interactive ensembles in the presence of unresolved scales[1]

Wim Wiegerinck — SNN Adaptive Intelligence *and* Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, The Netherlands

Frank Selten — Royal Netherlands Meteorological Institute, De Bilt, The Netherlands

1. Supermodels (SUMOs): interactive ensemble of existing models

- Proposal to improve climate modeling[2]
  - Assumption: models are good but imperfect
  - Model combination for improvement
  - Alternative to conventional noninteractive ensemble methods
- Supermodel = Ensemble of *dynamically coupled* models

Individual model dynamics

$$\dot{x}^i_\mu = f^i_\mu(x_\mu)$$

(This poster) SUMO coupling by weighted averaging

$$\dot{x}^i = \sum_\mu w^i_\mu f^i_\mu(\boldsymbol{x})$$

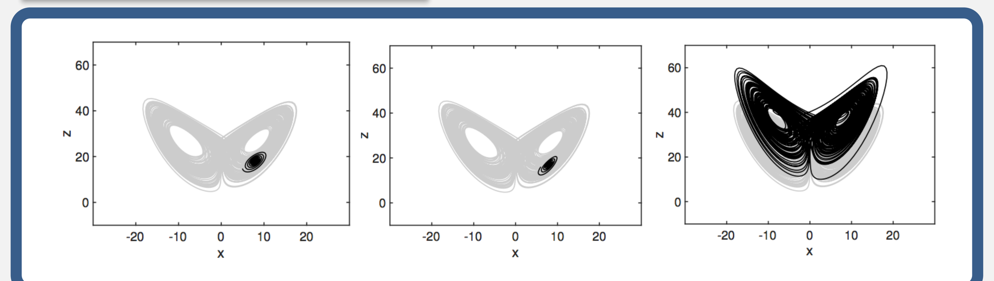
- SUMO couplings need to be optimized
  e.g. by minimizing short term prediction error E
  → Very successful in simulations, good attractors[2]

- However: perfect model class scenario
- Unrealistic assumption?
- What are the consequences
- What are alternative training methods

2. Perfect model class scenario

- Assumed Ground Truth (GT) in same model class as imperfect models (IMs)
- So IMs have imperfect model parameters
- But in the same perfect model class
- *Example:* Lorenz 63 experiment (L63) below [2]

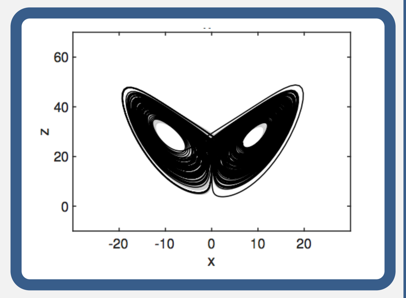
Standard L63



Top row, black: IMPERFECT MODELS - L63 with perturbed parameters

Right, black: SUMO

Grey: GROUND TRUTH - L63 with standard parameters



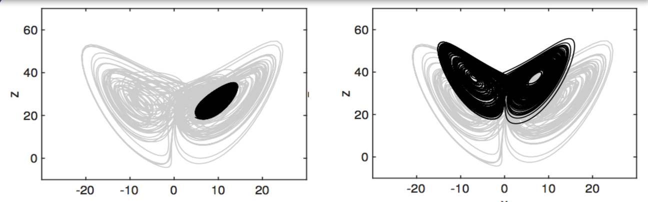
$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z\,.$$

	σ	ρ	β
Truth	10	28	8/3
Model 1	13.25	19	3.5
Model 2	7	18	3.7
Model 3	6.5	38	1.7
SUMO	9.9	29.7	3.1

3. Imperfect model class scenario

- GT is more complex than IMs (unresolved scales)
- Example: Chaotically driven L63, see below
- Short term prediction learning → wrong attractor
- Remedied by attractor learning

Chaotically driven L63



Grey: GROUND TRUTH - Chaotically driven L63

Black: Imperfect models: L63 with constant forcing -

$$\dot{x}_\mu = \sigma_\mu(y_\mu - x_\mu) + \alpha_\mu$$
$$\dot{y}_\mu = x_\mu(\rho_\mu - z_\mu) - y_\mu$$
$$\dot{z}_\mu = x_\mu y_\mu - \beta_\mu z_\mu + \gamma_\mu$$

	σ	ρ	β	α	γ
Model 1	10	28	8/3	25	0
Model 2	6.5	38	1.6	0	10

$$\left.\begin{aligned}\dot{x}_v &= \sigma(y_v - x_v) + \epsilon z_h,\\ \dot{y}_v &= x_v(\rho - z_v) - y_v,\\ \dot{z}_v &= x_v y_v - \beta z_v + \delta(x_h + \eta),\end{aligned}\right\}\ \text{visible}$$

$$\left.\begin{aligned}\dot{x}_h &= \sigma(y_h - x_h),\\ \dot{y}_h &= x_h(\rho - z_h) - y_h,\\ \dot{z}_h &= x_h y_h - \beta z_h.\end{aligned}\right\}\ \text{hidden}$$

$\sigma = 10,\ \rho = 28,\ \beta = 8/3,$

$\epsilon = 1,\ \delta = 5,\ \eta = 2.$

Short term prediction error E *- cheap to compute*

$$E(\boldsymbol{w}) = \frac{1}{K}\sum_{i=1}^{K}\sum_{t=t_i}^{t_i+\Delta T}|\boldsymbol{x}(t;\boldsymbol{w}) - \boldsymbol{x}_{gt}(t)|^2\Delta t$$

Short runs, with x initialized in x_{gt} at each t_i

Attractor errors W and V *- expensive*

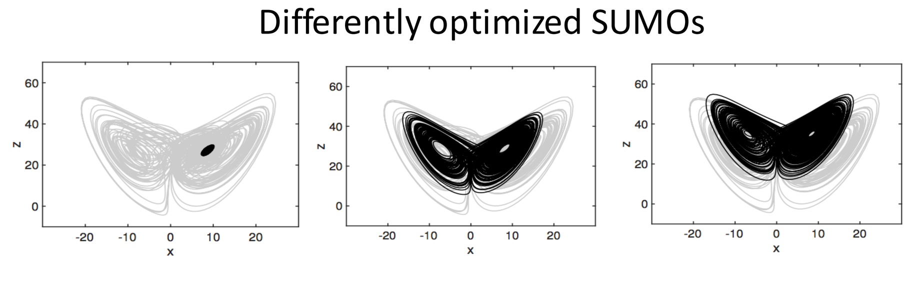
requires model roll out

$$W^2 = |\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0|^2 + \mathrm{Tr}\left(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\right)$$

$$V^2 = \mathrm{Tr}\left(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\right)$$

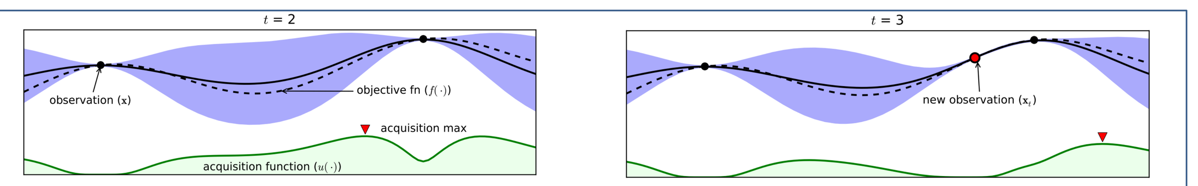
After model roll out, mean and covariances of model and data are compared

Differently optimized SUMOs



4. Bayesian optimization

- For optimization of expensive cost functions
- Models cost function as well uncertainty using Gaussian process regression
- Selects new point based on expected improvement



from: Brochu, Eric, Vlad M. Cora, and Nando De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." *arXiv preprint arXiv:1012.2599* (2010).

5. QG3 Model

- Spectral three-level quasi-geostrophic model simulating winter-time atmosphere in the Northern hemisphere (QG3), from [3]
- GT: T42 resolution model
- IMs: T21 resolution model with perturbed parameters

τ_E	Time scale in days of the Ekman damping (linear damping on vorticity at lowest level)
α_1	Parameter of the land-sea mask dependent Ekman damping (more friction over land)
α_2	Parameter of the orography dependent Ekman damping (more friction over steep slopes)
τ_R	Time scale of the radiative cooling of temperature in days
τ_h	Time scale in days of the scale selective horizontal diffusion at the three levels for the smallest wavenumber
p_h	Power of the laplacian for the scale selective diffusion, the higher the more the damping is restricted to the smallest waves
h_0	Scale height of the topography in km
R_1	Rossby radius of deformation of the 200-500 hPa layer (in earth radius units)
R_2	Rossby radius of deformation of the 500-800 hPa layer (in earth radius units)

	GT	M_1	M_2	M_3
τ_E	3.0	2.0	4.0	4.0
α_1	0.5	0.2	0.8	1.0
α_2	0.5	0.2	0.3	0.1
τ_R	25	40	20	30
τ_h	3.0	5.0	4.0	2.0
p_h	4.0	4.0	2.0	3.0
h_0	3.0	9.0	5.0	2.0
R_1	0.110	0.115	0.120	0.100
R_2	0.070	0.072	0.080	0.060

SUMO equations for potential vorticity (PV)

$$\frac{\partial q_1^s}{\partial t} = \sum_\mu w_1^\mu\left[-\boldsymbol{v}^\mu_{\psi_1^\mu}\cdot\nabla q_1^s - D_1^\mu(\psi_1^\mu,\psi_2^\mu) + S_1^\mu\right]$$

$$\frac{\partial q_2^s}{\partial t} = \sum_\mu w_2^\mu\left[-\boldsymbol{v}^\mu_{\psi_2^\mu}\cdot\nabla q_2^s - D_2^\mu(\psi_1^\mu,\psi_2^\mu,\psi_3^\mu) + S_2^\mu\right]$$

$$\frac{\partial q_3^s}{\partial t} = \sum_\mu w_3^\mu\left[-\boldsymbol{v}^\mu_{\psi_3^\mu}\cdot\nabla q_3^s - D_3^\mu(\psi_2^\mu,\psi_3^\mu) + S_3^\mu\right]$$

Velocity v and stream function ψ are calculated from the PV q by a linear transformation that is different for each model μ

Quantity of interest (in this study) STD of PV

$$m(x,y,z) = \frac{1}{T}\sum_t q(x,y,z,t)$$

$$\sigma^2(x,y,z) = \frac{1}{T}\sum_t(q(x,y,z,t) - m(x,y,z))^2$$

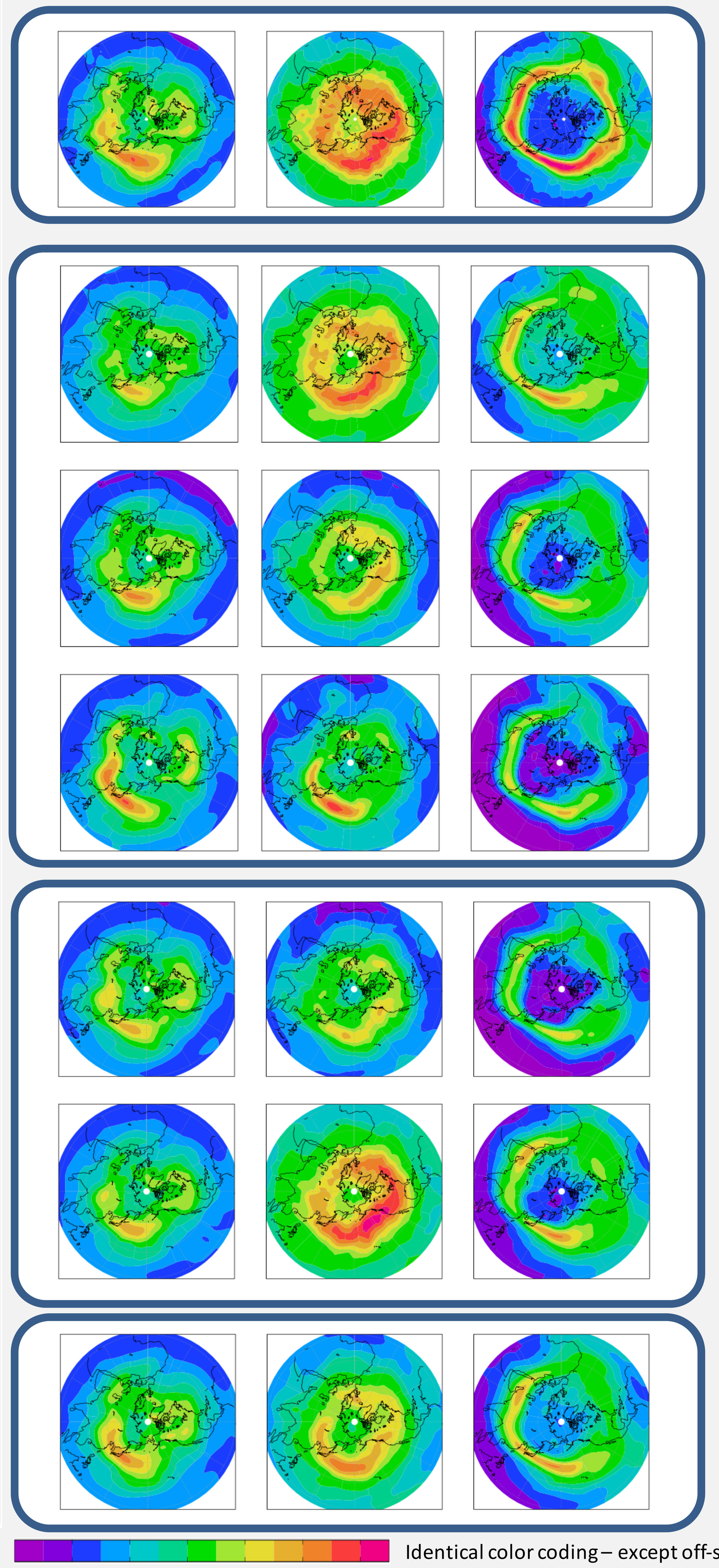
SUMOs optimized for E and U, where

$$U^2 = |\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_0|^2$$

QG3 Model

Normalized errors E and U (both x 10^7)

	E	U
GT	-	21(1)
M_1	159(2)	63(2)
M_2	159(2)	88(2)
M_3	162(2)	122(2)
SUMO(E)	151(2)	105(5)
SUMO(U)	154(2)	53(1)
PE	-	69(1)



Level-wise spatial distribution of standard deviation in potential vorticity.

6. CONCLUSIONS

- Nonlinear dynamical systems: reducing error on the level of short term prediction does not necessarily lead to improved climate properties
- Attractor learning may needed (Bayesian optimization)
- The cost function that is minimized may have a strong influence on the result
- Be careful with perfect model scenario simulations, they can give over-optimistic results

References

[1] Wiegerinck, W., and F. M. Selten. "Attractor learning in synchronized chaotic systems in the presence of unresolved scales." *Chaos: An Interdisciplinary Journal of Nonlinear Science*27.12 (2017): 126901.

[2] van den Berge, L. A., et al. "A multi-model ensemble method that combines imperfect models through learning, Earth Syst. Dynam., 2, 161–177." (2011).

[3] Marshall, John, and Franco Molteni. "Toward a dynamical understanding of planetary-scale flow regimes." *Journal of the atmospheric sciences* 50.12 (1993): 1792-1818.

European Geosciences Union | General Assembly 2018 | Vienna | Austria | 8 -13 April 2018

This work is partly funded by STERCP (ERC project 648982)