Data assimilation in slow-fast systems using stochastic subgridscale forecast models

Georg Gottwald, Lewis Mitchell

University of Sydney

Durham, August 5th, 2011

Stochastic parametrizations for data assimilation

GENERAL QUESTION: Can reduced stochastic climate models be beneficial for forecasting and prediction?

Our setting here: Data assimilation with Ensemble Kalman filters

Under what circumstances and why can stochastic reduced models be beneficial as forecast models in an ensemble Kalman filter setting? Can we achieve

- computational gain
- better skill?

Stochastic homogenization (*Khasminsky '66, Kurtz '73, Papanicolaou '76*) has been recently taken up in the context of climate models (works by *Crommelin, Franzke, Majda, Timofejev, Vanden-Eijnden*).

IDEA: Consider $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$

$$dx = \frac{1}{\varepsilon} f_0(x,y)\,dt + f_1(x,y)\,dt$$
$$dy = \frac{1}{\varepsilon^2} g_0(x,y)\,dt + \frac{1}{\varepsilon}\sigma(x,y)\,dW_t$$

(For purely deterministic dynamics see *Melbourne and Stuart, Nonlinearity 2011*)

Assume the fast y-process is ergodic, and the average of f_0 over this measure is zero; then the statistics of the slow x-dynamics can be approximated in the limit $\varepsilon \to 0$ by

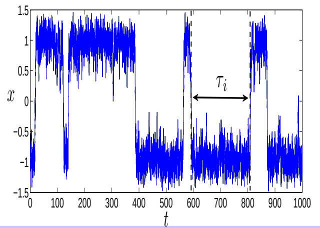
$$dX = F(X)\,dt + \Sigma(X)\,dB_t$$

Toy Model

We study the skew product system of a chaotically forced bistable system (*Givon et al., Nonlinearity* **17** *(2004)*)

$$\frac{dx}{dt} = x - x^3 + \frac{4}{90\varepsilon} y_2$$

$$\frac{dy_1}{dt} = \frac{10}{\varepsilon^2}(y_2 - y_1) \qquad \frac{dy_2}{dt} = \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1 y_3) \qquad \frac{dy_3}{dt} = \frac{1}{\varepsilon^2}\left(y_1 y_2 - \frac{8}{3} y_3\right)$$



This model is

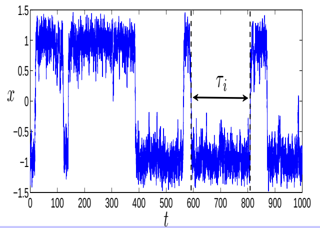
- simple
- deterministic
- and has slow as fast metastable states

Toy Model

We study the skew product system of a chaotically forced bistable system (*Givon et al., Nonlinearity* **17** *(2004)*)

$$\frac{dx}{dt} = x - x^3 + \frac{4}{90\varepsilon} y_2$$

$$\frac{dy_1}{dt} = \frac{10}{\varepsilon^2}(y_2 - y_1) \qquad \frac{dy_2}{dt} = \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1 y_3) \qquad \frac{dy_3}{dt} = \frac{1}{\varepsilon^2}\left(y_1 y_2 - \frac{8}{3} y_3\right)$$



This model is

- simple
- deterministic
- and has slow as fast metastable states

Homogenization yields

$$dx = (x - x^3)dt + \sigma dW$$

where

$$\frac{\sigma^2}{2} = -\left(\frac{4}{90}\right)^2 \int_0^\infty y_2(t) \lim_{T\to\infty} \frac{1}{T} \int_0^T y_2(t+s)\, ds\, dt$$

has to be numerically estimated.

Parameter estimation (*Siegert et al.* 1998)

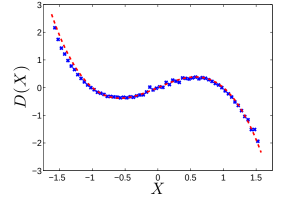
Assume that the slow dynamics of the deterministic system is modelled (on a coarse time scale) by a Langevin equation

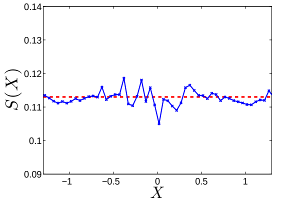
$$dx = d(x)\, dt + \sigma(x)\, dW_t$$

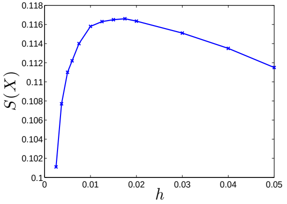
Estimate drift and diffusion from a long trajectory; partition phase space into bins $[X, X + \Delta X]$, sample at coarse sampling time $h \gg dt$

$$D(X) = \frac{1}{h}\langle (x^{n+1} - x^n)\rangle\Big|_{x^n \in (X+\Delta X)} \xrightarrow[h \to 0]{} \mathrm{d}(X)$$
$$S(X) = \frac{1}{h}\langle (x^{n+1} - x^n)^2\rangle\Big|_{x^n \in (X+\Delta X)} \xrightarrow[h \to 0]{} \sigma^2(X)$$

Parameter estimation (*Siegert et al.* 1998)



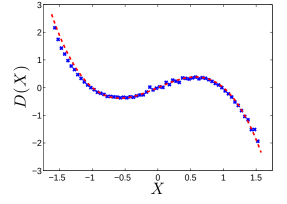


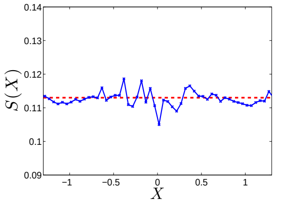


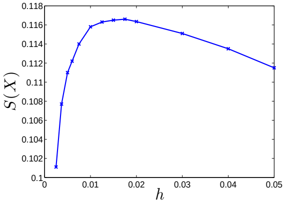
$$dx = d(x)\,dt + \sigma(x)\,dW_t$$

(Sensitivity of the estimated coefficients to subsampling time; implies uncertainty of climate model!)

Parameter estimation (*Siegert et al.* 1998)





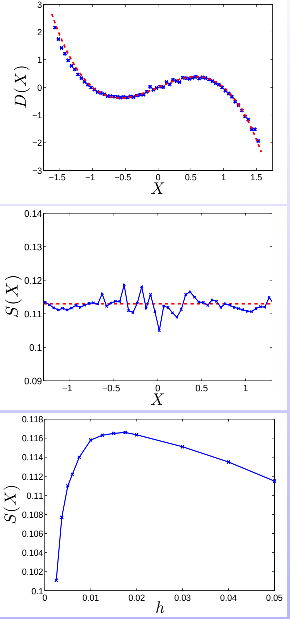


$$dx = d(x)\,dt + \sigma(x)\,dW_t$$

How to choose the subsampling time h?

(Sensitivity of the estimated coefficients to subsampling time; implies uncertainty of climate model!)

Parameter estimation (*Siegert et al.* 1998)



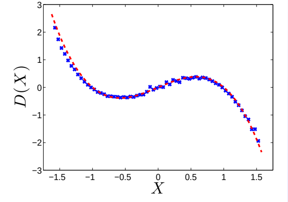
$$dx = d(x)\,dt + \sigma(x)\,dW_t$$

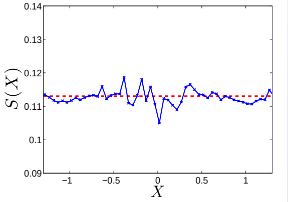
How to choose the subsampling time h?

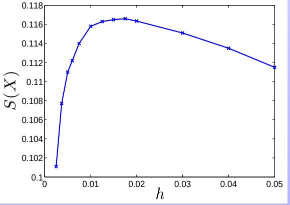
- If h is chosen too small, diffusion coefficient does not exist
- If h is chosen too large,
  - $S(X) \approx X^2(1-X^2)^2\,h^2 + \sigma^2 \xrightarrow[h\to 0]{} \mathrm{d}^2(X)$
  - $D(X) \approx \frac{1}{h}\int (x - X)\hat{\rho}(x)\,dx = -\frac{X}{h}$

(Sensitivity of the estimated coefficients to subsampling time; implies uncertainty of climate model!)

Parameter estimation (*Siegert et al.* 1998)







$$dx = d(x)\,dt + \sigma(x)\,dW_t$$

How to choose the subsampling time h?

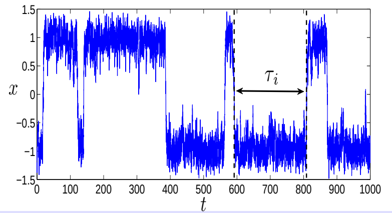
- If h is chosen too small, diffusion coefficient does not exist
- If h is chosen too large,
  - $S(X) \approx X^2(1-X^2)^2\,h^2 + \sigma^2 \xrightarrow[h\to 0]{} \mathrm{d}^2(X)$
  - $D(X) \approx \frac{1}{h}\int (x - X)\hat{\rho}(x)\,dx = -\frac{X}{h}$

Rule of thumb: Choose $h \approx 3T_f$, where T_f is the characteristic time of the fast dynamics

$$\sigma^2 = 0.113$$

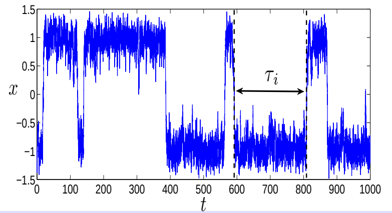
(Sensitivity of the estimated coefficients to subsampling time; implies uncertainty of climate model!)

Characteristic time scales



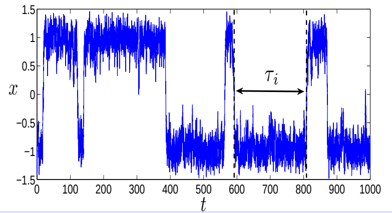
- Autocorrelation time τ_{corr}
  - $C(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(s)x(\tau+s)\,ds$: $\quad (\tau_{\text{corr}} = 208)$

Characteristic time scales



- Autocorrelation time $\tau_{\rm corr}$
  - $C(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(s)x(\tau+s)\,ds$: $(\tau_{\rm corr} = 208)$
- Mean sojourn time $\bar{\tau}$ (Mean exit time $\tau_e = \bar{\tau}/2$)
  - average over individual τ_i: $(\bar{\tau} = 218)$
  - assume Poisson process $P_c(\tau_i) = 1 - \exp\left(-\frac{\tau_i}{\bar{\tau}}\right)$: $(\bar{\tau} = 214)$
  - homogenized model: $\mathcal{L}_{\rm clim}\bar{\tau} = -2$: $(\bar{\tau} = 234)$

Characteristic time scales



- Autocorrelation time τ_{corr}
  - $C(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(s)x(\tau+s)\,ds$: $(\tau_{\text{corr}} = 208)$
- Mean sojourn time $\bar{\tau}$ (Mean exit time $\tau_e = \bar{\tau}/2$)
  - average over individual τ_i: $(\bar{\tau} = 218)$
  - assume Poisson process $P_c(\tau_i) = 1 - \exp\left(-\frac{\tau_i}{\bar{\tau}}\right)$: $(\bar{\tau} = 214)$
  - homogenized model: $\mathcal{L}_{\text{clim}}\bar{\tau} = -2$: $(\bar{\tau} = 234)$
- Mean transit time $\bar{\tau}_t$
  - average over individual $\tau_{t,i}$: $(\bar{\tau}_t = 5.9)$
  - homogenized model: $(\bar{\tau} = 5.66)$

Accuracy and sensitivity of homogenized model

How good are climate models to reproduce the statistics?

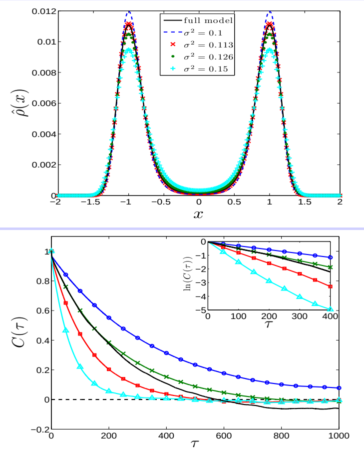
Accuracy and sensitivity of homogenized model

How good are climate models to reproduce the statistics?

In the limit $\varepsilon \longrightarrow 0$ there are rigorous theorems (*Kurtz '73, Papanicolaou '76, Melbourne & Stuart '11*)

How good are climate models to reproduce the statistics?

In the limit $\varepsilon \to 0$ there are rigorous theorems (*Kurtz '73, Papanicolaou '76, Melbourne & Stuart '11*)

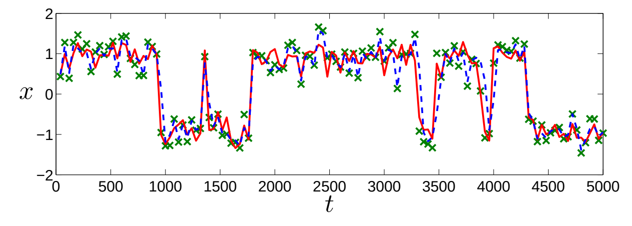


The climate model is not sensitive to uncertainties in $\varepsilon < 0.05$ but very sensitive to changes in drift and diffusion coefficients:

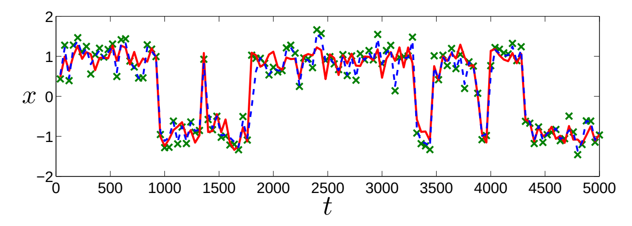
	full model	climate model ($\sigma^2 = 0.1$)	climate model ($\sigma^2 = 0.113$)	climate model ($\sigma^2 = 0.126$)	climate model ($\sigma^2 = 0.15$)
$\tau_{\rm corr}$	208.3	353.9	221.7	129.0	70.5
τ_e	108.6	205.7	117.8	75.6	40.8
τ_t	5.9	5.86	5.66	5.48	5.17
$\lambda_{\rm max}^{-1}$	0.0103	n.a.	n.a.	n.a	n.a
λ_{LS}^{-1}	233.7	588.2	398.6	206.4	108.6

Numerical results

Use the climate model as a forecast model in an Ensemble Transform Kalman filter (ETKF) setting (*Tippet et al. 2003*). Only the slow variable x is observed.

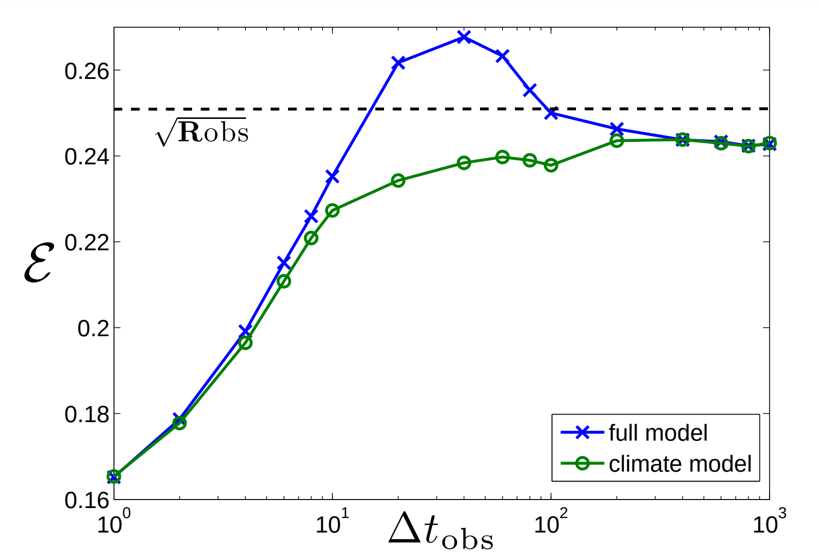


Full deterministic model



Reduced climate model ($\sigma^2 = 0.126 > 0.113$)

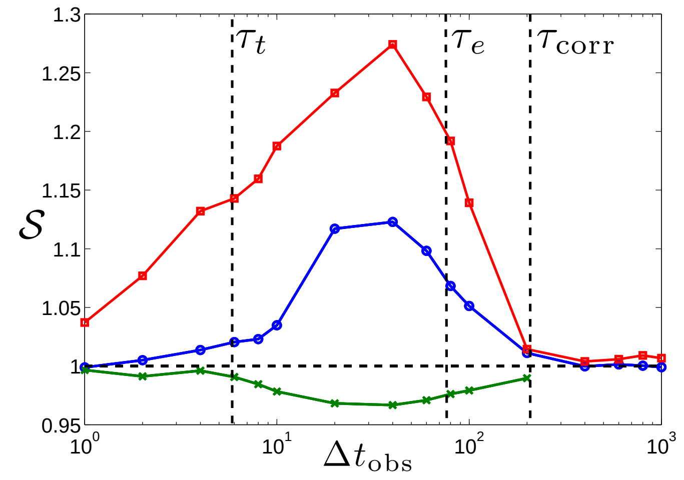
Numerical results



Numerical results

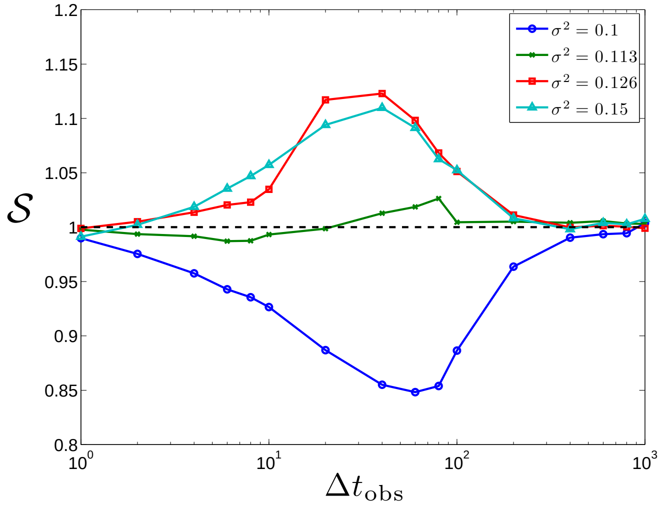
We define the *skill* $\mathcal{S} = \frac{\mathcal{E}_{\text{full}}}{\mathcal{E}_{\text{climate}}}$, $\mathcal{S} > 1$ is good!

Blue - all analyses, Green - metastable states, Red - transitions



Numerical results

$\sigma^2 = 0.1$, $\sigma^2 = 0.113$, $\sigma^2 = 0.126$, $\sigma^2 = 0.15$



Numerical results

The numerical results suggests that stochastic climate models are

- beneficial for observation time intervals $\Delta t_{\mathrm{obs}} \in (\tau_t, \tau_{\mathrm{corr}})$
- good at capturing the transitions between slow metastable states
- perform better than full system for diffusion larger than the "correct" value: $\sigma^2 > 0.113$

Numerical results

The numerical results suggests that stochastic climate models are

- beneficial for observation time intervals $\Delta t_{\text{obs}} \in (\tau_t, \tau_{\text{corr}})$
- good at capturing the transitions between slow metastable states
- perform better than full system for diffusion larger than the "correct" value: $\sigma^2 > 0.113$

So why, if the climate model fails to accurately reproduce the statistics of the full model, does it perform better?

Numerical results

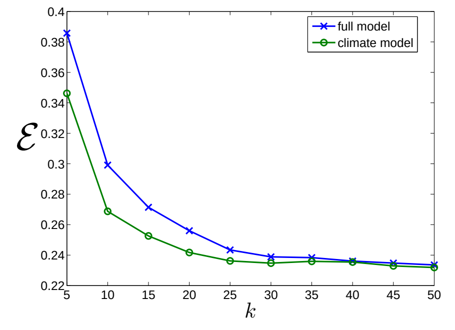
Small ensemble sizes $\longrightarrow$ underestimation of $\mathbf{P}_f$ $\nearrow$ larger ensembles

$\searrow$ covariance inflation

Numerical results

Small ensemble sizes $\longrightarrow$ underestimation of $\mathbf{P}_f$ $\nearrow$ larger ensembles / $\searrow$ covariance inflation

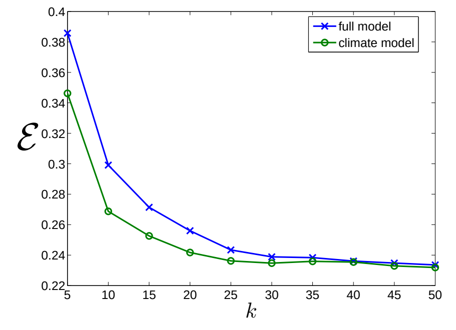
Increasing ensemble size k:



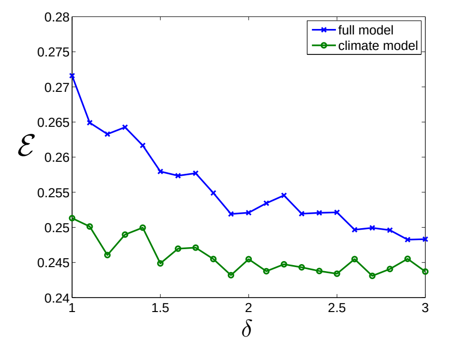
Numerical results

Small ensemble sizes $\longrightarrow$ underestimation of $\mathbf{P}_f$ $\nearrow$ larger ensembles $\searrow$ covariance inflation

Increasing ensemble size k:



Increasing covariance inflation δ:



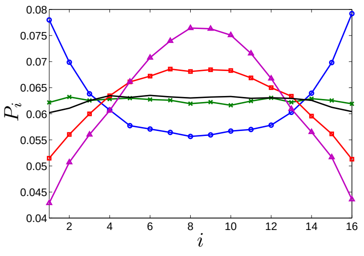
Reliability and Talagrand diagrams

- sort the forecast ensemble $\mathbf{X}_f = [x_{f,1}, x_{f,2}, ..., x_{f,k}]$ and create bins $(-\infty, x_{f,1}]$, $(x_{f,1}, x_{f,2}]$, ... , $(x_{f,k}, \infty)$ at each forecast step
- increment whichever bin the actual truth falls into at each forecast step

Convex histogram: underestimating ensemble
Concave histogram: overestimating ensemble
Flat histogram: reliable ensemble for which each ensemble member has equal probability of being nearest to the truth

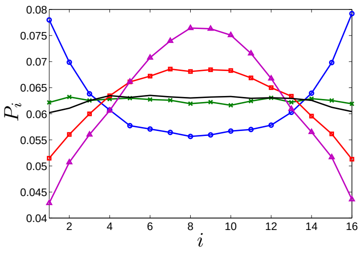
Reliability and Talagrand diagrams

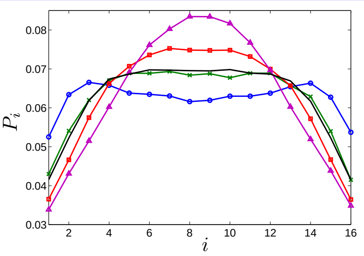
$\sigma^2 = 0.1$, $\sigma^2 = 0.113$, $\sigma^2 = 0.126$, $\sigma^2 = 0.15$



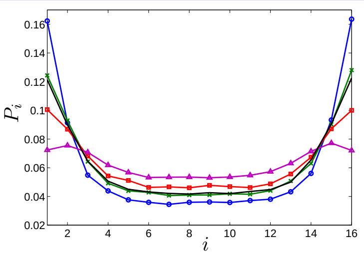
Reliability and Talagrand diagrams

$$\sigma^2 = 0.1,\ \sigma^2 = 0.113,\ \sigma^2 = 0.126,\ \sigma^2 = 0.15$$





Metastable states



Transitions

Summary and outlook

We have

- studied a homogenized stochastic climate model from a chaotic deterministic model with slow and fast metastable states (regimes)

Summary and outlook

We have

- studied a homogenized stochastic climate model from a chaotic deterministic model with slow and fast metastable states (regimes)
- shown that the effectiveness of the stochastic climate model to reproduce the full dynamics is very sensitive to the model parameters

Summary and outlook

We have

- studied a homogenized stochastic climate model from a chaotic deterministic model with slow and fast metastable states (regimes)
- shown that the effectiveness of the stochastic climate model to reproduce the full dynamics is very sensitive to the model parameters
- shown that the skill in an ensemble data assimilation scheme is not due to the accuracy of the model but is a finite size effect and is entirely due to the increased variance of a stochastic model which acts as forecast error covariance inflation

Summary and outlook

We have

- studied a homogenized stochastic climate model from a chaotic deterministic model with slow and fast metastable states (regimes)
- shown that the effectiveness of the stochastic climate model to reproduce the full dynamics is very sensitive to the model parameters
- shown that the skill in an ensemble data assimilation scheme is not due to the accuracy of the model but is a finite size effect and is entirely due to the increased variance of a stochastic model which acts as forecast error covariance inflation

We would like to

- explore the usefulness of climate models in more realistic settings

Summary and outlook

We have

- studied a homogenized stochastic climate model from a chaotic deterministic model with slow and fast metastable states (regimes)
- shown that the effectiveness of the stochastic climate model to reproduce the full dynamics is very sensitive to the model parameters
- shown that the skill in an ensemble data assimilation scheme is not due to the accuracy of the model but is a finite size effect and is entirely due to the increased variance of a stochastic model which acts as forecast error covariance inflation

We would like to

- explore the usefulness of climate models in more realistic settings
- study the effectiveness of stochastic climate models in other data assimilation schemes