Data-based stochastic subgrid-scale modelling

Frank Kwasniok

Exeter Climate Systems (XCS),
College of Engineering, Mathematics and Physical Sciences,
University of Exeter, Exeter, UK

Workshop on
Representing model uncertainty and error
in numerical weather and climate prediction models
ECMWF, Reading, 20–24 June 2011

Outline

Outline

Stochastic climate models / Stochastic parametrisation

- Dynamics of weather and climate encompass wide range of spatial and temporal scales which are coupled through the nonlinear nature of the governing equations of motion.
- Finite computational resources limit the spatial resolution of weather and climate prediction models.
- Closure or parametrisation problem
- Can the effect of unresolved scales and processes onto the resolved scales be modelled by stochastic terms? Can the effect of model uncertainty and error be accounted for by stochastic terms?
- Time scale separation is weak in weather and climate.

Previous work

- Proposition of stochastic climate models:
  Hasselmann 1976
- Early studies:
  Lemke 1977; Egger 1981
- Linear stochastic climate models:
  Penland 1990; Farrell and Ioannou 1993; Newman et al. 1997; Whitaker and Sardeshmukh 1998; Zhang and Held 1999; Winkler et al. 2001; del Sole 2004
- Nonlinear approaches:
  Majda et al. 1999; Buizza et al. 1999; Palmer 2001; Lin and Neelin 2002; Majda and Khouider 2002; Majda et al. 2003; Shutts 2005; Plant and Craig 2008
- Purely data-driven schemes:
  Wilks 2005; Crommelin and Vanden-Eijnden 2008

Outline

The Lorenz '96 system

$$\dot{X}_k = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F + B_k$$
$$\dot{Y}_{j,k} = \frac{1}{\varepsilon}\left[Y_{j+1,k}(Y_{j-1,k} - Y_{j+2,k}) - Y_{j,k} + h_y X_k\right]$$

$k = 1, \ldots, K; \qquad j = 1, \ldots, J$

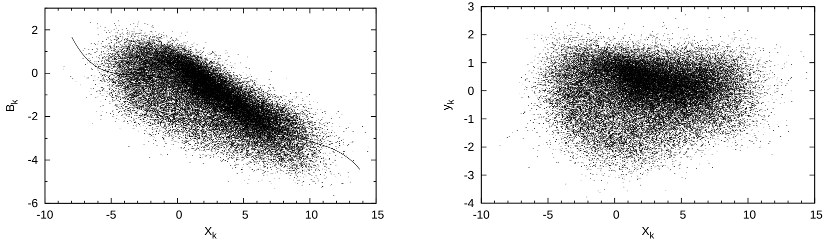
Subgrid term: $B_k = \frac{h_x}{J}\sum_{j=1}^{J} Y_{j,k}$

Cyclic boundary conditions:
$X_k = X_{k+K},\ Y_{j,k} = Y_{j,k+K},\ Y_{j+J,k} = Y_{j,k+1}$

Parameter setting:
$K = 18,\ J = 20,\ \varepsilon = 0.5,\ F = 10,\ h_x = -1,\ h_y = 1$
(*Crommelin and Vanden-Eijnden 2008*)

18 + 360 = 378 variables

Scatterplot B_k and B'_k versus X_k



- strong multiplicity of subgrid term given X_k
- conditional mean explains only about 52% of the variance

Outline

Subgrid-scale parametrisation

Climate state vector: $\mathbf{z} = (\mathbf{x}, \mathbf{y})$

True tendency of resolved variables: $\dot{\mathbf{x}} = \mathbf{R}(\mathbf{x}) + \mathbf{U}(\mathbf{x}, \mathbf{y})$

Parametrisation of unresolved tendency: $\mathbf{U}(\mathbf{x}, \mathbf{y}) \sim \mathbf{f}(\mathbf{x}) + \boldsymbol{\xi}(\mathbf{x})$

Canonical choice: $\mathbf{f}(\mathbf{x}) = \langle \mathbf{U}(\mathbf{x}, \mathbf{y}) | \mathbf{x} \rangle$

Simple parametrisations:

Deterministic: $\mathbf{U}(\mathbf{x}, \mathbf{y}) \sim \langle \mathbf{U}(\mathbf{x}, \mathbf{y}) | \mathbf{x} \rangle$

AR(1) scheme: $\mathbf{U}(\mathbf{x}, \mathbf{y}) \sim \langle \mathbf{U}(\mathbf{x}, \mathbf{y}) | \mathbf{x} \rangle$ + AR(1) process
(*Wilks 2005*)

Cluster-weighted subgrid model

Cluster-weighted modelling (*Gershenfeld et al. 1999*)

$$p(\mathbf{c}, \xi) = \sum_{m=1}^{M} w_m \, p(\mathbf{c}|m) \, p(\xi|\mathbf{v}, m)$$

Predictive probability density:

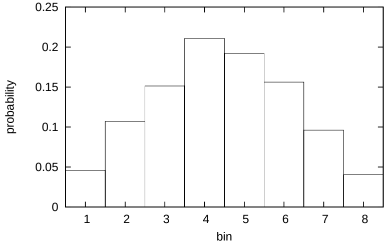
$$p(\xi|\mathbf{c}) = \sum_{m=1}^{M} g_m(\mathbf{c}) \, p(\xi|\mathbf{v}, m)$$

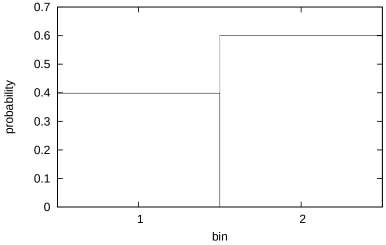
$$\text{with } g_m(\mathbf{c}) = \frac{w_m \, p(\mathbf{c}|m)}{\sum_{n=1}^{M} w_n \, p(\mathbf{c}|n)}$$

Here: Cluster-weighted Markov chain model with $\mathbf{c} = (X_k, \Delta X_k)$
(cf.,*Crommelin and Vanden-Eijnden 2008*)

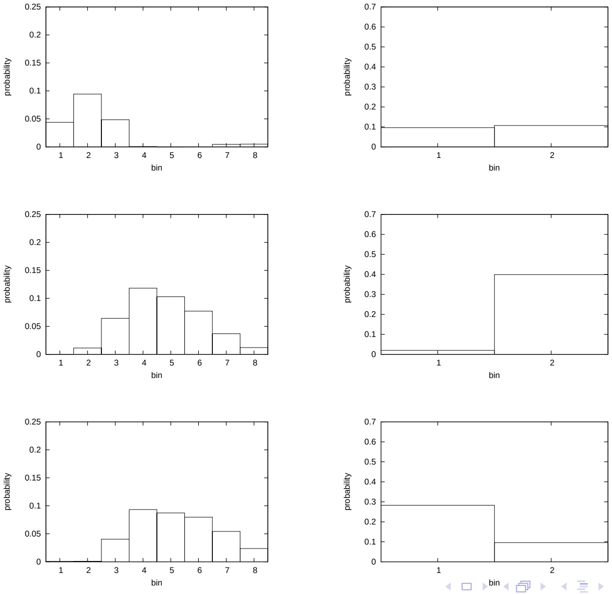
Outline

Climatology (or invariant measure) of X_k and ΔX_k

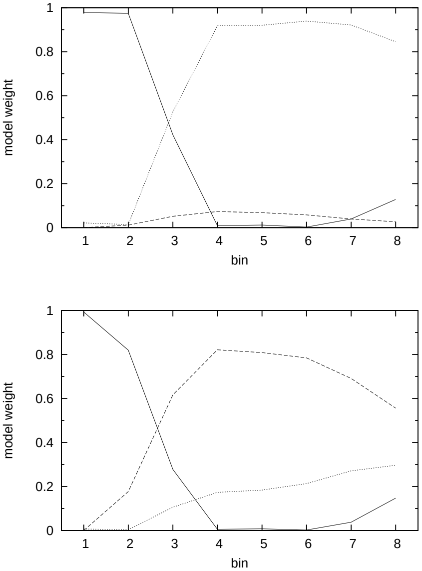




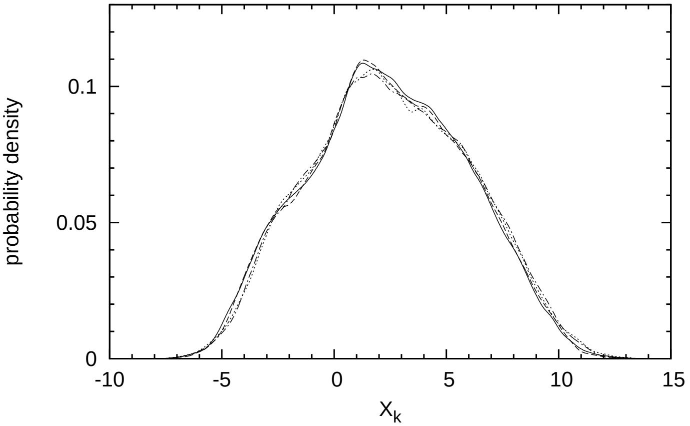
Cluster states, $w_1 = 0.203$, $w_2 = 0.418$, $w_3 = 0.379$



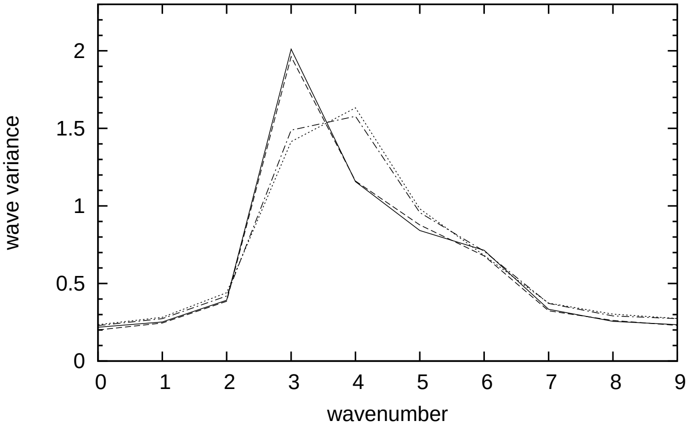
Local model weights g_m



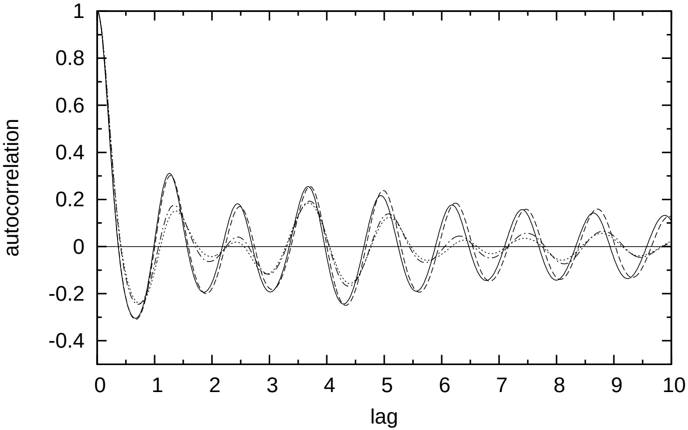
Probability density function of X_k



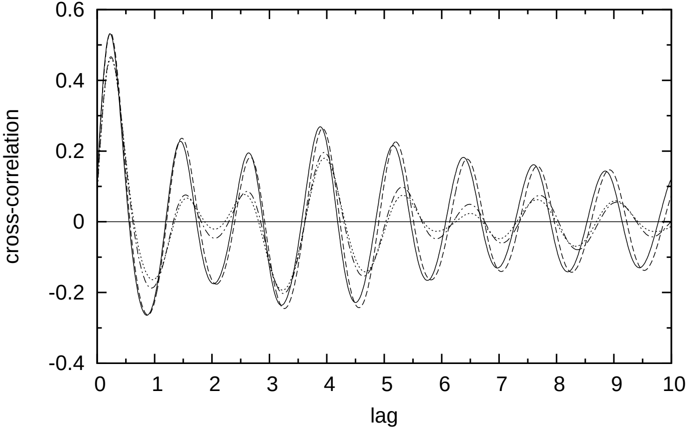
Wave variances



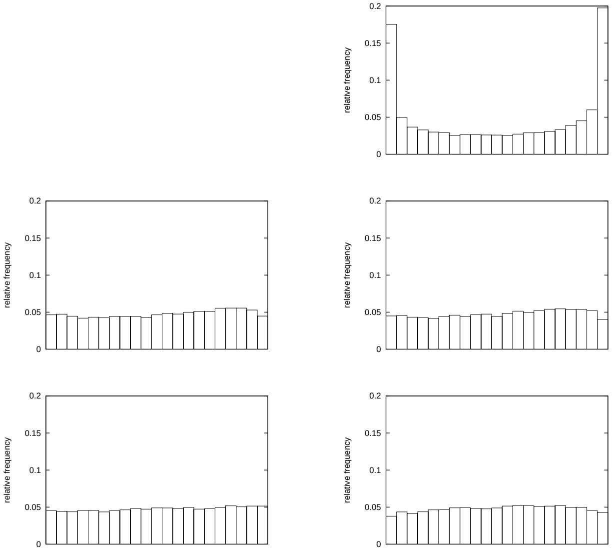
Autocorrelation function of X_k



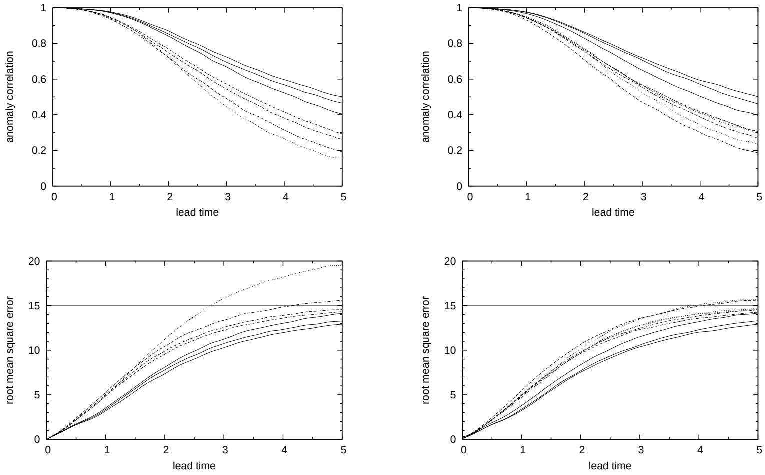
Cross-correlation function of X_k and X_{k+1}



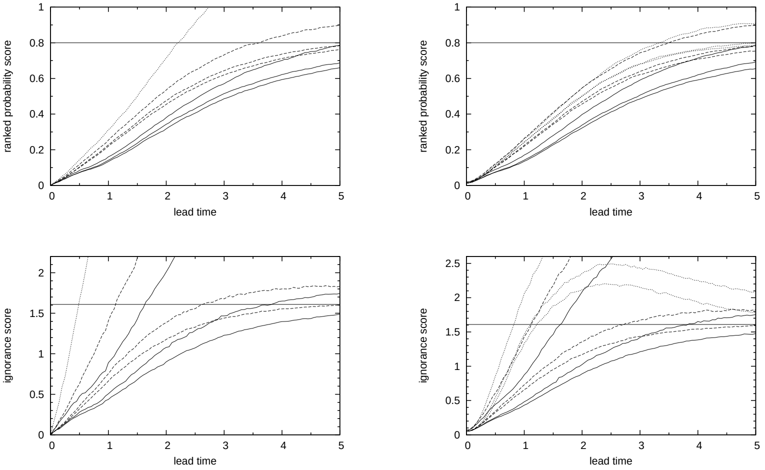
Rank histograms



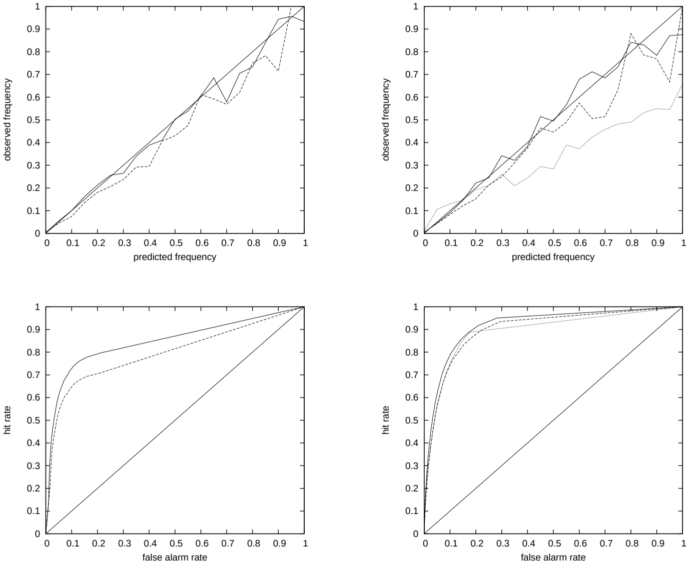
Predictive skill of the ensemble mean



Skill of probabilistic forecasts, 5 categories



Prediction of extreme events (> 95% quantile)



Remarks

- Scheme could be used locally on a grid in an atmosphere or ocean model.
- Choice of clustering variables **c**:
  - Conditioning on next higher resolved local scale.
  - Conditioning on global physically meaningful patterns.
- Combining purely data-driven predictive subgrid schemes with parametrisations based on physical reasoning or stochastic dynamical systems theory.

References

- Wilks D. S. (2005): Effects of stochastic parametrizations in the Lorenz '96 system, Quarterly Journal of the Royal Meteorological Society 131, 389–407.
- Crommelin D., Vanden-Eijnden E. (2008): Subgrid-scale parametrization with conditional Markov chains, Journal of the Atmospheric Sciences 65, 2661–2675.
- Kwasniok F. (2011): Data-based stochastic subgrid-scale parametrisation: an approach using cluster-weighted modelling, Philosophical Transactions of the Royal Society A, submitted.