Statistical properties of deterministic dynamical systems and their applications in weather and climate forecasting

Georg Gottwald

joint work with Jason Frank, Brent Giggins, John Harlim, Ian Melbourne, Lewis Mitchell, Karsten Peters, Caroline Wormell and Jeroen Wouters

THE UNIVERSITY OF SYDNEY

Colloquium (somewhere) 29 May 2020

Statistical limit laws for deterministic dynamical systems

I) Heuristics and a few theorems

II) Some applications

- data assimilation
- ensemble forecasting - bred vectors
- sensitivity to perturbations - Linear Response Theory
- numerical integration of deterministic multi-scale systems
- parametrisation of tropical convection

Motivation for stochastic parametrisation:

- prediction: computational cost in running model

$$\left.\begin{aligned} \dot{x} &= f(x,y) \\ \dot{y} &= \frac{1}{\varepsilon} g(x,y) \\ \\ x &\in \mathbb{R}^n \\ y &\in \mathbb{R}^{m} \\ \\ \varepsilon &\ll 1 \end{aligned}\right\}$$

stiff high-dimensional deterministic
multi-scale problem

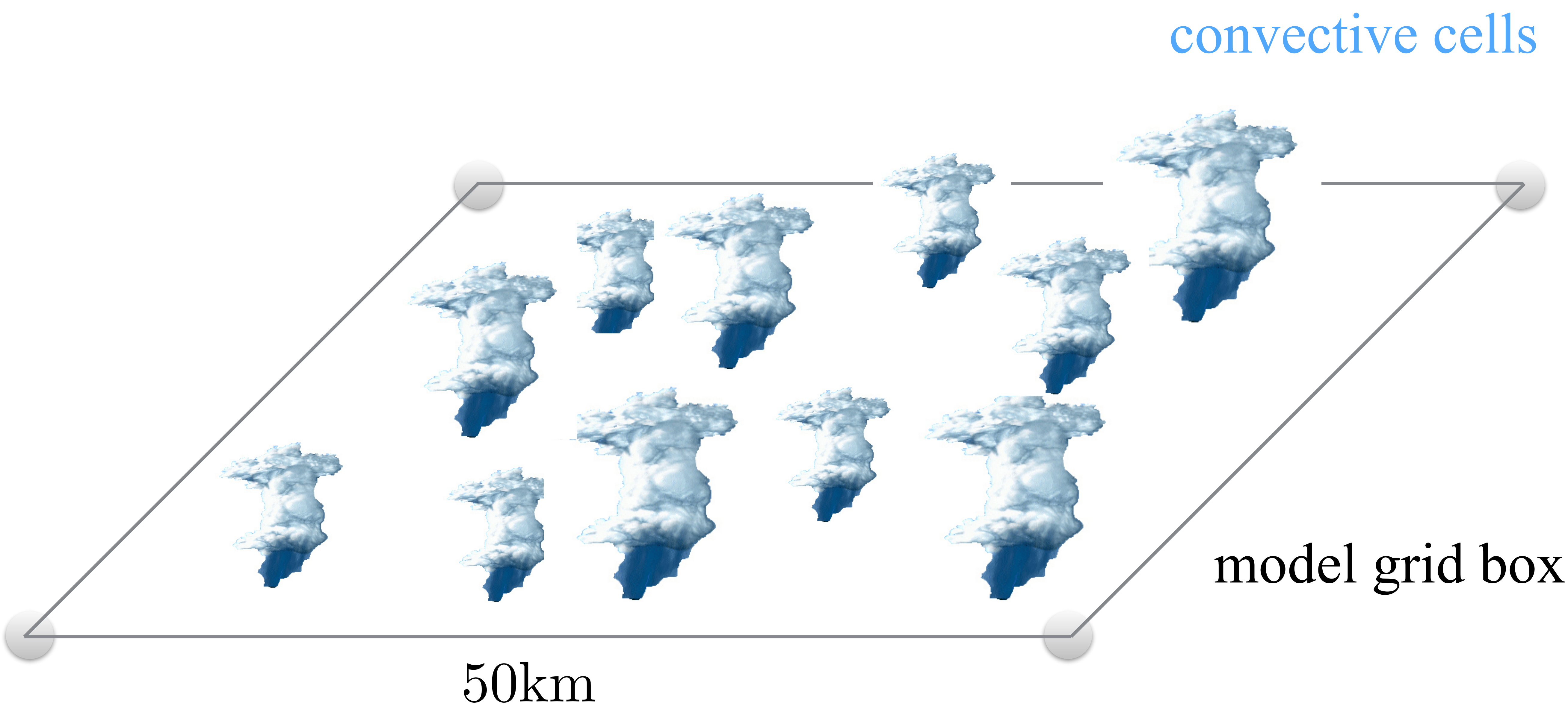
$$dX = F(X)dt + \Sigma\, dW_t$$

$$X \in \mathbb{R}^n$$

lower-dimensional stochastic problem

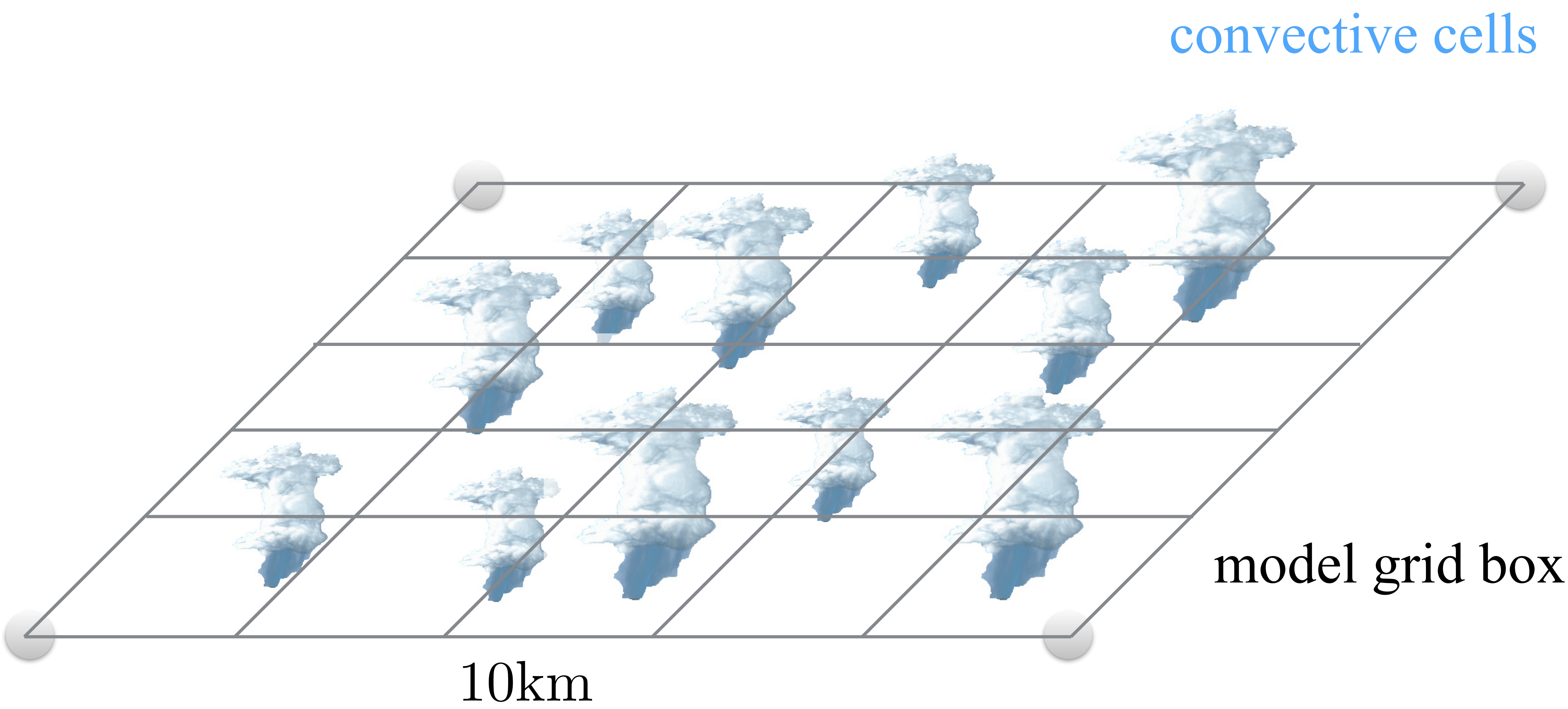
Motivation for stochastic parametrisation:

- prediction: computational cost in running model
- increase of resolution necessitates stochastic approach



Motivation for stochastic parametrisation:

- prediction: computational cost in running model
- increase of resolution necessitates stochastic approach



Heuristics for why the fast process can be replaced by noise

$$
\begin{aligned}
dx^{(\varepsilon)} &= f(x^{(\varepsilon)}, y^{(\varepsilon)})\, dt \\
dy^{(\varepsilon)} &= \frac{1}{\varepsilon} g(y^{(\varepsilon)})\, dt
\end{aligned}
$$

Heuristics for why the fast process can be replaced by noise

$$
\begin{aligned}
dx^{(\varepsilon)} &= f(x^{(\varepsilon)}, y^{(\varepsilon)})\, dt \\
dy^{(\varepsilon)} &= \frac{1}{\varepsilon} g(y^{(\varepsilon)})\, dt
\end{aligned}
$$

Integrate the slow equation

$$
\begin{aligned}
x^{(\varepsilon)}(t) &= x^{(\varepsilon)}(0) + \int_0^t f(x^{(\varepsilon)}, y^{(\varepsilon)}(s))\, ds \\
&= x^{(\varepsilon)}(0) + \varepsilon \int_0^{\frac{t}{\varepsilon}} f(x^{(\varepsilon)}, y^{(\varepsilon=1)}(\tau))\, d\tau \\
&= x^{(\varepsilon)}(0) + \frac{1}{n} \int_0^{nt} f(x^{(\varepsilon)}, y^{(\varepsilon=1)}(\tau))\, d\tau
\end{aligned}
$$

Heuristics for why the fast process can be replaced by noise

$$dx^{(\varepsilon)} = f(x^{(\varepsilon)}, y^{(\varepsilon)})\, dt$$
$$dy^{(\varepsilon)} = \frac{1}{\varepsilon} g(y^{(\varepsilon)})\, dt$$

Integrate the slow equation

$$x^{(\varepsilon)}(t) = x^{(\varepsilon)}(0) + \int_0^t f(x^{(\varepsilon)}, y^{(\varepsilon)}(s))\, ds$$
$$= x^{(\varepsilon)}(0) + \varepsilon \int_0^{\frac{t}{\varepsilon}} f(x^{(\varepsilon)}, y^{(\varepsilon=1)}(\tau))\, d\tau$$
$$= x^{(\varepsilon)}(0) + \frac{1}{n} \int_0^{nt} f(x^{(\varepsilon)}, y^{(\varepsilon=1)}(\tau))\, d\tau$$

Envoking Birkhoff's Ergodic Theorem

$$X(t) = X(0) + \int_0^t F(X(s))\, ds$$
$$F(X) = \int f(x, y)\, \mu(dy)$$

Averaged deterministic dynamics

law of large numbers

Heuristics for why the fast process can be replaced by noise

$$dx^{(\varepsilon)} = f(x^{(\varepsilon)}, y^{(\varepsilon)})\, dt$$
$$dy^{(\varepsilon)} = \frac{1}{\varepsilon} g(y^{(\varepsilon)})\, dt$$

Integrate the slow equation

$$\begin{aligned} x^{(\varepsilon)}(t) &= x^{(\varepsilon)}(0) + \int_0^t f(x^{(\varepsilon)}, y^{(\varepsilon)}(s))\, ds \\ &= x^{(\varepsilon)}(0) + \varepsilon \int_0^{\frac{t}{\varepsilon}} f(x^{(\varepsilon)}, y^{(\varepsilon=1)}(\tau))\, d\tau \\ &= x^{(\varepsilon)}(0) + \frac{1}{n} \int_0^{nt} f(x^{(\varepsilon)}, y^{(\varepsilon=1)}(\tau))\, d\tau \end{aligned}$$

Envoking Birkhoff's Ergodic Theorem

$$X(t) = X(0) + \int_0^t F(X(s))\, ds$$
$$F(X) = \int f(x, y)\, \mu(dy)$$

Averaged deterministic dynamics

law of large numbers

Heuristics for why the fast process can be replaced by noise

$$dx^{(\varepsilon)} = f(x^{(\varepsilon)}, y^{(\varepsilon)})\, dt$$

$$dy^{(\varepsilon)} = \frac{1}{\varepsilon} g(y^{(\varepsilon)})\, dt$$

Heuristics for why the fast process can be replaced by noise

$$dx^{(\varepsilon)} = \frac{1}{\varepsilon} f(y^{(\varepsilon)})\, dt$$

$$dy^{(\varepsilon)} = \frac{1}{\varepsilon^2} g(y^{(\varepsilon)})\, dt$$

go to long *diffusive* time scale

Heuristics for why the fast process can be replaced by noise

$$dx^{(\varepsilon)} = \frac{1}{\varepsilon} f(y^{(\varepsilon)})\, dt$$
$$dy^{(\varepsilon)} = \frac{1}{\varepsilon^2} g(y^{(\varepsilon)})\, dt$$

Integrate the slow equation

$$\begin{aligned} x^{(\varepsilon)}(t) &= x^{(\varepsilon)}(0) + \frac{1}{\varepsilon} \int_0^t f(y^{(\varepsilon)}(s))\, ds \\ &= x^{(\varepsilon)}(0) + \varepsilon \int_0^{\frac{t}{\varepsilon^2}} f(y^{(\varepsilon=1)}(\tau))\, d\tau \\ &= x^{(\varepsilon)}(0) + \frac{1}{\sqrt{n}} \int_0^{nt} f(y^{(\varepsilon=1)}(\tau))\, d\tau \end{aligned}$$

Heuristics for why the fast process can be replaced by noise

$$dx^{(\varepsilon)} = \frac{1}{\varepsilon} f(y^{(\varepsilon)})\, dt$$

$$dy^{(\varepsilon)} = \frac{1}{\varepsilon^2} g(y^{(\varepsilon)})\, dt$$

Integrate the slow equation

$$x^{(\varepsilon)}(t) = x^{(\varepsilon)}(0) + \frac{1}{\varepsilon}\int_0^t f(y^{(\varepsilon)}(s))\, ds$$

$$= x^{(\varepsilon)}(0) + \varepsilon \int_0^{\frac{t}{\varepsilon^2}} f(y^{(\varepsilon=1)}(\tau))\, d\tau$$

$$= x^{(\varepsilon)}(0) + \frac{1}{\sqrt{n}}\int_0^{nt} f(y^{(\varepsilon=1)}(\tau))\, d\tau$$

Assuming $\int f(y)\mu(dy) = 0$ and invoking the Central Limit Theorem

$$X(t) = X(0) + W_t$$

$$dX = dW_t$$

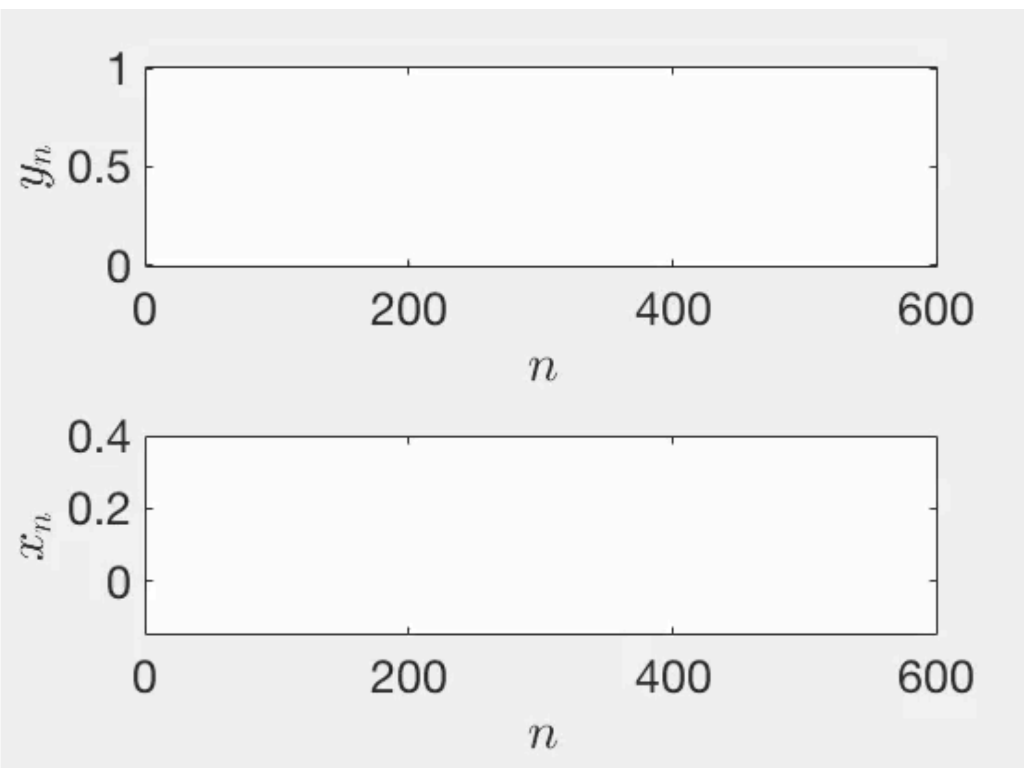
Homogenised stochastic equation

central limit theorem

Homogenisation in action

$$x_{n+1} = x_n + \varepsilon(y_n - \tfrac{1}{2})$$
$$y_{n+1} = 4y_n(1 - y_n)$$

strong chaos

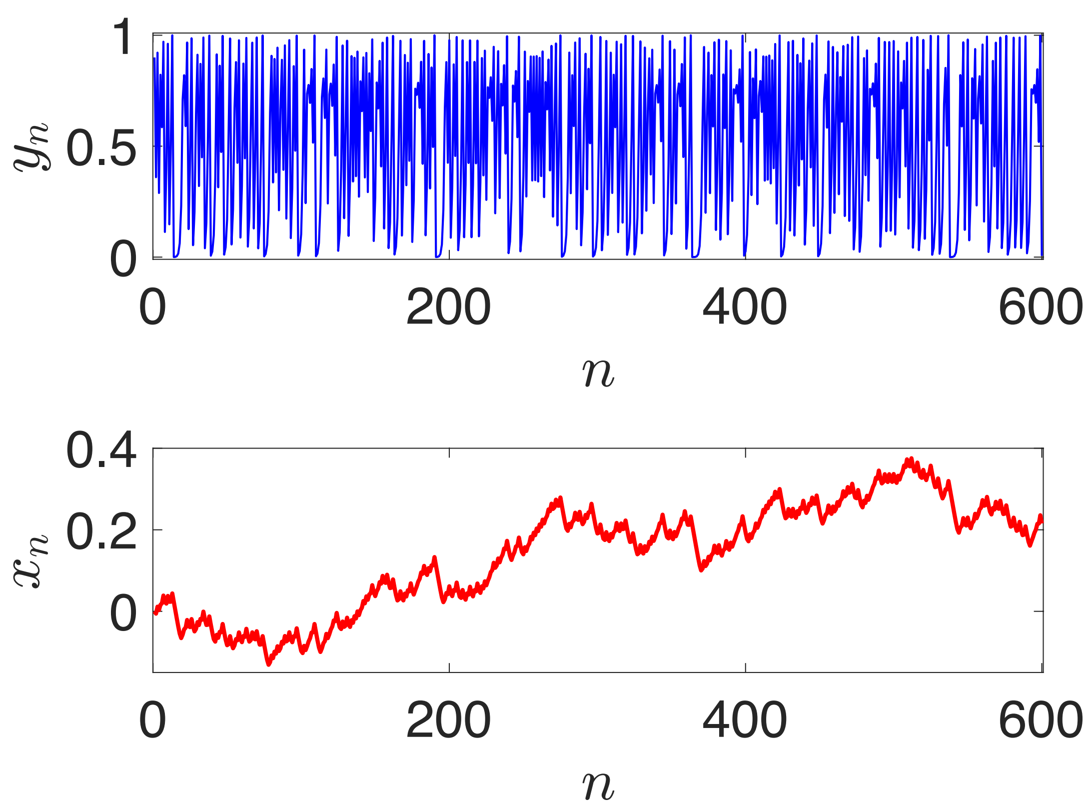


Brownian motion

Homogenisation in action

$$x_{n+1} = x_n + \varepsilon(y_n - \tfrac{1}{2})$$
$$y_{n+1} = 4y_n(1 - y_n)$$

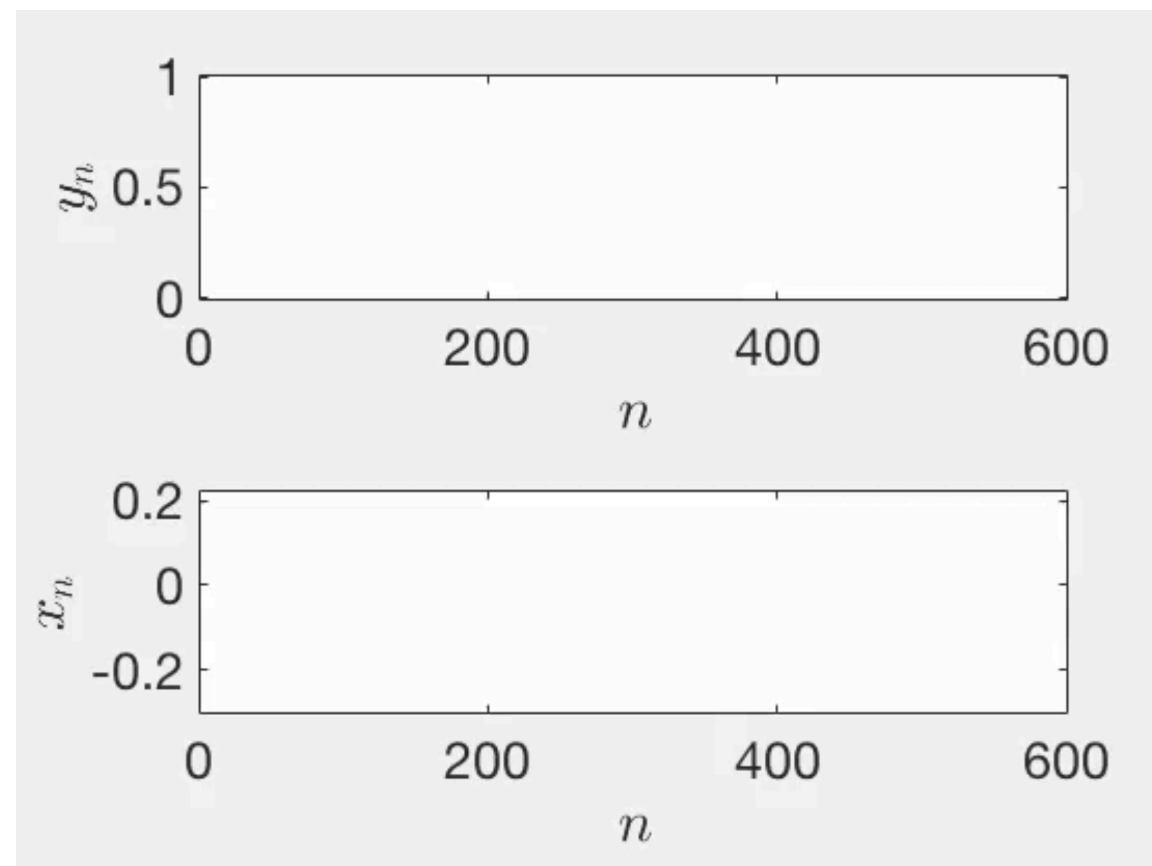
strong chaos



Brownian motion

$$x_{n+1} = x_n + \varepsilon(y^\star - y_n)$$
$$y_{n+1} = \begin{cases} y_n(1 + 2^\gamma y_n^\gamma) & 0 \le y_n \le \frac{1}{2} \\ 2y_n - 1 & \frac{1}{2} \le y_n \le 1 \end{cases}$$

weak chaos



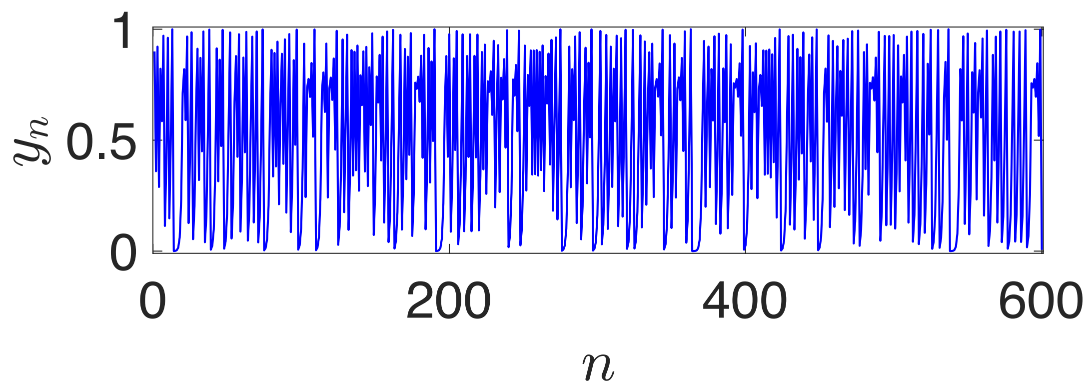
Homogenisation in action

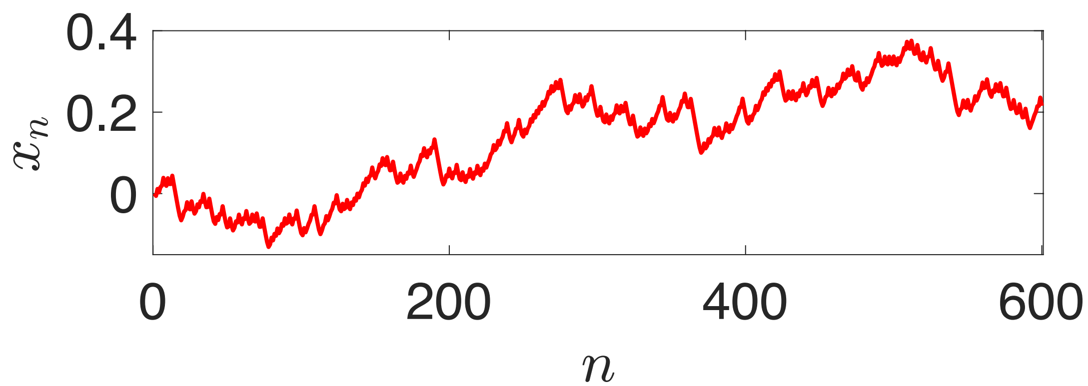
$$x_{n+1} = x_n + \varepsilon(y_n - \tfrac{1}{2})$$
$$y_{n+1} = 4y_n(1 - y_n)$$

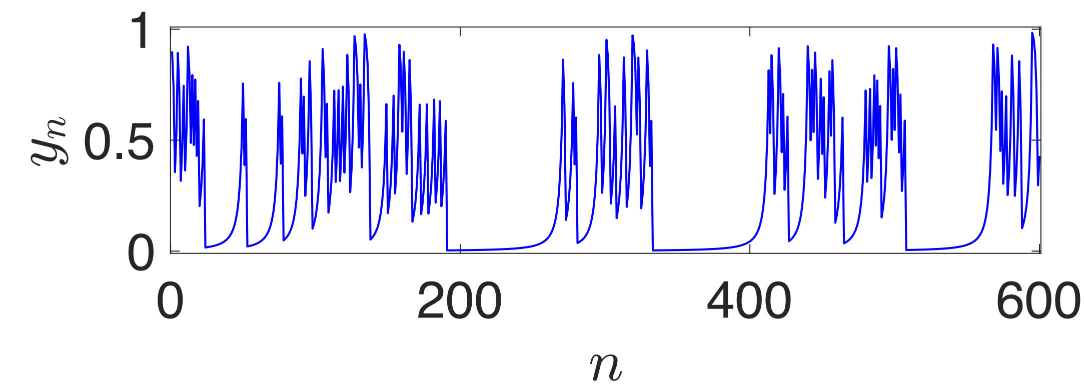
strong chaos

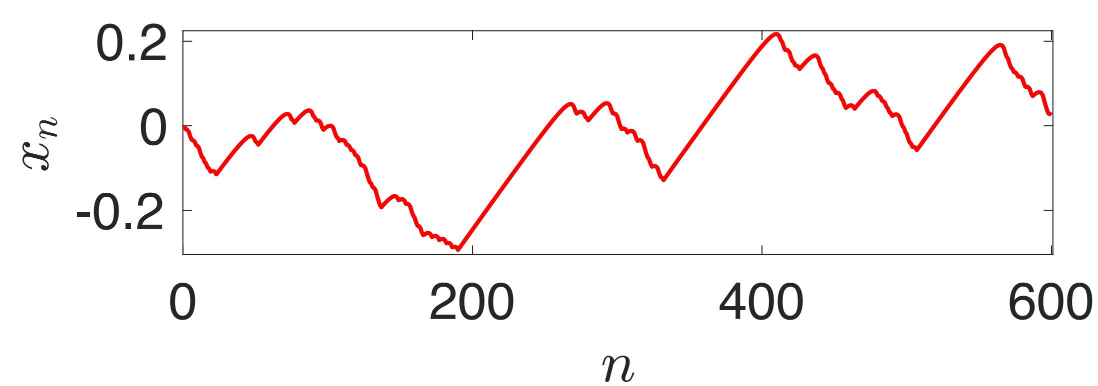
$$x_{n+1} = x_n + \varepsilon(y^\star - y_n)$$
$$y_{n+1} = \begin{cases} y_n(1 + 2^\gamma y_n^\gamma) & 0 \le y_n \le \frac{1}{2} \\ 2y_n - 1 & \frac{1}{2} \le y_n \le 1 \end{cases}$$

weak chaos









Brownian motion

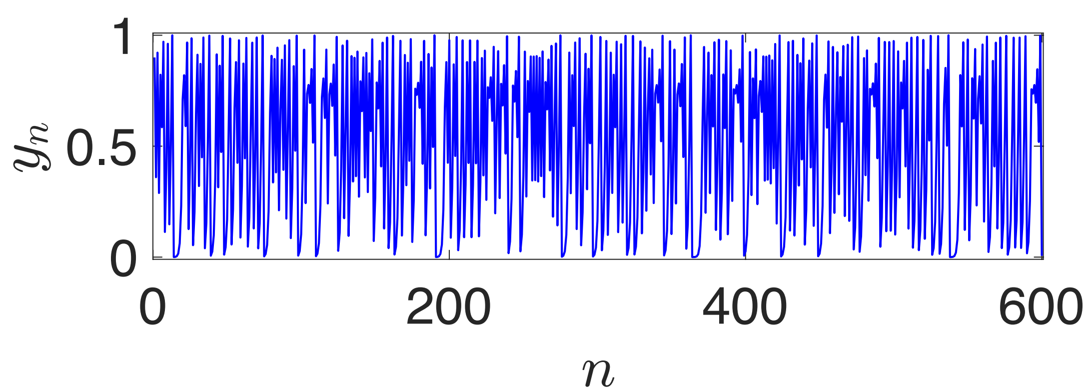
Homogenisation in action

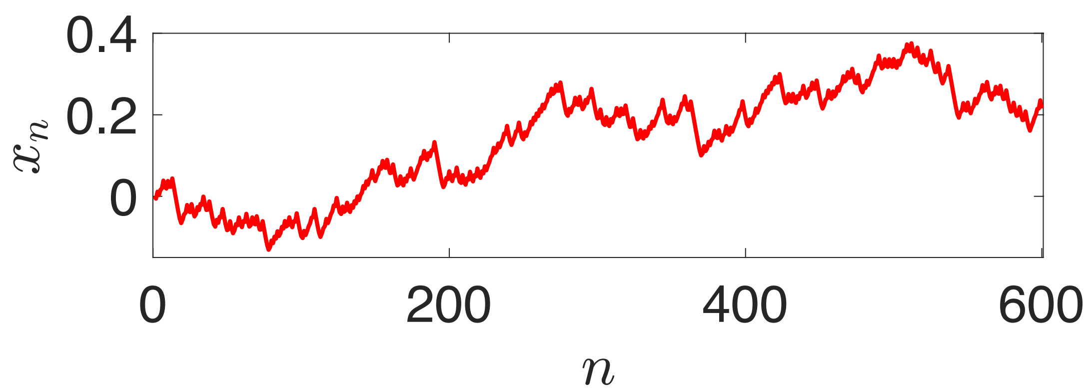
$$x_{n+1} = x_n + \varepsilon(y_n - \tfrac{1}{2})$$
$$y_{n+1} = 4y_n(1 - y_n)$$

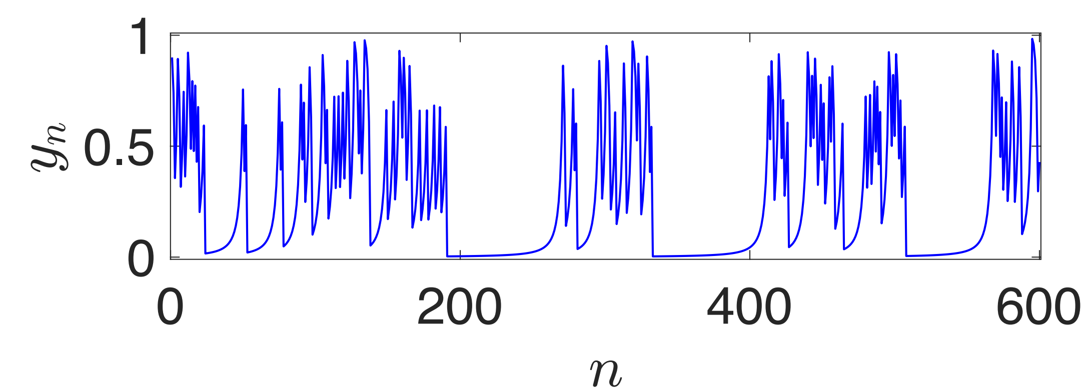
strong chaos

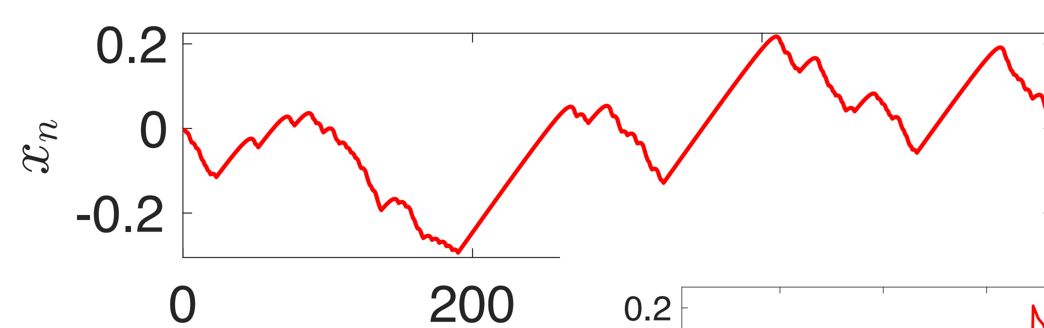
$$x_{n+1} = x_n + \varepsilon(y^\star - y_n)$$
$$y_{n+1} = \begin{cases} y_n(1 + 2^\gamma y_n^\gamma) & 0 \le y_n \le \frac{1}{2} \\ 2y_n - 1 & \frac{1}{2} \le y_n \le 1 \end{cases}$$

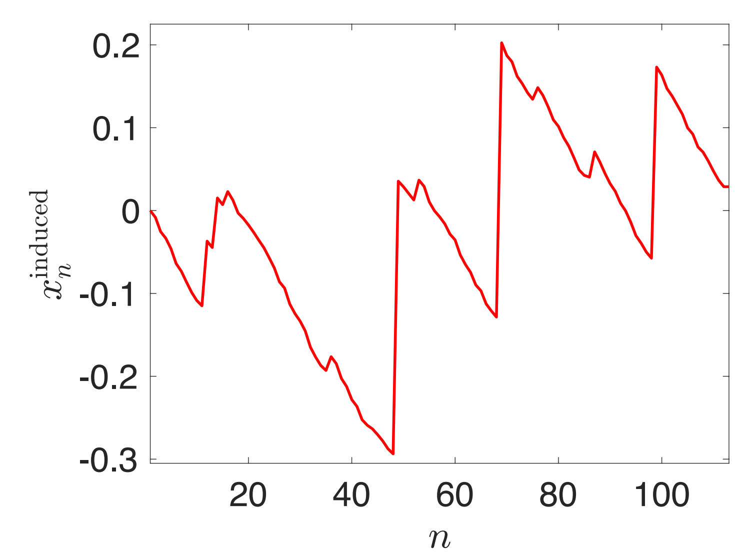
weak chaos











Brownian motion

(GAG, & Melbourne, Proc Roy Soc A (2013))

inducing

α-stable noise

$S(\alpha, \beta, \eta, \mu)$

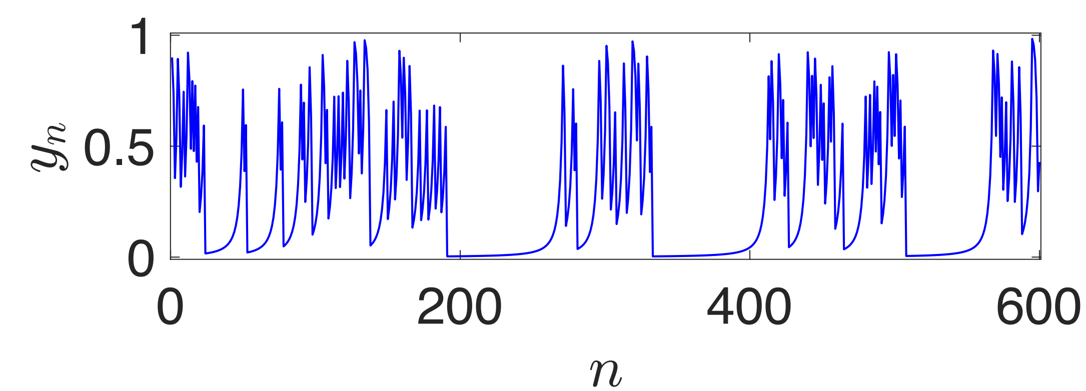
Homogenisation in action

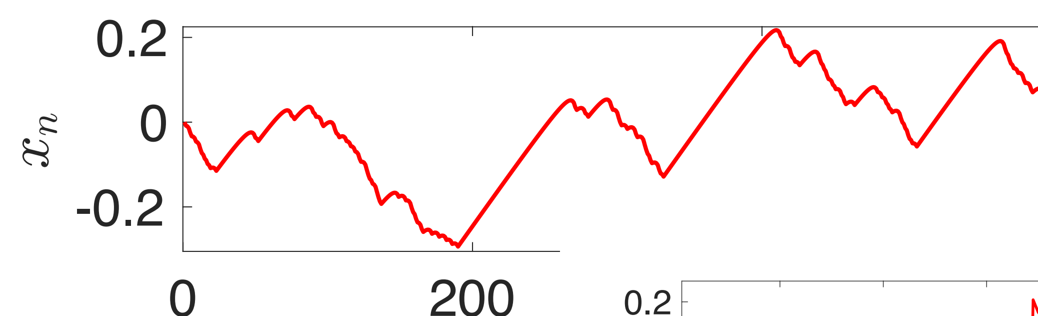
$$x_{n+1} = x_n + \varepsilon(y_n - \tfrac{1}{2})$$
$$y_{n+1} = 4y_n(1 - y_n)$$

strong chaos

$$x_{n+1} = x_n + \varepsilon(y^\star - y_n)$$
$$y_{n+1} = \begin{cases} y_n(1 + 2^\gamma y_n^\gamma) & 0 \le y_n \le \frac{1}{2} \\ 2y_n - 1 & \frac{1}{2} \le y_n \le 1 \end{cases}$$

weak chaos





Can be used to devise a test for anomalous diffusion in time series

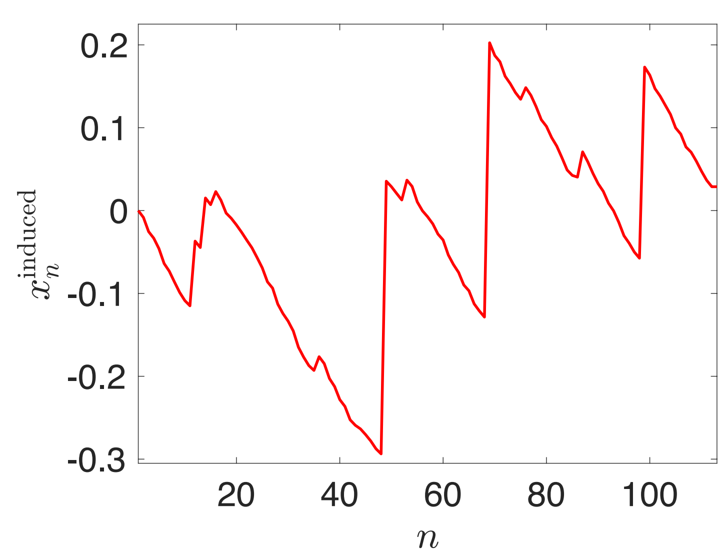
(GAG & Melbourne, J Stat Mech (2016))

(GAG, & Melbourne, Proc Roy Soc A (2013))

inducing

α-stable noise

$S(\alpha, \beta, \eta, \mu)$



Homogenisation

resolved/slow: $dx = \frac{1}{\varepsilon} f_0(x,y)\,dt + f_1(x,y)\,dt$

unresolved/fast: $dy = \frac{1}{\varepsilon^2} g(x,y)\,dt + \frac{1}{\varepsilon}\sigma(x,y)\,dW_t$

Assumptions:

- fast y-process is ergodic with measure μ_x (mild chaoticity assumptions)
- $\int f_0(x,y)d\mu_x = 0$

Then, in the limit of $\varepsilon \to 0$, the statistics of the slow x-dynamics is approximated by

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

where the diffusion matrix is given by a Green-Kubo formula

$$\frac{1}{2}\Sigma\Sigma^T = \int_0^\infty C(s)ds$$

with the auto-correlation matrix $C(t) = \mathbb{E}^{\mu_x}[f_0(x,y)f_0(x,y(t))]$ and

$$F(X) = \int f_1(x,y)\,d\mu_x + \int_0^\infty \int \nabla_x f_0(x,y(s)) \otimes f_0(x,y)d\mu_x \; ds$$

formally:

$$d\mu = \rho(x,y)dx$$

$$\rho(x,y) = \hat{\rho}(x)\rho_\infty(y|x) + \varepsilon\rho_1(x,y) + \ldots$$

Open problems and challenges

- slow dynamics couples back into the fast dynamics

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(x,y)$$

What can go wrong?

If the fast invariant measure μ_x does not depend smoothly on x ("no linear response") even averaging does not "*work*"

$$F(X) = \underbrace{\int f_1(x,y)\mu_x(dy)}_{\text{non-Lipschitz}}$$

non-Lipschitz
uniqueness of solutions not guaranteed

Open problems and challenges

- slow dynamics couples back into the fast dynamics
- finite time scale separation

Theory works in the limit $\varepsilon \to 0$
but in many physical applications ε is not so small

Where do we need the limit?

Averaging: Large deviation principle: $$|\frac{1}{T}\int_0^T f_1(x,y(s))ds - F(x)|$$

Homogenisation: Central Limit Theorem (Weak Invariance Principle)

$$W_\varepsilon(t) = \varepsilon \int_0^{\frac{t}{\varepsilon^2}} f_0(y(s))ds \to_w W(t) \quad \text{as} \quad \varepsilon \to 0$$

Finite ε effects are finite size effects

The Central Limit Theorem and the Edgeworth expansion

The Central Limit Theorem

Assume X_i are *i.i.d.* random variables

$$S_n := \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu) \to_d \mathcal{N}(0,1)$$

where $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathbb{E}[X_i^2]$

For finite n there are *deviations* to the CLT

These are described by the **Edgeworth expansion**

$$\rho_n(x) = \Phi_{0,\sigma^2}(x) \times \left(1 + \frac{1}{6\sqrt{n}} \frac{\gamma}{\sigma^3} H_3(x/\sigma)\right) + o(\frac{1}{\sqrt{n}})$$

where $H_3(x) = x^3 - 3x$ is the third Hermite polynomial

and γ/σ^3 is the skewness of X_i

The Central Limit Theorem and the Edgeworth expansion

The Central Limit Theorem

Assume X_i are *i.i.d.* random variables

$$S_n := \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu) \to_d \mathcal{N}(0,1)$$

where $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathbb{E}[X_i^2]$

For finite n there are *deviations* to the CLT

These are described by the **Edgeworth expansion**

$$\rho_n(x) = \Phi_{0,\sigma^2}(x) \times \left(1 + \frac{1}{6\sqrt{n}} \frac{\gamma}{\sigma^3} H_3(x/\sigma)\right) + o(\frac{1}{\sqrt{n}})$$

where $H_3(x) = x^3 - 3x$ is the third Hermite polynomial

and γ/σ^3 is the skewness of X_i

this is not a density!

The Central Limit Theorem and the Edgeworth expansion

The Central Limit Theorem

Assume X_i are *i.i.d.* random variables

$$S_n := \frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(X_j - \mu) \to_d \mathcal{N}(0,1)$$

where $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathbb{E}[X_i^2]$

For finite n there are *deviations* to the CLT

These are described by the **Edgeworth expansion**

$$\rho_n(x) = \Phi_{0,\sigma^2}(x) \times \left(1 + \frac{1}{6\sqrt{n}}\frac{\gamma}{\sigma^3}H_3(x/\sigma)\right) + o(\frac{1}{\sqrt{n}})$$

where $H_3(x) = x^3 - 3x$ is the third Hermite polynomial

and γ/σ^3 is the skewness of X_i

can be pushed to any order involving higher-order moments

The Central Limit Theorem and the Edgeworth expansion

The Central Limit Theorem

Assume X_i are stationary *weakly dependent* random variables

$$S_n := \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu) \to_d \mathcal{N}(0,1)$$

where $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathbb{E}[X_i^2] + 2\sum_{j=1}^{\infty} \mathbb{E}[X_1 X_{j+1}]$

For finite n there are *deviations* to the CLT

These are described by the **Edgeworth expansion**

$$\rho_n(x) = \Phi_{0,\sigma^2+\delta\sigma^2/n}(x) \times \left(1 + \frac{1}{\sqrt{n}} \delta\kappa H_3(x/\sigma)\right) + o(\frac{1}{\sqrt{n}})$$

where H_3 is the third Hermite polynomial and $\delta\sigma^2$ and $\delta\kappa$ are integrals of correlation functions of X_i *(Götze & Hipp (1983))*

Stochastic Parametrisation using the Edgeworth expansion

Given a multi-scale dynamical system

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g(y)$$

(I) determine the Edgeworth expansion coefficients $\sigma^2_{\rm GK},\ \delta\kappa$ associated with $f_0(x,y)$

Stochastic Parametrisation using the Edgeworth expansion

Given a multi-scale dynamical system

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

(I) determine the Edgeworth expansion coefficients $\sigma^2_{\mathrm{GK}},\ \delta\kappa$ associated with $f_0(x,y)$

(II) model the multi-scale system by the surrogate stochastic process

$$\dot{X} = \frac{1}{\varepsilon} A(\eta) + F(X)$$
$$d\eta = -\frac{1}{\varepsilon^2}\gamma\eta\, dt + \frac{1}{\sqrt{\varepsilon}} dW_t$$

with $A(\eta) = a\eta^2 + b\eta + c$

1d Ornstein-Uhlenbeck process

where the parameters $a,\ b,\ c,\ \gamma$ are determined such that the Edgeworth expansion coefficients associated with $A(\eta)$ match $\sigma^2_{\mathrm{GK}},\ \delta\kappa$

Stochastic Parametrisation using the Edgeworth expansion

Given a multi-scale dynamical system

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

(I) determine the Edgeworth expansion coefficients $\sigma^2_{\mathrm{GK}},\ \delta\kappa$ associated with $f_0(x,y)$

(II) model the multi-scale system by the surrogate stochastic process

$$\dot{X} = \frac{1}{\varepsilon} A(\eta) + F(X)$$
$$d\eta = -\frac{1}{\varepsilon^2}\gamma\eta\, dt + \frac{1}{\sqrt{\varepsilon}} dW_t$$

with $A(\eta) = a\eta^2 + b\eta + c$

1d Ornstein-Uhlenbeck process

where the parameters $a,\ b,\ c,\ \gamma$ are determined such that the Edgeworth expansion coefficients associated with $A(\eta)$ match $\sigma^2_{\mathrm{GK}},\ \delta\kappa$

Remark: By construction the homogenised limit system of the original and the surrogate system are the same!

How to calculate the Edgeworth coefficients?

The three time scales of multi-scale systems

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

How to calculate the Edgeworth coefficients?

The three time scales of multi-scale systems

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

nontrivial fast dynamics

trivial slow dynamics $x(t) = x_0$

How to calculate the Edgeworth coefficients?

The three time scales of multi-scale systems

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

nontrivial fast dynamics

trivial slow dynamics $x(t) = x_0$

fast dynamics has equilibrated

trivial slow dynamics $x(t) = x_0$

How to calculate the Edgeworth coefficients?

The three time scales of multi-scale systems

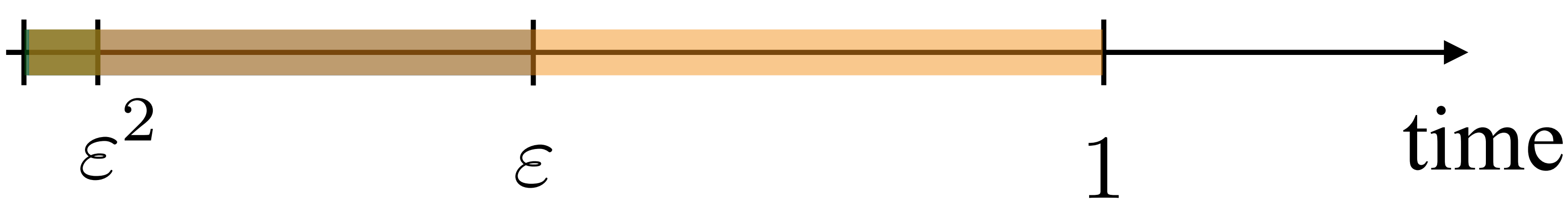
$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

nontrivial fast dynamics

trivial slow dynamics $x(t) = x_0$

diffusive time scale: CLT

$$dX = F(X)\,dt + \sigma(X) \circ dW_t$$



fast dynamics has equilibrated

trivial slow dynamics $x(t) = x_0$

How to calculate the Edgeworth coefficients?

The three time scales of multi-scale systems

$$\dot{x} = \frac{1}{\varepsilon} f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x,y)$$

nontrivial fast dynamics

trivial slow dynamics $x(t) = x_0$

diffusive time scale: CLT

$$dX = F(X)\,dt + \sigma(X) \circ dW_t$$

fast dynamics has equilibrated

trivial slow dynamics $x(t) = x_0$

expect deviations of CLT on timescale $t = \varepsilon$

$$\frac{x(t) - x_0}{\sqrt{t}} \to \sigma(x_0) W_t$$

How to calculate the Edgeworth coefficients?

Consider $\rho_t(x(t)|x(0) = x_0) = \int dxdy\, e^{\mathcal{L}t}\delta_{x_0}(x)\mu(dy)$ for $t = \varepsilon$

transfer operator

$$\dot{x} = \frac{1}{\varepsilon} f_0(x, y) + f_1(x, y)$$

$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y) + \frac{1}{\varepsilon} g_1(x, y)$$

$$\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2$$

$$\mathcal{L}_0\rho = -\partial_y (g_0\rho)\,,\ \mathcal{L}_1\rho = -\,\partial_x (f_0\rho) - \partial_y (g_1\rho)\,,\ \mathcal{L}_2\rho = -\partial_x (f_1\rho)$$

How to calculate the Edgeworth coefficients?

Consider $\rho_t(x(t)|x(0)=x_0) = \int dxdy\, e^{\mathcal{L}t}\delta_{x_0}(x)\mu(dy)$ for $t=\varepsilon$

transfer operator

$$\dot{x} = \frac{1}{\varepsilon}f_0(x,y) + f_1(x,y)$$
$$\dot{y} = \frac{1}{\varepsilon^2}g_0(y) + \frac{1}{\varepsilon}g_1(x,y)$$

$$\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2$$
$$\mathcal{L}_0\rho = -\partial_y\left(g_0\rho\right),\ \mathcal{L}_1\rho = -\,\partial_x\left(f_0\rho\right) - \partial_y\left(g_1\rho\right),\ \mathcal{L}_2\rho = -\partial_x\left(f_1\rho\right)$$

Calculate asymptotically, using successive applications of the Duhamel-Dyson formula, up to $\mathcal{O}(\varepsilon^n)$:

$$\frac{\mathbb{E}[x(\varepsilon)-x_0]}{\sqrt{\varepsilon}} = \sqrt{\epsilon}\,\xi = \sqrt{\epsilon}\langle f_1(x_0)\rangle$$

$$\frac{\mathbb{E}[\hat{x}^2]}{\varepsilon} = \sigma^2_{\mathrm{GK}} - 2\varepsilon\int_0^{\frac{t}{\varepsilon^2}} ds\left(s\langle f_0 e^{\mathcal{L}_0 s} f_0\rangle - \langle f_0 e^{\mathcal{L}_0 s} f_1\rangle\right) + \cdots$$

$\hat{x} = x - \mathbb{E}[x]$

$$\frac{\mathbb{E}[\hat{x}^3]}{\varepsilon^{\frac{3}{2}}} = \sqrt{\varepsilon}\int_0^{\frac{t}{\varepsilon^2}} ds_1\, ds_2\, \langle f_0 e^{\mathcal{L}_0 s_1} f_0 e^{\mathcal{L}_0 s_2} f_0\rangle$$

Theorem (Wouters & GAG, 2019)

The Edgeworth expansion of the transition probability $\pi_\varepsilon(\xi, t = \varepsilon, x_0)$ *for the deterministic multi-scale system up to* $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ *is given in the limit* $t = \varepsilon \ll 1$ *and* $t/\varepsilon^2 \to \infty$ *by*

$$\pi_\varepsilon(\xi, t = \varepsilon, x_0) = \mathbf{n}_{0,\sigma^2}(\xi) \left(1 + \sqrt{\varepsilon} \left(\frac{c_{\frac{1}{2}}^{(1)}}{\sigma} H_1\left(\frac{\xi}{\sigma}\right) + \frac{c_{\frac{1}{2}}^{(3)}}{3!\sigma^3} H_3\left(\frac{\xi}{\sigma}\right) \right) \right.$$
$$\left. + \varepsilon \left(\frac{c_1^{(2)} + {c_{\frac{1}{2}}^{(1)}}^2}{2\sigma^2} H_2\left(\frac{\xi}{\sigma}\right) + \frac{c_1^{(4)} + 4 c_{\frac{1}{2}}^{(1)} c_{\frac{1}{2}}^{(3)}}{4!\sigma^4} H_4\left(\frac{\xi}{\sigma}\right) + \frac{{c_{\frac{1}{2}}^{(3)}}^2}{2(3!\sigma^3)^2} H_6\left(\frac{\xi}{\sigma}\right) \right) \right)$$
$$+ \mathcal{O}(\varepsilon^{\frac{3}{2}}) .$$

It involves only the cumulants $c_\varepsilon^{(p)}$ *with* $p \leq 4$ *with explicit expressions. These cumulants only involve the leading order measure* $\mu_{x_0}^{(0)}$ *and, in particular, do not involve the linear response term* $\mu_{x_0}^{(1)}$.

(Wouters & GAG, Proc Roy Soc A (2019))

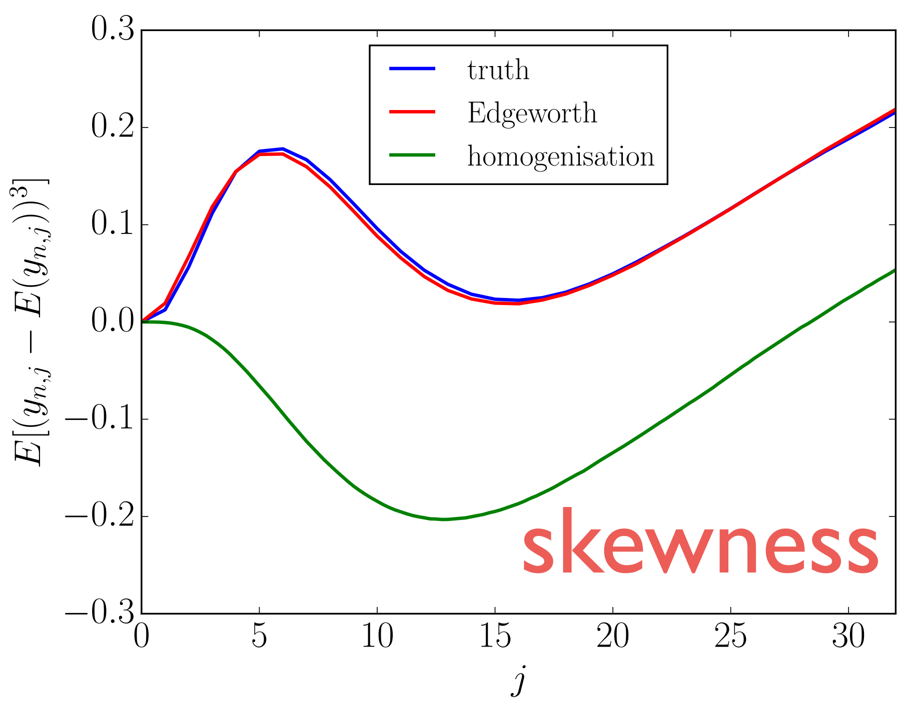
$$
\begin{aligned}
x_{j+1}^{(\varepsilon)} &= x_j^{(\varepsilon)} + \varepsilon f_0(y_j) + \varepsilon^2 f_1(x_j^{(\varepsilon)}) \\
y_{j+1} &= p\, y_j \pmod 1
\end{aligned}
$$

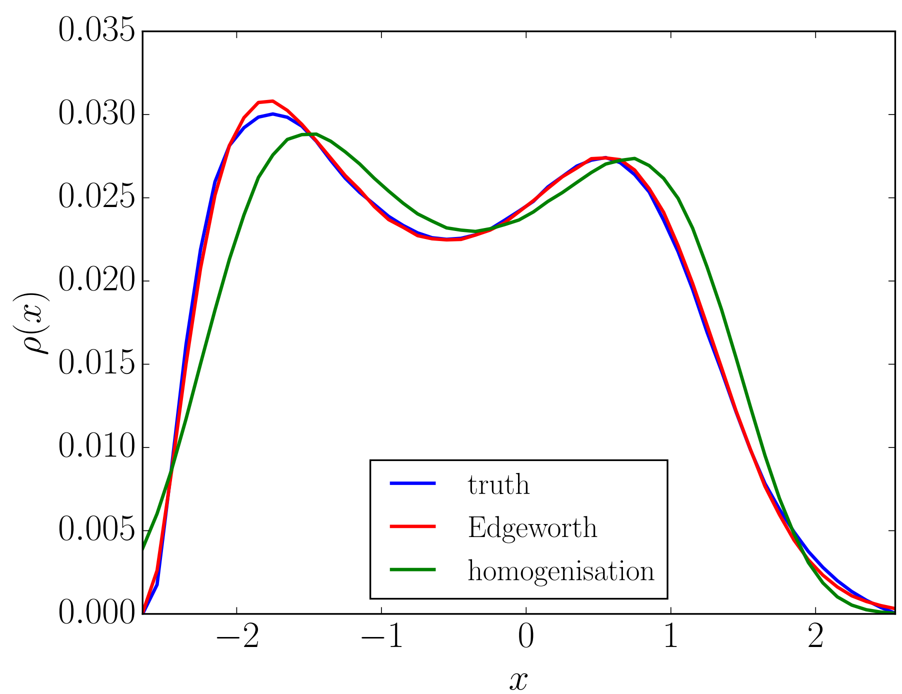
Triad

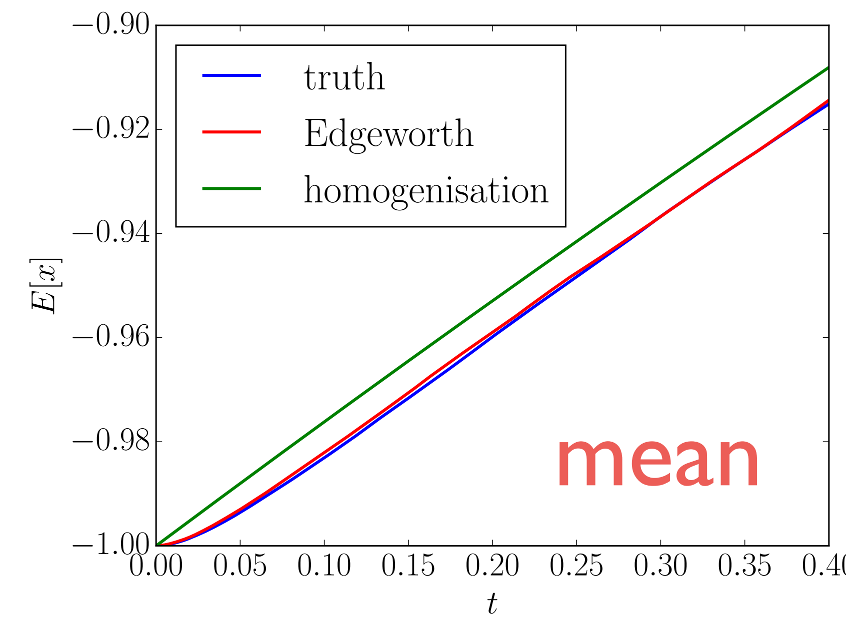
(Majda et al)

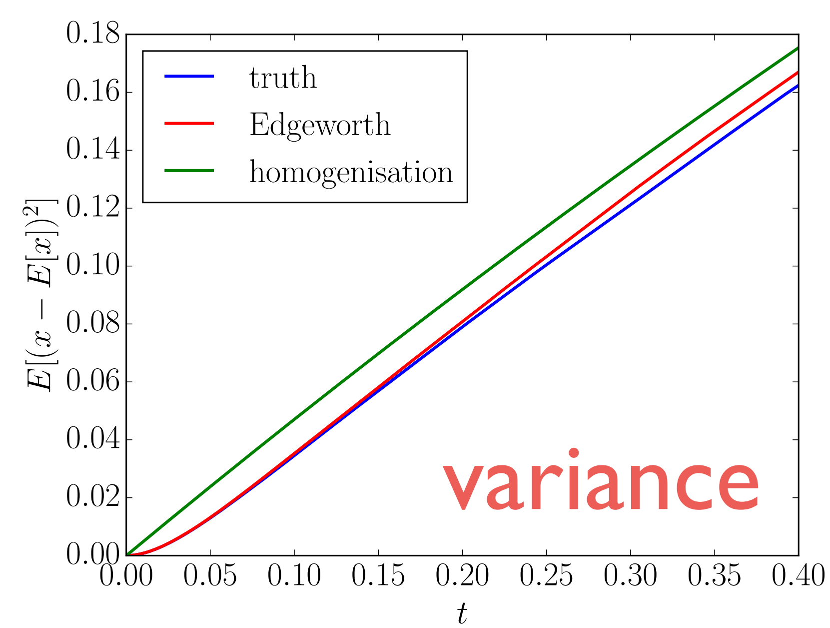
$$
\begin{aligned}
\dot{x} &= \frac{1}{\varepsilon} B_0 y_1 y_2 \\
\dot{y}_1 &= \frac{1}{\varepsilon} B_1 x y_2 - \frac{1}{\varepsilon^2}\gamma_1 y_1 - \frac{1}{\varepsilon}\sigma_1 \dot{W}_1 \\
\dot{y}_2 &= \frac{1}{\varepsilon} B_2 x y_1 - \frac{1}{\varepsilon^2}\gamma_2 y_2 - \frac{1}{\varepsilon}\sigma_2 \dot{W}_2
\end{aligned}
$$

backcoupling









Summary

We have used the Edgeworth expansion to push stochastic model reduction past the limit of infinite time scale separation, going beyond the Central Limit Theorem

We have developed a machinery to calculate the Edgeworth corrections for continuous time deterministic systems

The fast dynamics are replaced by a stochastic surrogate process, the parameters of which are tuned to match the Edgeworth expansion corrections of the full multi-scale system

(Wouters & GAG, Proc Roy Soc A (2019) + SIAM MMS (2019))

Summary

We have used the Edgeworth expansion to push stochastic model reduction past the limit of infinite time scale separation, going beyond the Central Limit Theorem

We have developed a machinery to calculate the Edgeworth corrections for continuous time deterministic systems

The fast dynamics are replaced by a stochastic surrogate process, the parameters of which are tuned to match the Edgeworth expansion corrections of the full multi-scale system

Outlook:

- Use the strategy for the triad system to apply Edgeworth expansion to the barotropic vorticity equation
- Use Edgeworth expansions in a data-driven approach
- Prove the corrections rigorously (start with stochastic fast dynamics)

Applications of Statistical Limit Theorems

- Data assimilation - Ensemble Kalman Filters
- Ensemble forecasting - Stochastically perturbed bred vectors
- Linear response theory
- Numerical integration of deterministic multi-scale systems
- Parametrisation of tropical convection

I - Data Assimilation: Ensemble Kalman Filters

using the reduced stochastic model as forecast model leads to reliable ensembles via dynamics-informed inflation

Homogenisation

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

$$\frac{dy_1}{dt} = \frac{10}{\varepsilon^2}(y_2 - y_1)$$

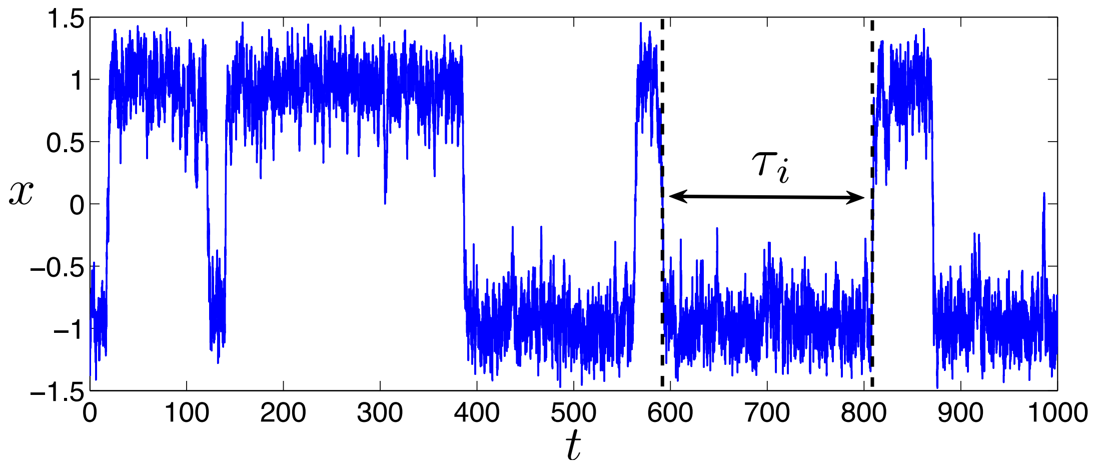
$$\frac{dy_2}{dt} = \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1 y_3)$$

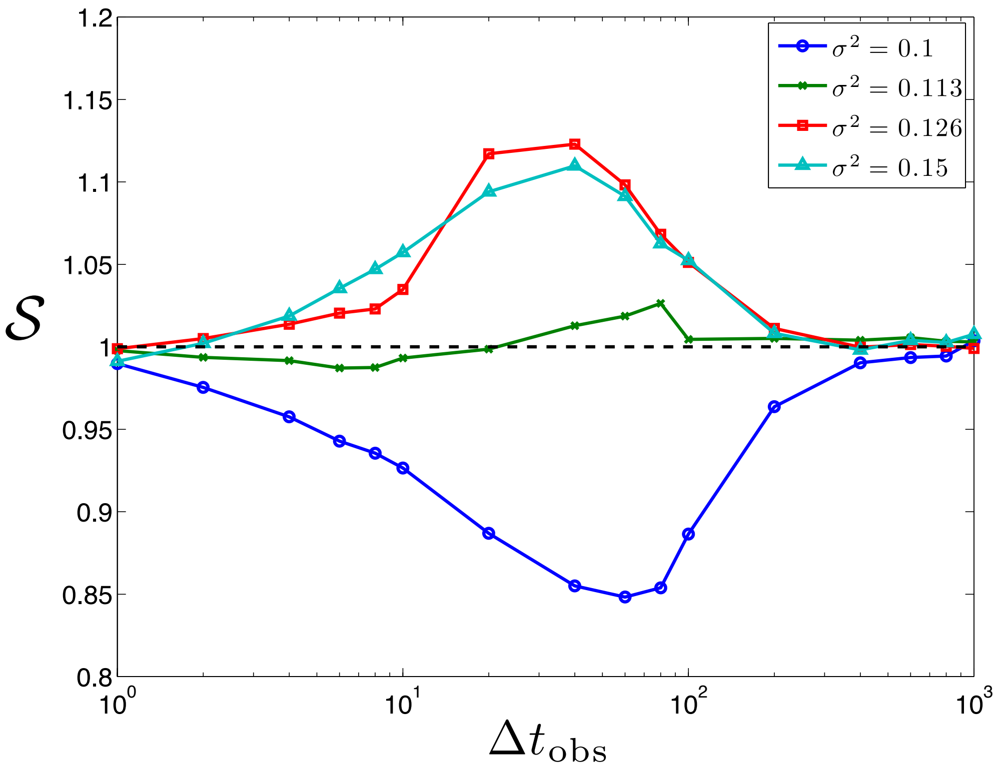
$$\frac{dy_3}{dt} = \frac{1}{\varepsilon^2}\left(y_1 y_2 - \frac{8}{3} y_3\right)$$

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

$$\sigma^2 = 2\left(\frac{4}{90}\right)^2 \int_0^\infty \mathbb{E}[y_2(0) y_2(t)]dt$$

$$\approx 0.113$$





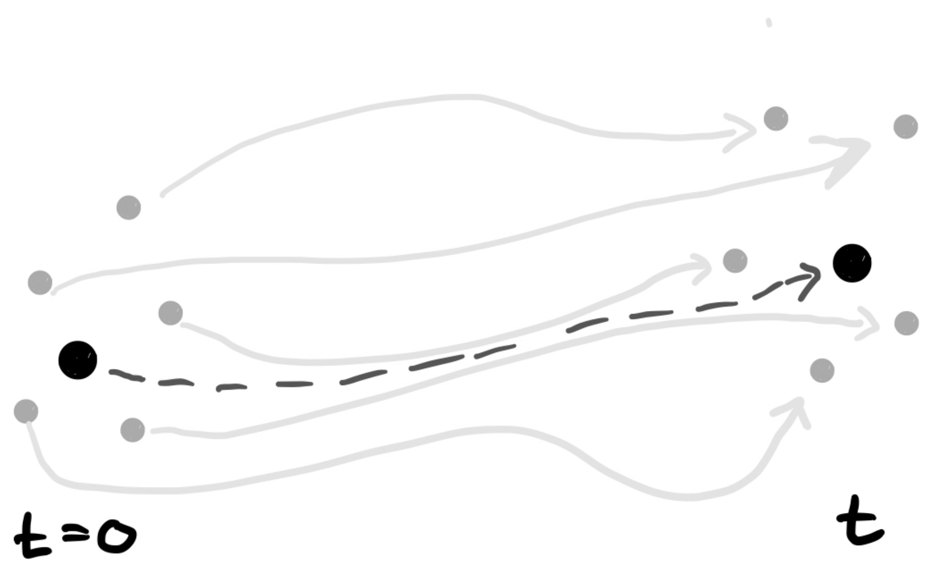
(Mitchell and GAG, JAS (2012); GAG & Harlim, Proc Roy Soc A (2014))

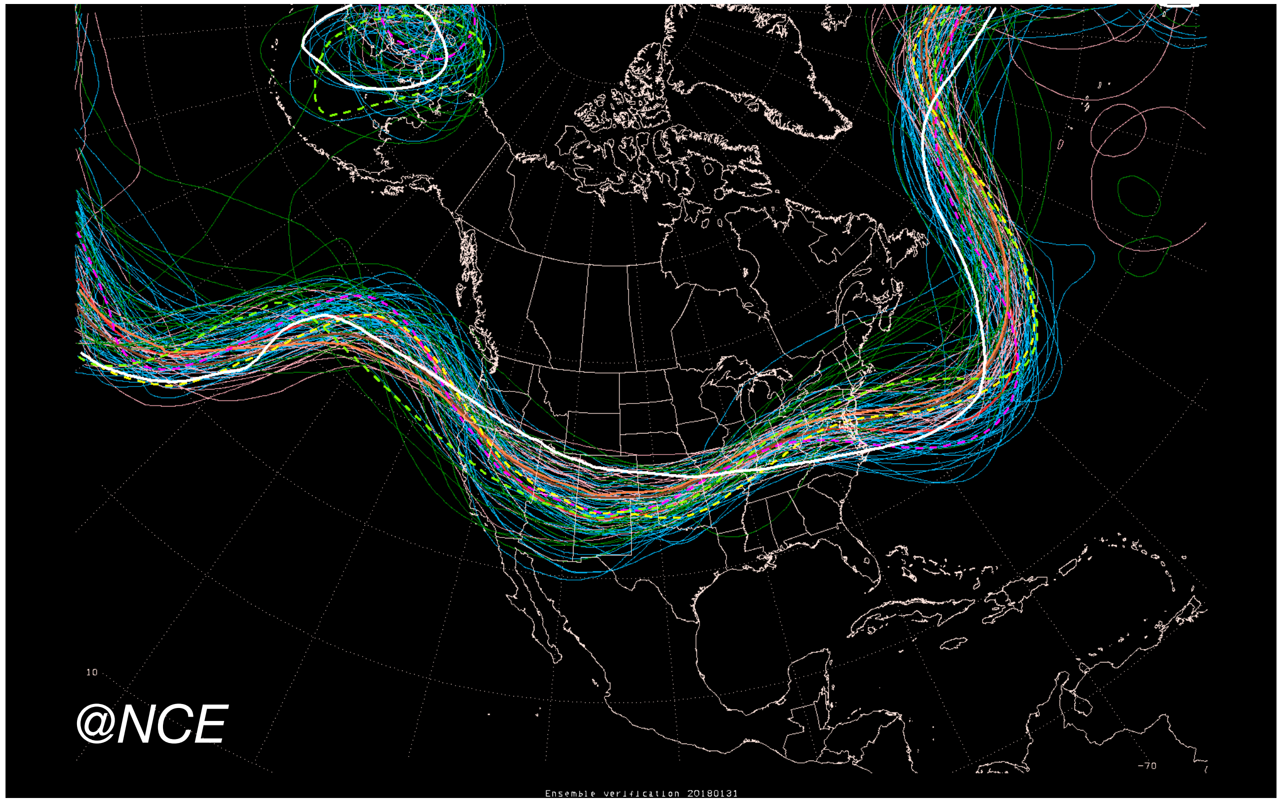
II - Ensemble forecasting

In chaotic systems a single forecast is not meaningful

Probabilistic forecast

Ensemble forecast



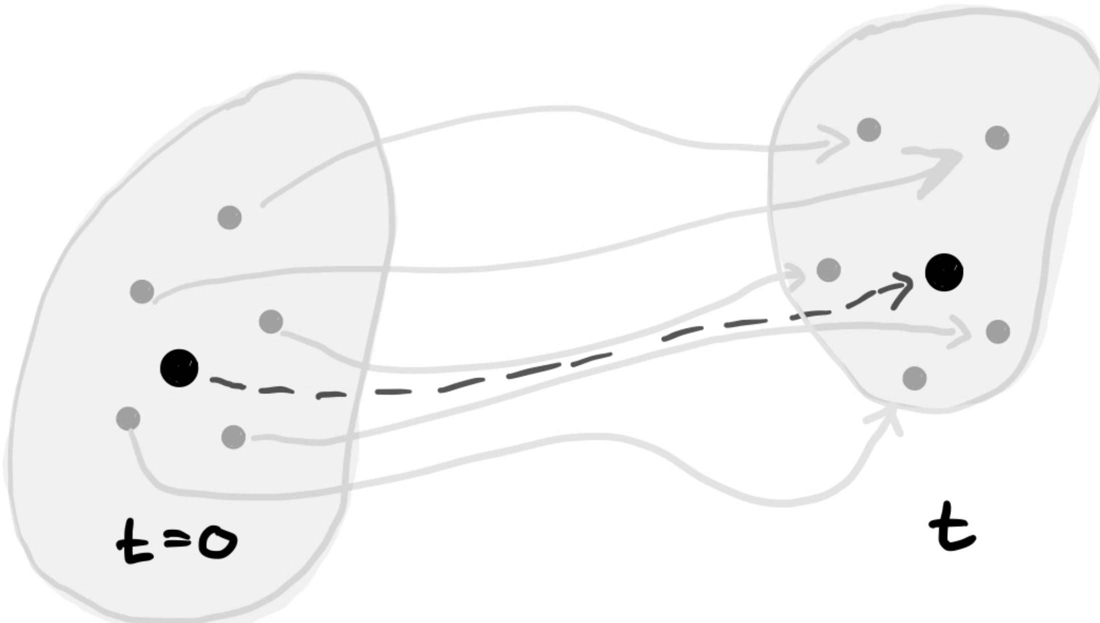


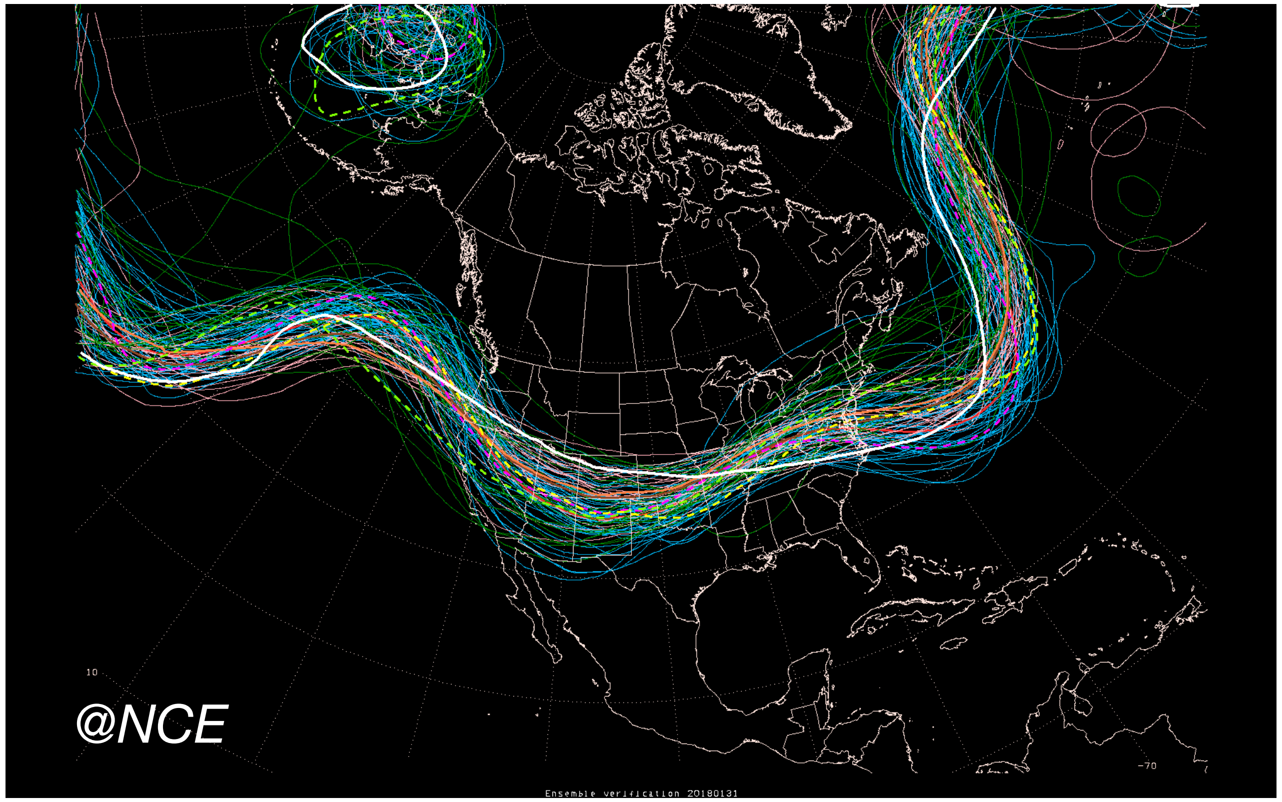
II - Ensemble forecasting

In chaotic systems a single forecast is not meaningful

Probabilistic forecast

Ensemble forecast





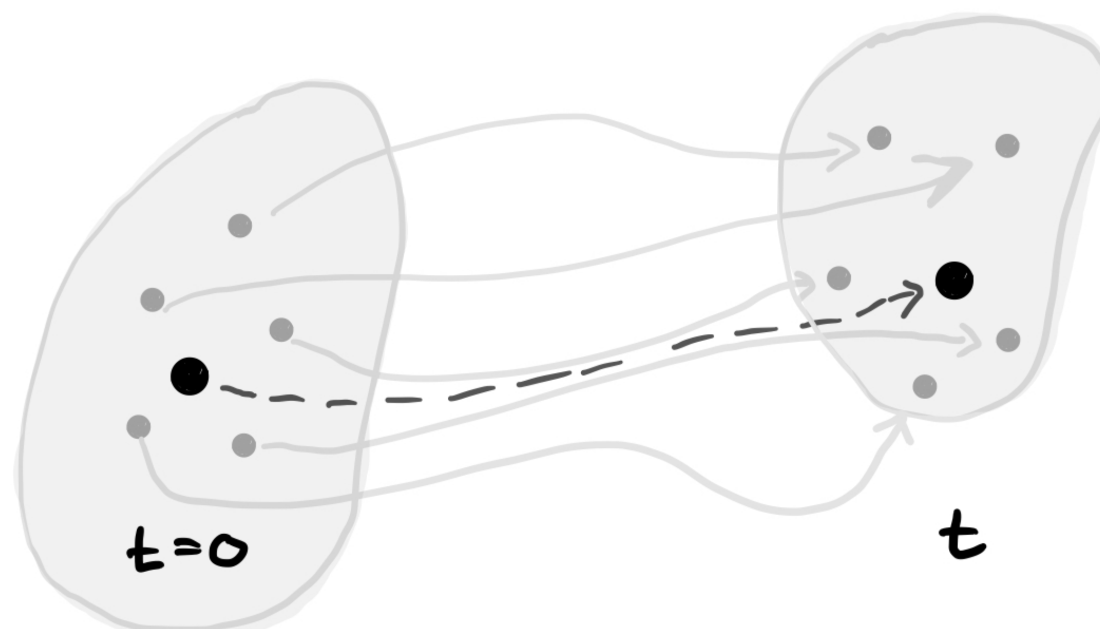
$$p(x,0) \approx \frac{1}{N}\sum_{j=1}^{N} \delta_{x_j} \longmapsto p(x,t) \approx \frac{1}{N}\sum_{j=1}^{N} \delta_{x_j(t)}$$

Use ensemble mean and spread to estimate forecast and its uncertainty

II - Ensemble forecasting

A good ensemble should have (at least) these 4 properties (Pazó *et al 2010*) :

- **Evolve into areas in phase space with large measure**
- **Forecast skill**
- **Reliability**
- **Dynamic adaptation**

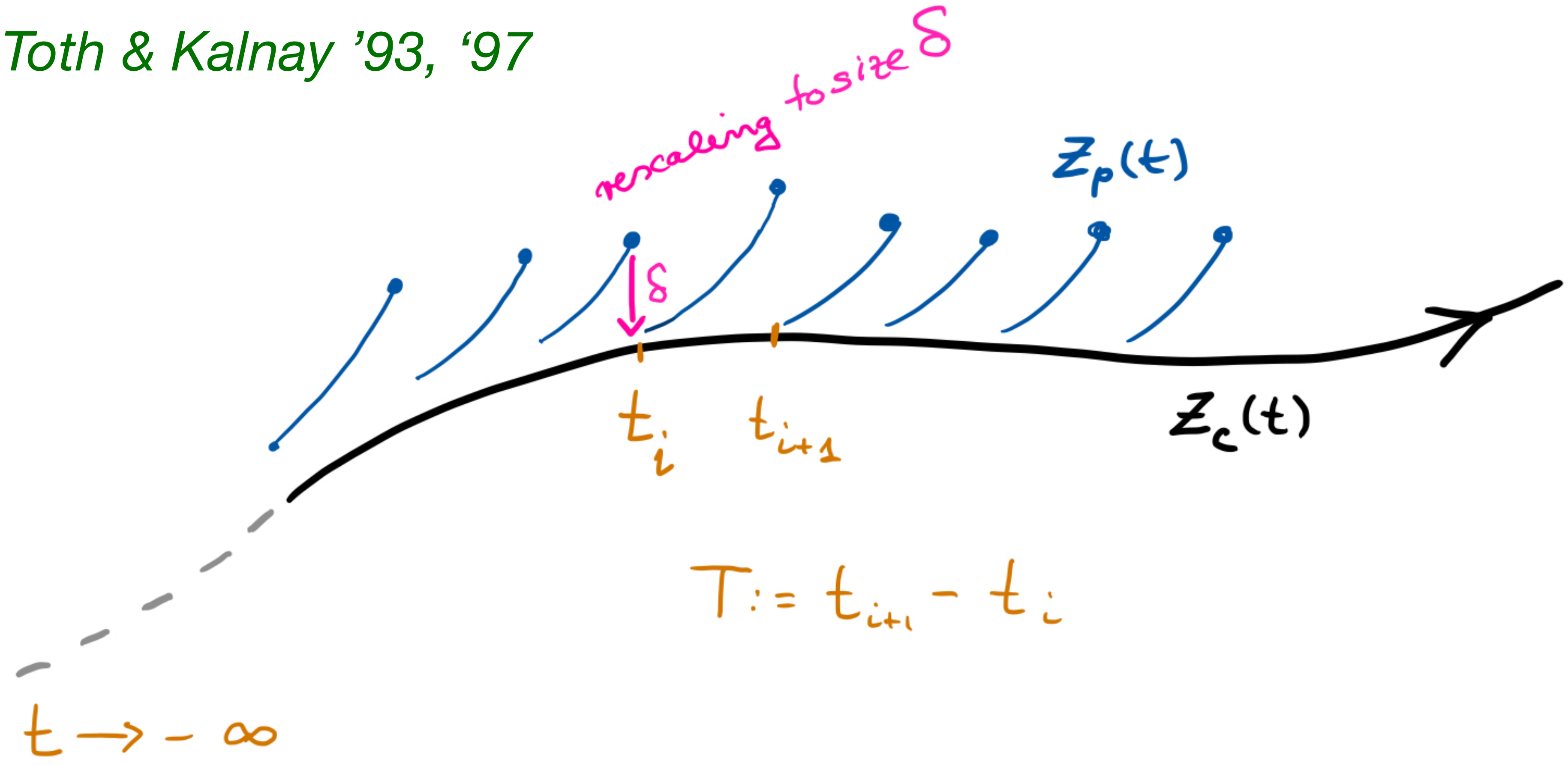


$$p(x,0) \approx \frac{1}{N}\sum_{j=1}^{N}\delta_{X_j} \longmapsto p(x,t) \approx \frac{1}{N}\sum_{j=1}^{N}\delta_{X_j(t)}$$

Bred vectors (BV)

Toth & Kalnay '93, '97

$$z_p(t_i) = z_c(t_i) + \delta \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}$$

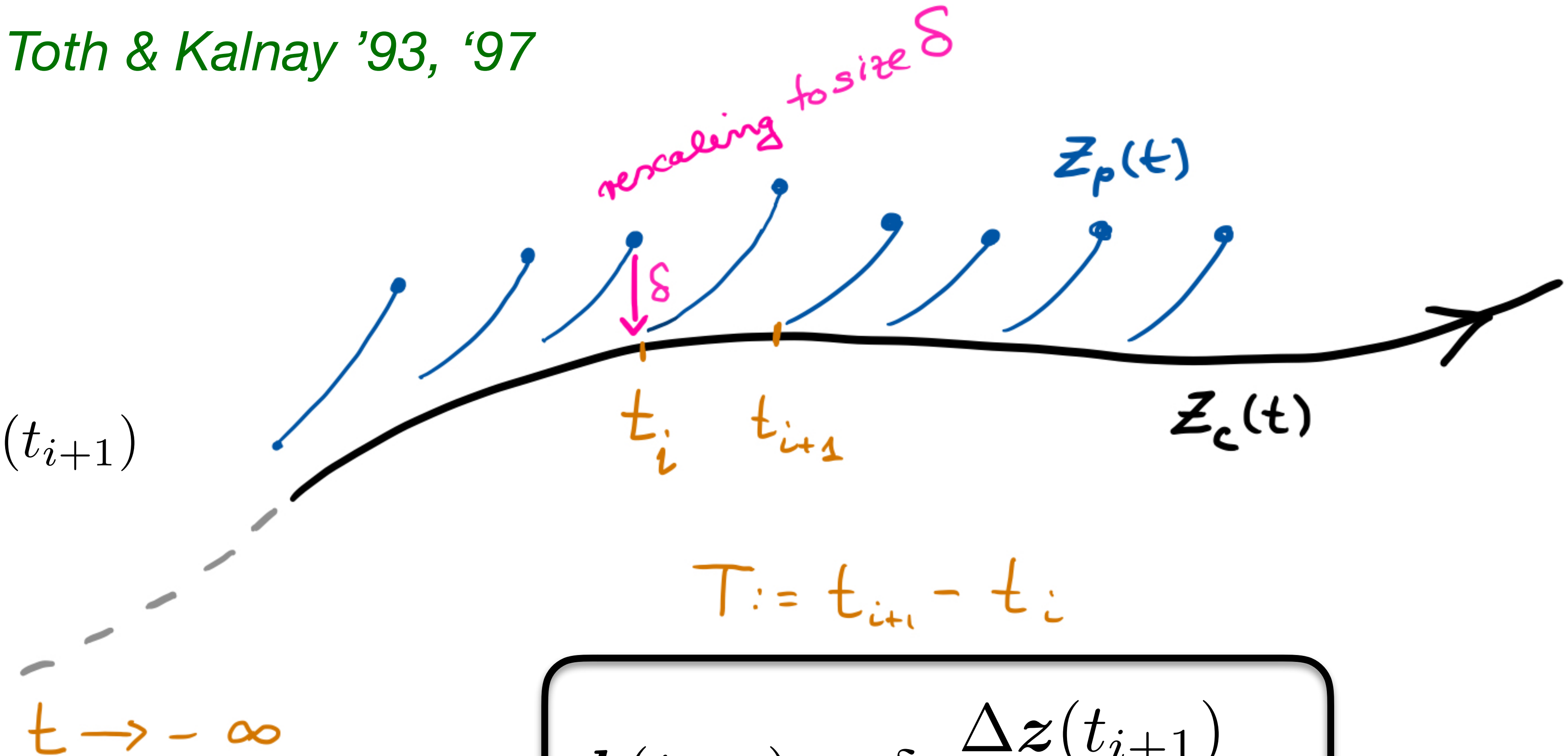


Bred vectors (BV)

Toth & Kalnay '93, '97

$$\boldsymbol{z}_p(t_i) = \boldsymbol{z}_c(t_i) + \delta \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}$$

$$\Delta \boldsymbol{z}(t_{i+1}) = \boldsymbol{z}_p(t_{i+1}) - \boldsymbol{z}_c(t_{i+1})$$



BV

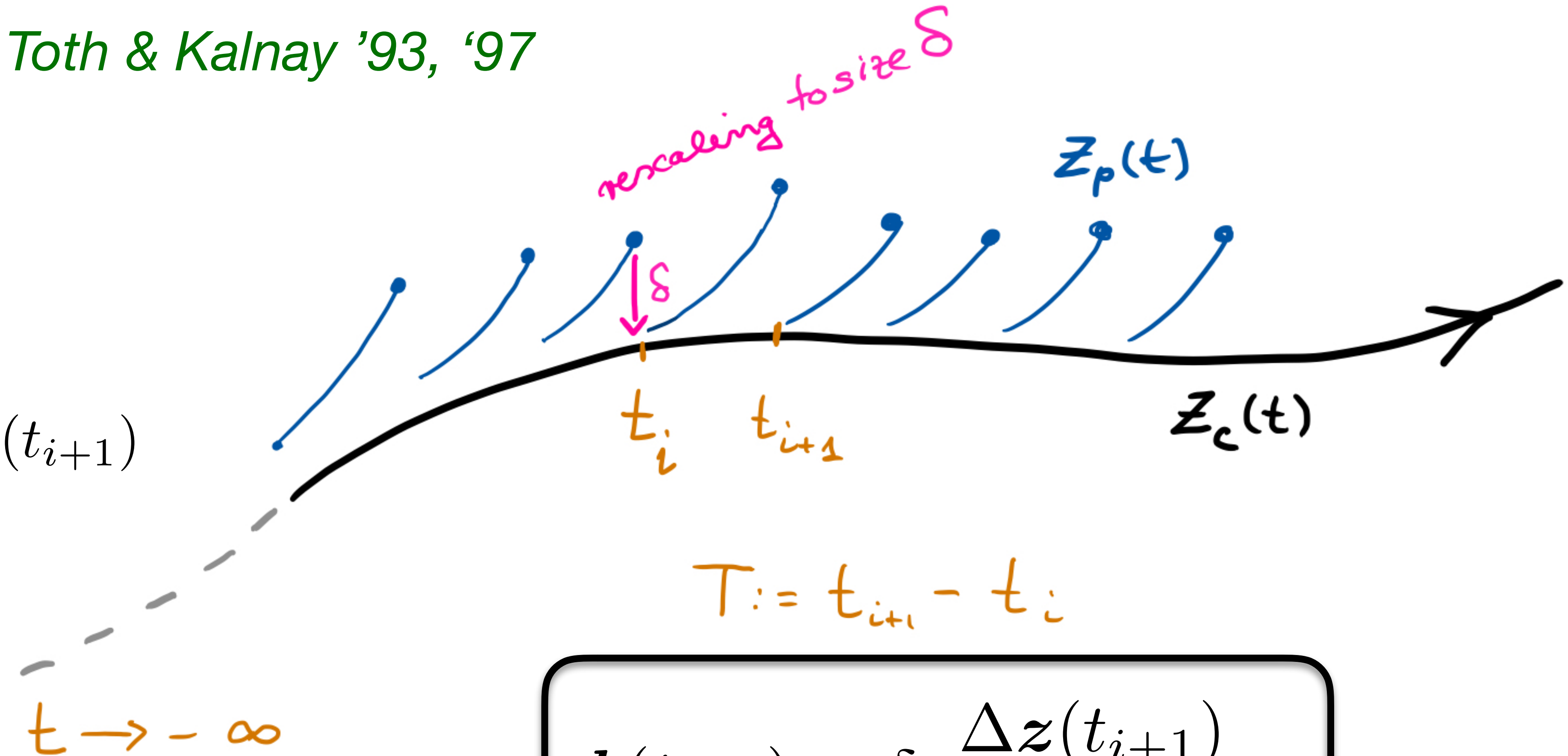
$$\boldsymbol{b}(t_{i+1}) = \delta \frac{\Delta \boldsymbol{z}(t_{i+1})}{\|\Delta \boldsymbol{z}(t_{i+1})\|}$$

Bred vectors (BV)

Toth & Kalnay '93, '97

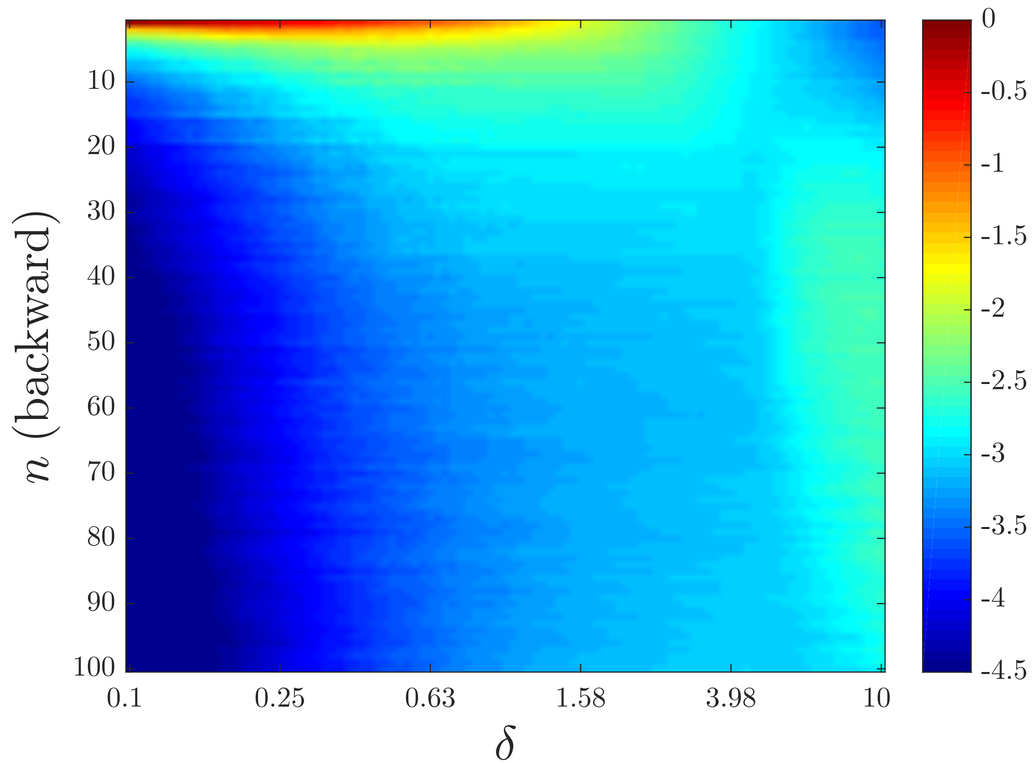
$$z_p(t_i) = z_c(t_i) + \delta \frac{b}{\|b\|}$$

$$\Delta z(t_{i+1}) = z_p(t_{i+1}) - z_c(t_{i+1})$$



BV

$$b(t_{i+1}) = \delta \frac{\Delta z(t_{i+1})}{\|\Delta z(t_{i+1})\|}$$



Advantages

- computationally cheap
- dynamically consistent

Disadvantages

- collapse to a low-dimensional subspace alignment with leading Lyapunov vector *bad spread*

Covariant Lyapunov Vectors l_n

$$\pi_n^i(t_i) = \left| \frac{b^i(t_j)}{\|b^i(t_j)\|} \cdot \frac{l_n(t_j)}{\|l_n(t_j)\|} \right|$$

Way out: Stochastically Perturbed Bred Vectors (SPBVs)

In multi-scale systems with time-scale separation $1/\varepsilon$ we can approxmiate

$$\rho(X, Y, t) = \hat{\rho}(X, t)\rho_\infty(Y|X) + \mathcal{O}(\varepsilon)$$

X: slow
Y: fast

Way out: Stochastically Perturbed Bred Vectors (SPBVs)

In multi-scale systems with time-scale separation $1/\varepsilon$ we can approxmiate

$$\rho(X,Y,t) = \hat{\rho}(X,t)\rho_\infty(Y|X) + \mathcal{O}(\varepsilon)$$

X: slow
Y: fast

$$\mathrm{BV} = \left(\begin{array}{c} \mathrm{slow}_1 \approx 0 \\ \vdots \\ \mathrm{slow}_K \approx 0 \\ \hline \mathrm{fast}_1 \\ \vdots \\ \mathrm{fast}_D \end{array}\right)$$

Way out: Stochastically Perturbed Bred Vectors (SPBVs)

In multi-scale systems with time-scale separation $1/\varepsilon$ we can approxmiate

$$\rho(X,Y,t) = \hat{\rho}(X,t)\rho_\infty(Y|X) + \mathcal{O}(\varepsilon)$$

X: slow
Y: fast

$$\mathrm{BV} = \begin{pmatrix} \mathrm{slow}_1 \approx 0 \\ \vdots \\ \mathrm{slow}_K \approx 0 \\ \hline \mathrm{fast}_1 \\ \vdots \\ \mathrm{fast}_D \end{pmatrix} \longrightarrow \mathrm{SPBV} = \begin{pmatrix} \mathrm{slow}_1 \approx 0 \\ \vdots \\ \mathrm{slow}_K \approx 0 \\ \hline (1+\eta)\,\mathrm{fast}_1 \\ \vdots \\ (1+\eta)\,\mathrm{fast}_D \end{pmatrix} \quad \eta \sim \mathcal{N}(0,\sigma^2)$$

Way out: Stochastically Perturbed Bred Vectors (SPBVs)

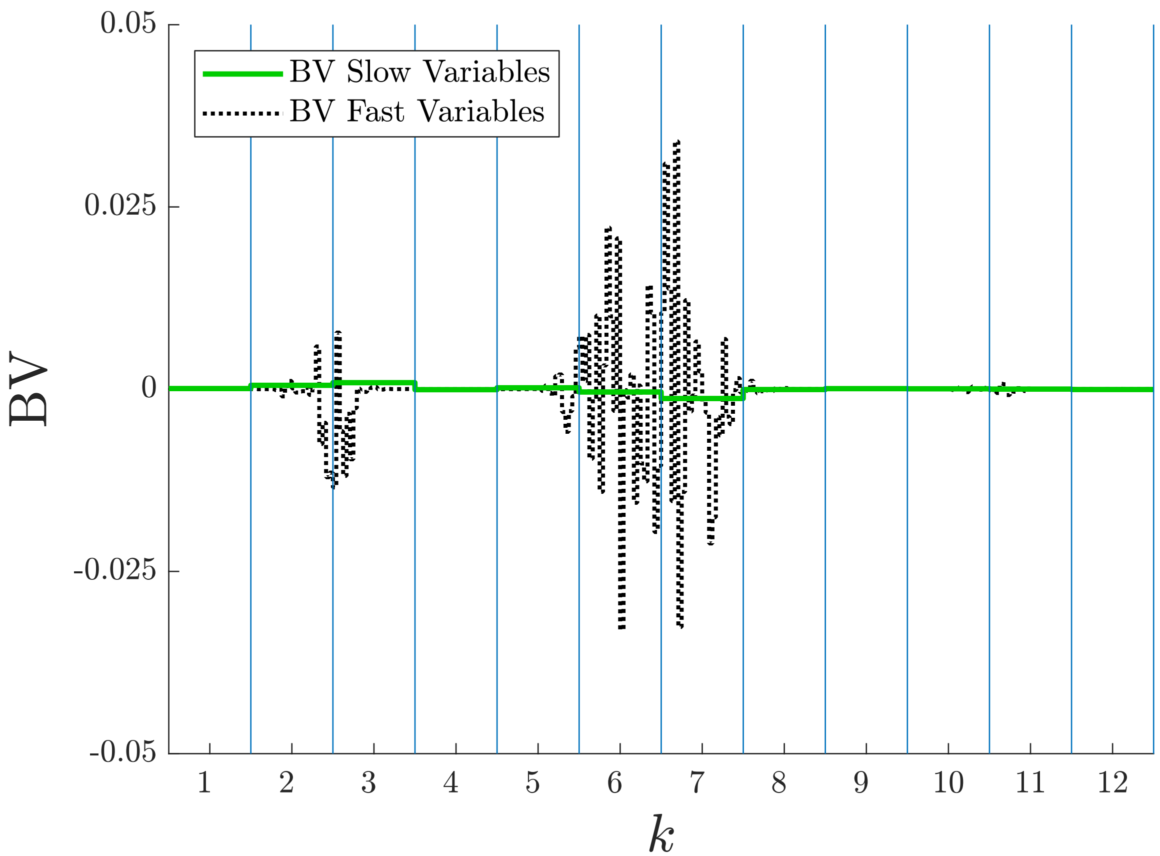
In multi-scale systems with time-scale separation $1/\varepsilon$ we can approxmiate

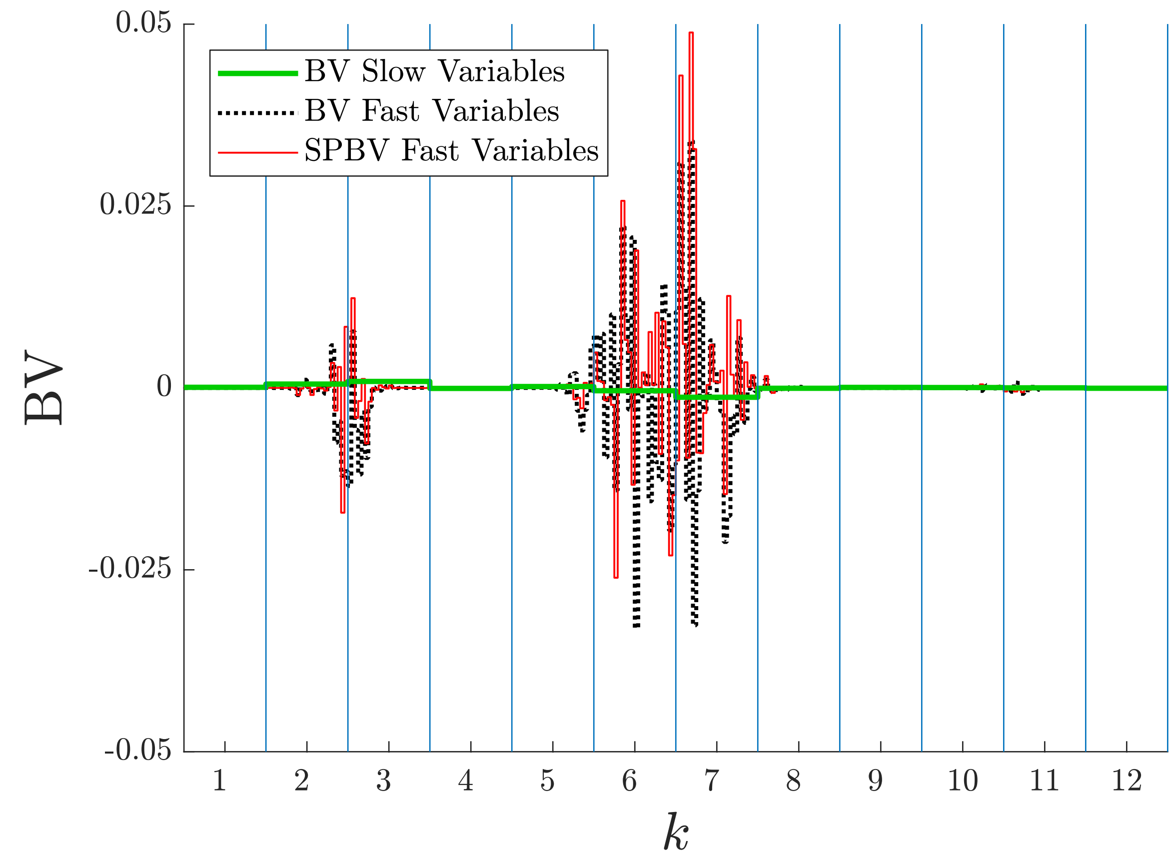
$$\rho(X,Y,t) = \hat{\rho}(X,t)\rho_\infty(Y|X) + \mathcal{O}(\varepsilon)$$

X: slow
Y: fast

$$\text{BV} = \left(\begin{array}{c} \text{slow}_1 \approx 0 \\ \vdots \\ \text{slow}_K \approx 0 \\ \hline \text{fast}_1 \\ \vdots \\ \text{fast}_D \end{array}\right) \longrightarrow \text{SPBV} = \left(\begin{array}{c} \text{slow}_1 \approx 0 \\ \vdots \\ \text{slow}_K \approx 0 \\ \hline (1+\eta)\,\text{fast}_1 \\ \vdots \\ (1+\eta)\,\text{fast}_D \end{array}\right)$$

$$\eta \sim \mathcal{N}(0,\sigma^2)$$



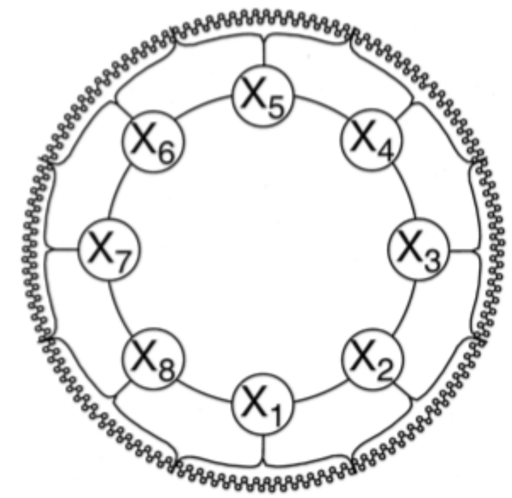


II - Ensemble forecasting

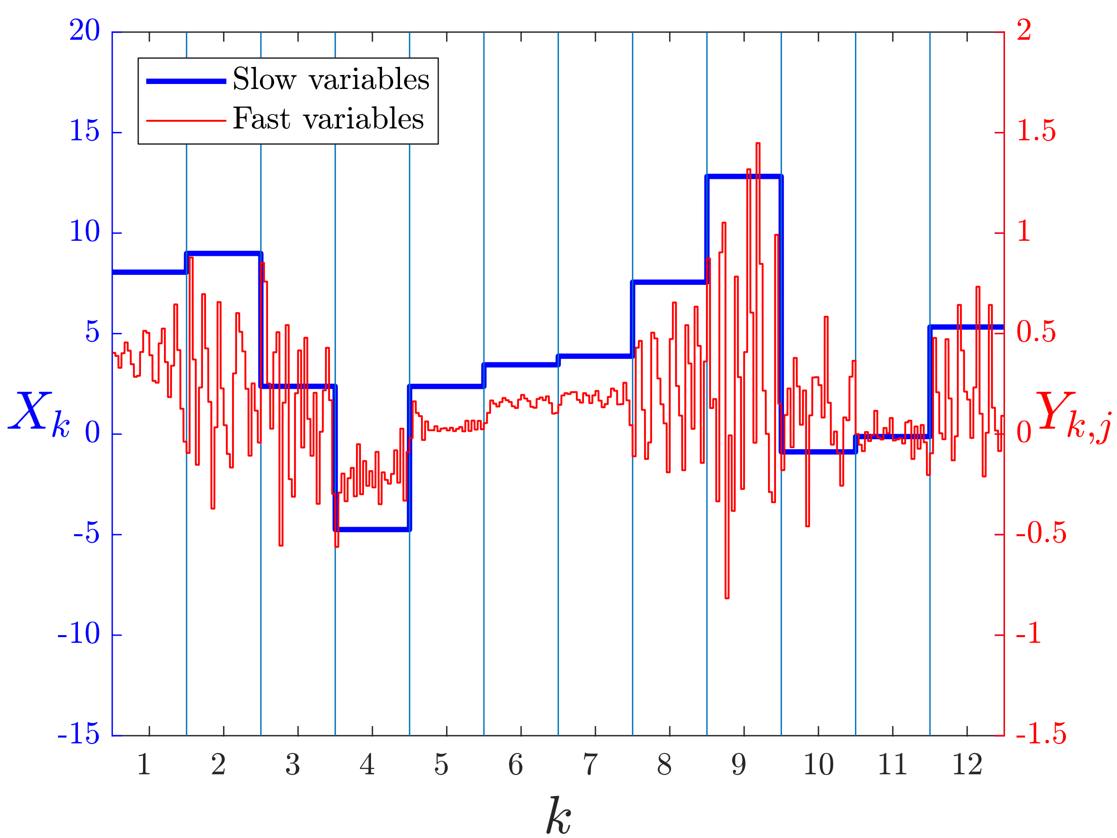
Multi-scale Lorenz 1996 model

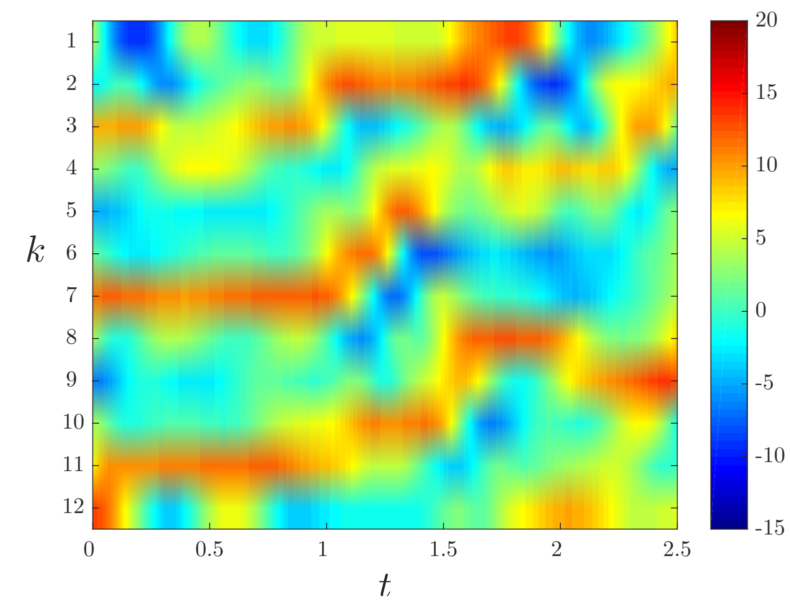
$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}$$

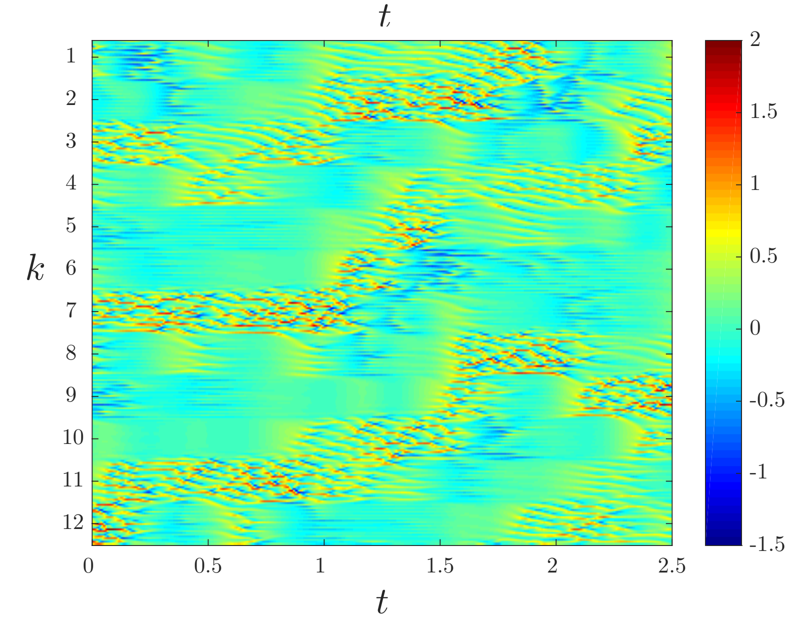
$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + \frac{hc}{b}X_k$$



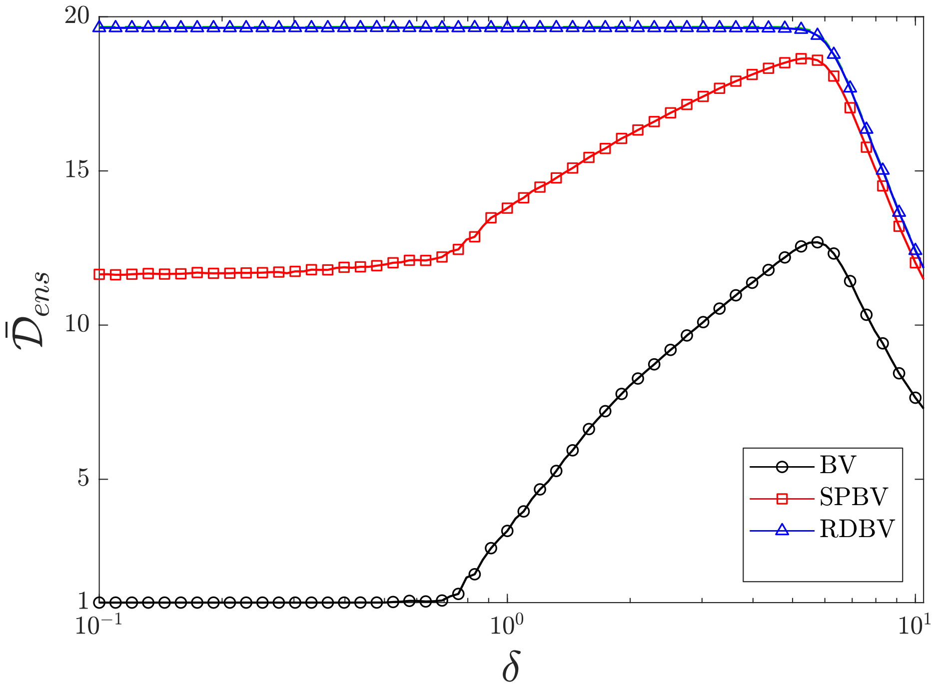
@ Wilks 2005

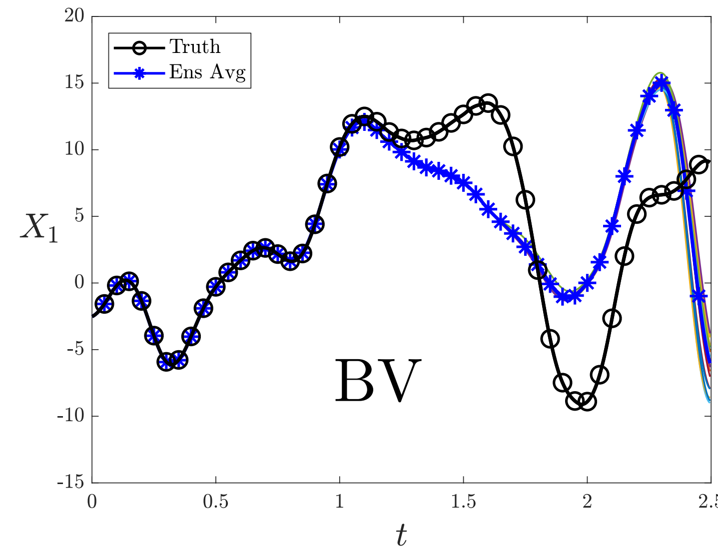


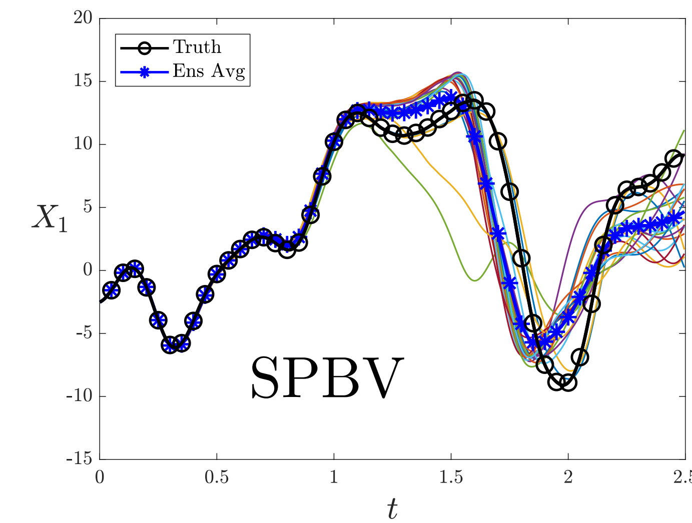




II - Ensemble forecasting





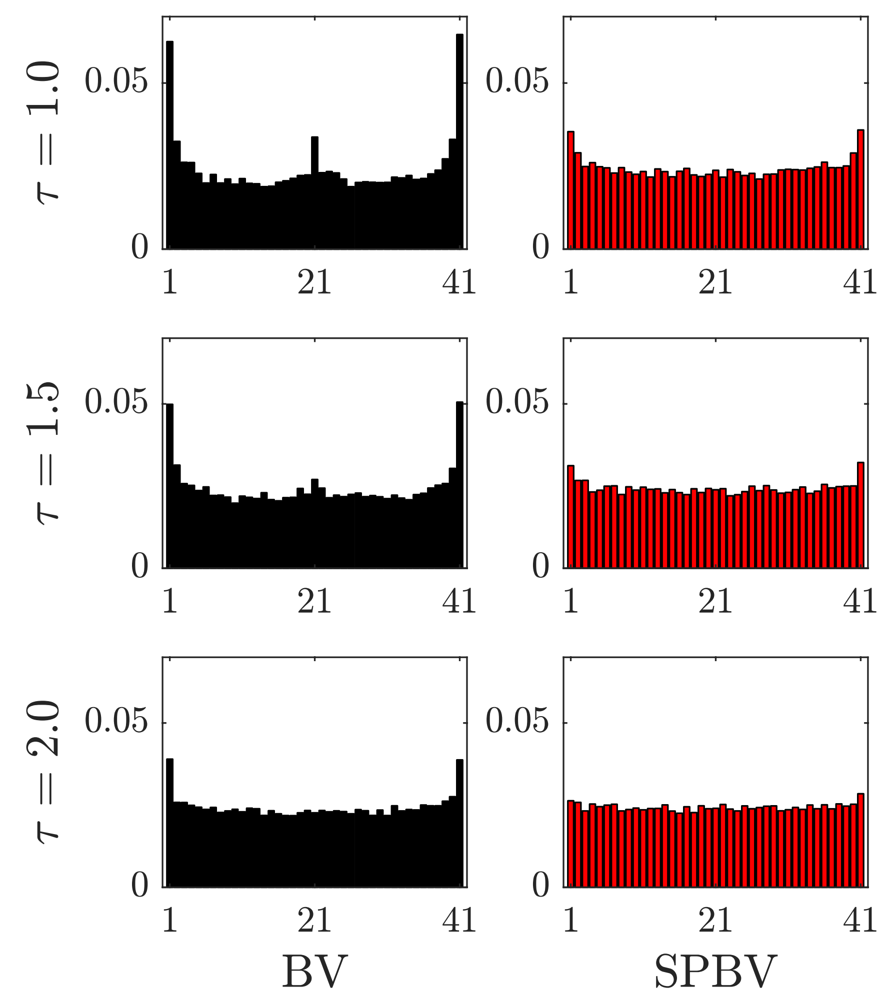


SPBVs

- good ensemble diversity
- reliable ensemble
- good forecast skill
- dynamically consistent
- computationally cheap

(Giggins and GAG, QJRMS (2019))

Talagrand diagram



III - Linear Response Theory

Given a chaotic dynamical system

parameter

$$\dot{x} = f(x, \varepsilon)$$

with a unique invariant physical measure μ_ε

What is the *change of the average* of an observable

$$\mathbb{E}^{\varepsilon}[\Psi] = \int_D \Psi(x)\, d\mu_\varepsilon$$

upon changing the parameter from its unperturbed state with ε_0 ?

$$\mathbb{E}^{\varepsilon}[\Psi] \approx \mathbb{E}^{\varepsilon_0}[\Psi] + \delta\varepsilon\, \mathbb{E}^{\varepsilon_0}[\Psi]' \qquad \varepsilon = \varepsilon_0 + \delta\varepsilon$$

using only information about the statistics of the unperturbed system

III - Linear Response Theory

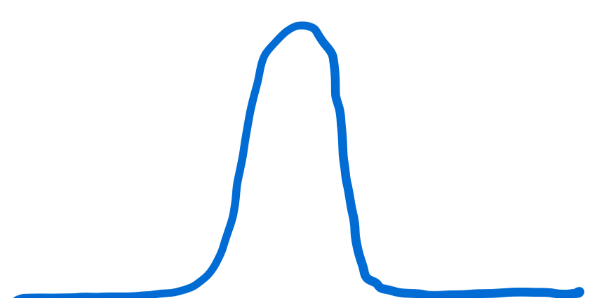
$$\mathbb{E}^{\varepsilon}[\Psi] \approx \mathbb{E}^{\varepsilon_0}[\Psi] + \delta\varepsilon\,\mathbb{E}^{\varepsilon_0}[\Psi]' \qquad \varepsilon = \varepsilon_0 + \delta\varepsilon$$

sufficient condition for linear response:

the invariant measure μ_ε is differentiable with respect to ε

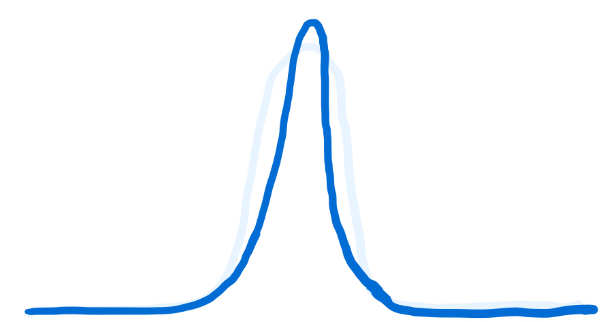
$$\mu_\varepsilon \approx \mu_{\varepsilon_0} + \mu'_\varepsilon(\varepsilon_0)\delta\varepsilon$$

<u>example:</u> Ornstein-Uhlenbeck process (stochastic)



$$dx = -\gamma x\,dt + \sigma\,dW_t$$

unperturbed



$$dx = -(\gamma + \varepsilon)x\,dt + \sigma\,dW_t$$

perturbed

Success stories in the Climate Sciences

Leith (1975)

toy models: *Majda et al ’07, ’10, Lucarini & Sarno ’11*

barotropic models: *Bell ’80, Gritsun & Dymnikov ’99, Abramov & Majda ‘09*

quasi-geostrophic models: Dymnikov & Gritsun ‘01

atmospheric models: *North et al ’04, Cionni et al ’04, Gritsun et al ’02/‘07, Gritsun & Branstator ’07, Ring & Plumb ’08, Gritsun ’10*

coupled climate models: *Langen & Alexeev ’05, Kirk & Davidoff ’09, Fuchs et al ’14, Ragone et al ‘15*

Success stories in the Climate Sciences

Leith (1975)

toy models: *Majda et al ’07, ’10, Lucarini & Sarno ’11*

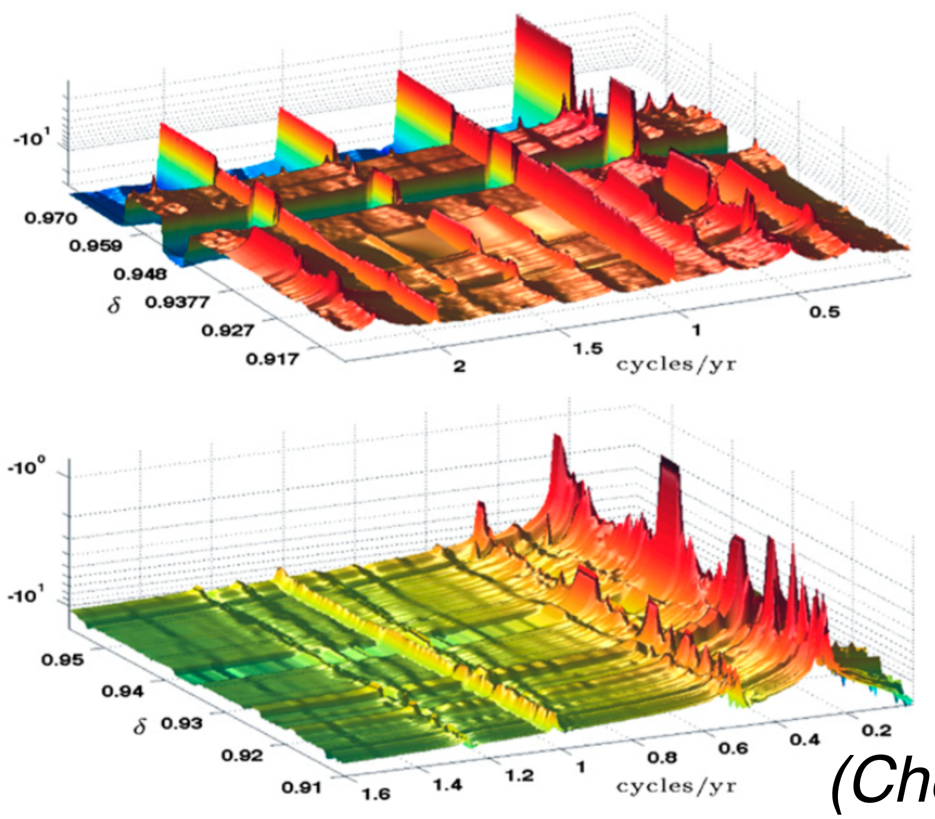
barotropic models: *Bell ’80, Gritsun & Dymnikov ’99, Abramov & Majda ‘09*

quasi-geostrophic models: Dymnikov & Gritsun ‘01

atmospheric models: *North et al ’04, Cionni et al ’04, Gritsun et al ’02/‘07, Gritsun & Branstator ’07, Ring & Plumb ’08, Gritsun ’10*

coupled climate models: *Langen & Alexeev ’05, Kirk & Davidoff ’09, Fuchs et al ’14, Ragone et al ‘15*

However, rough parameter dependency is known to exist in atmospheric and ocean dynamics



(Chekroun et al ’14)

Success stories in the Climate Sciences

Leith (1975)

toy models: *Majda et al '07, '10, Lucarini & Sarno '11*

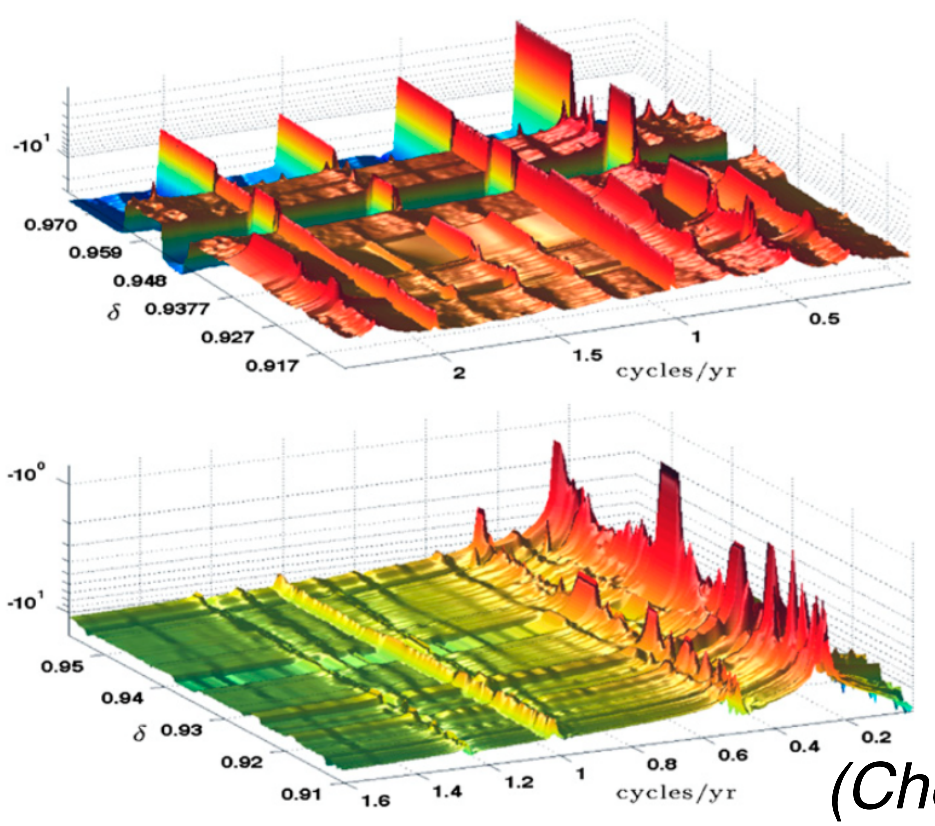
barotropic models: *Bell '80, Gritsun & Dymnikov '99, Abramov & Majda '09*

quasi-geostrophic models: Dymnikov & Gritsun '01

atmospheric models: *North et al '04, Cionni et al '04, Gritsun et al '02/'07, Gritsun & Branstator '07, Ring & Plumb '08, Gritsun '10*

coupled climate models: *Langen & Alexeev '05, Kirk & Davidoff '09, Fuchs et al '14, Ragone et al '15*

However, rough parameter dependency is known to exist in atmospheric and ocean dynamics



(Chekroun et al '14)

Note: even if linear response is **not** valid, this might not be detectable in a finite time series

(GAG, Wormell & Wouters '17)

What is known analytically?

- statistical mechanics: *Kubo '66* ✓
- stochastic dynamical systems: *Hänggi '78, Hairer & Majda '10* ✓

What is known analytically?

- statistical mechanics: *Kubo '66* ✓
- stochastic dynamical systems: *Hänggi '78, Hairer & Majda '10* ✓
- forced-dissipative deterministic dynamical systems (singular measures):
  - Axiom A (uniformly hyperbolic): *Ruelle '97, '98* ✓

What is known analytically?

- statistical mechanics: *Kubo '66* ✓
- stochastic dynamical systems: *Hänggi '78, Hairer & Majda '10* ✓
- forced-dissipative deterministic dynamical systems (singular measures):
  - Axiom A (uniformly hyperbolic): *Ruelle '97, '98* ✓
  - what about more general dynamical systems?

Baladi et al '08, '10, '14, '15 …

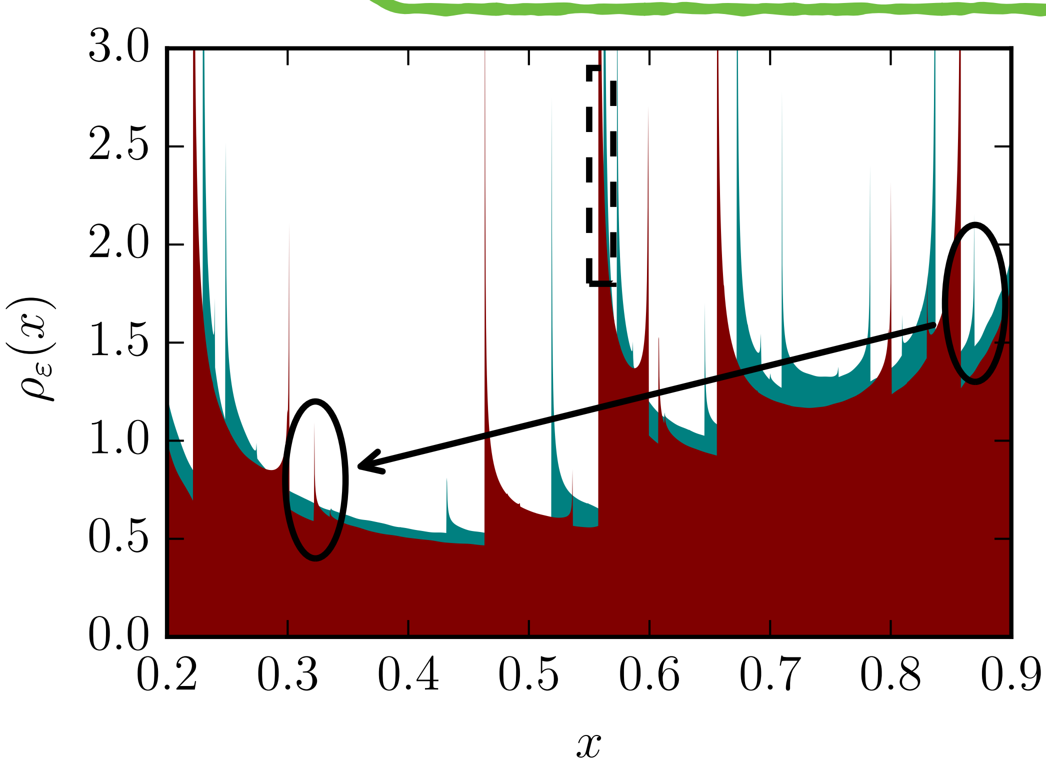
no linear response for the logistic map

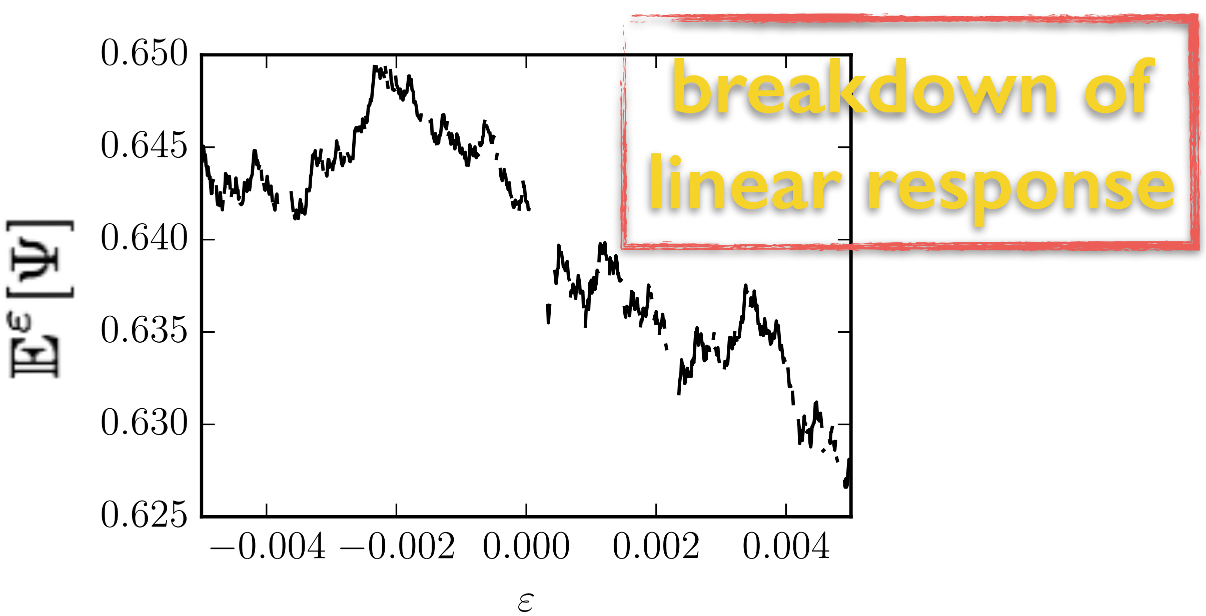
What is known analytically?

- statistical mechanics: *Kubo'66* ✓
- stochastic dynamical systems: *Hänggi '78, Hairer & Majda '10* ✓
- forced-dissipative deterministic dynamical systems (singular measures):
  - Axiom A (uniformly hyperbolic): *Ruelle '97, '98* ✓
  - what about more general dynamical systems?

Baladi et al '08, '10, '14, '15 …

no linear response for the logistic map





What is known analytically?

- statistical mechanics: *Kubo'66* ✓
- stochastic dynamical systems: *Hänggi '78, Hairer & Majda '10* ✓
- forced-dissipative deterministic dynamical systems (singular measures):
  - Axiom A (uniformly hyperbolic): *Ruelle '97, '98* ✓
  - what about more general dynamical systems?

Baladi et al '08, '10, '14, '15 …

no linear response for the logistic map

We address here the following conundrum

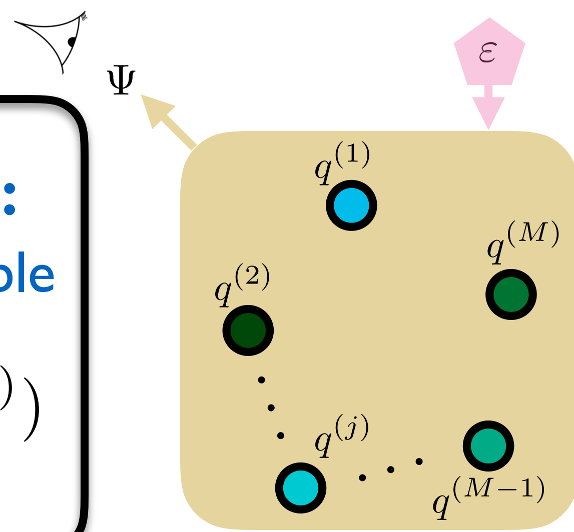
How can typical observables of high-dimensional systems obey linear response when their microscopic low-dimensional constituents typically do not?

III - Linear Response Theory

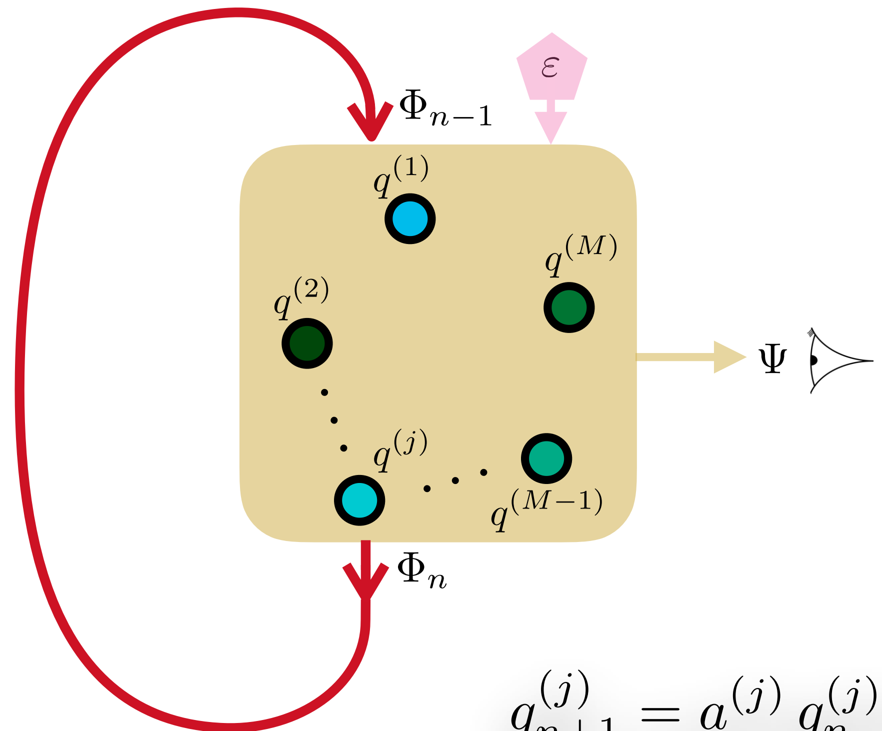
Object of interest:

macroscopic observable

$$\Psi_n = \frac{1}{M}\sum_{j=1}^{M}\psi(q_n^{(j)})$$



or

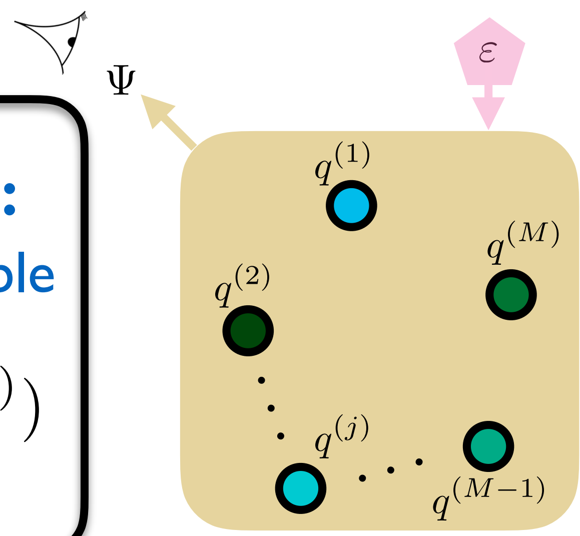


$$q_{n+1}^{(j)} = a^{(j)}\, q_n^{(j)}(1 - q_n^{(j)})$$

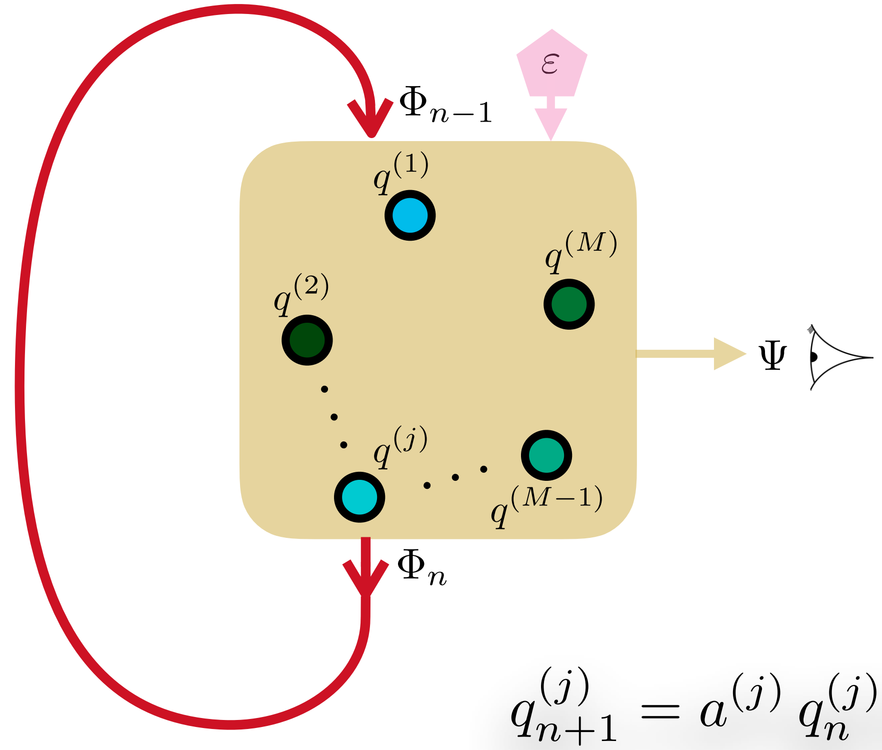
III - Linear Response Theory

Object of interest:
macroscopic observable

$$\Psi_n = \frac{1}{M}\sum_{j=1}^{M}\psi(q_n^{(j)})$$



or



$$q_{n+1}^{(j)} = a^{(j)}\, q_n^{(j)}(1 - q_n^{(j)})$$

statistical limit laws: $\Psi_n = \mathbb{E}\Psi + \frac{1}{\sqrt{M}}\zeta_n + o(1/\sqrt{M})$

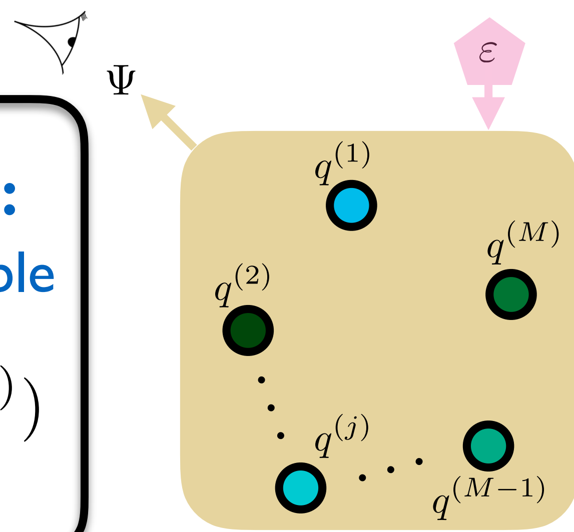
Gaussian process

III - Linear Response Theory

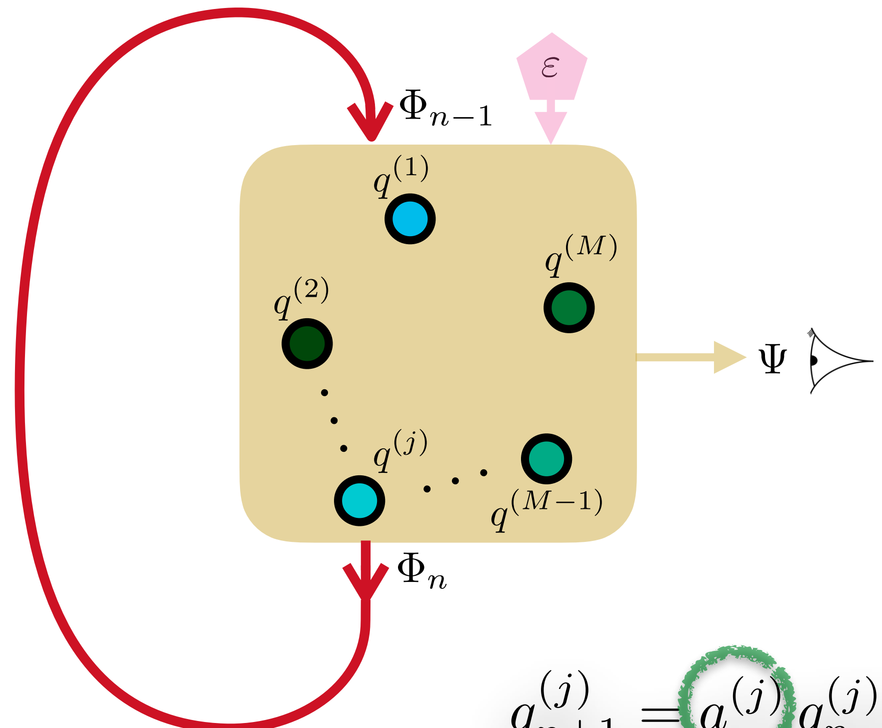
Object of interest:

macroscopic observable

$$\Psi_n = \frac{1}{M}\sum_{j=1}^{M}\psi(q_n^{(j)})$$



or



$$q_{n+1}^{(j)} = a^{(j)} q_n^{(j)}(1 - q_n^{(j)})$$

heterogeneity $\rightarrow$ $a^{(j)} \sim \nu(a)$

statistical limit laws: $\Psi_n = \mathbb{E}\Psi + \frac{1}{\sqrt{M}}\zeta_n + o(1/\sqrt{M})$

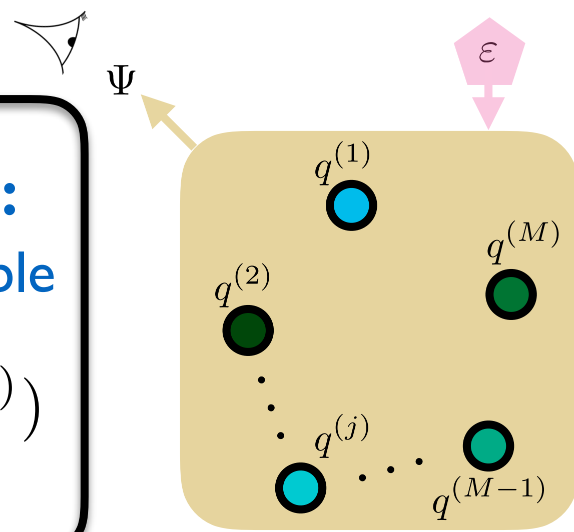
Gaussian process

$$\mathbb{E}\Psi_n = \langle \mathbb{E}\Psi_n \rangle + \frac{1}{\sqrt{M}}\eta + o(1/\sqrt{M})$$

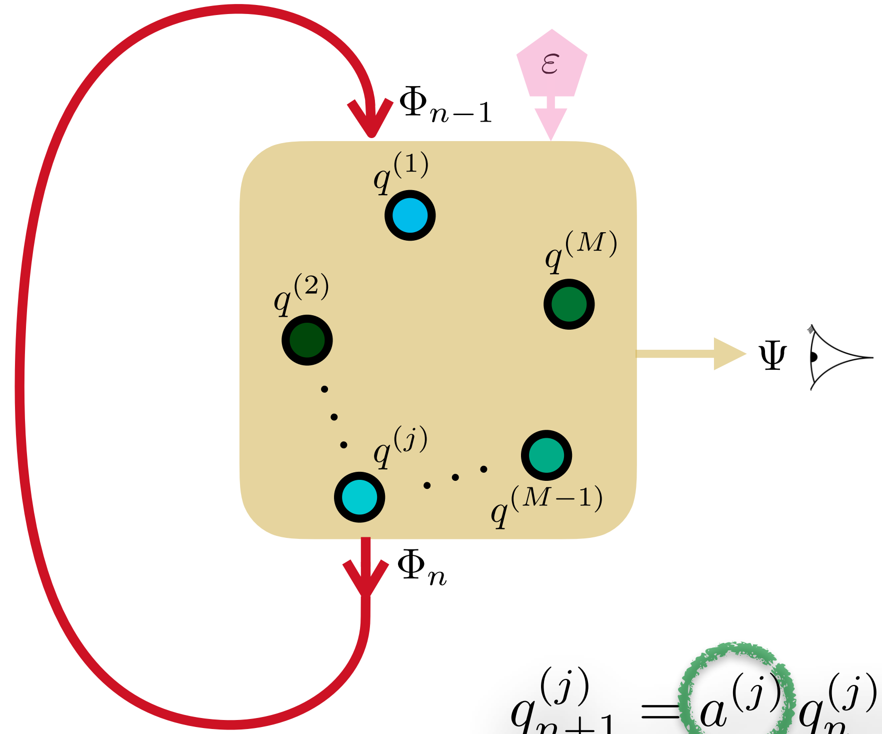
III - Linear Response Theory

Object of interest:
macroscopic observable

$$\Psi_n = \frac{1}{M}\sum_{j=1}^{M} \psi(q_n^{(j)})$$



or



$$q_{n+1}^{(j)} = a^{(j)} q_n^{(j)}(1-q_n^{(j)})$$

heterogeneity $\rightarrow$ $a^{(j)} \sim \nu(a)$

statistical limit laws: $\Psi_n = \mathbb{E}\Psi + \frac{1}{\sqrt{M}}\zeta_n + o(1/\sqrt{M})$ *Gaussian process*

$$\mathbb{E}\Psi_n = \langle \mathbb{E}\Psi_n \rangle + \frac{1}{\sqrt{M}}\eta + o(1/\sqrt{M})$$

Linear response holds for macroscopic observables provided

- Ψ_n is a stochastic process (diffusive limit)
- the $a^{(j)}$ are distributed according to a sufficiently smooth distribution $\nu(a)$ (heterogeneity)

IV - Numerical integration of multi-scale systems

How does the numerical time integrator affect the statistical behaviour of the simulation?

Discretisation

$$\dot{x} = \varepsilon\, h(x) f_0(y) + \varepsilon^2\, f(x,y)$$
$$\dot{y} = g(y)$$

$$x_{n+1} = x_n + \Delta t \varepsilon\, h(x_n) f_0(y_n) + \Delta t \varepsilon^2\, f(x_n, y_n)$$

Discretisation

$$\dot{x} = \varepsilon\, h(x) f_0(y) + \varepsilon^2\, f(x,y)$$
$$\dot{y} = g(y)$$

$$x_{n+1} = x_n + \Delta t \varepsilon\, h(x_n) f_0(y_n) + \Delta t \varepsilon^2\, f(x_n, y_n)$$

Homogenisation *(GAG & Melbourne (2013))* **Homogenisation**

$$dX = F(X)\, dt + \sigma h(X) \circ dW_t$$
$$F(X) = \int_\Lambda f(X,y)\, d\mu$$
$$\frac{1}{2}\sigma^2 = \int_0^\infty \mathbb{E}[f_0(y) f_0(\varphi^t y)]\, dt$$

$$dX = \left(F(X) - \frac{1}{2}\Delta t\, h(X) h'(X)\, \mathbb{E}[f_0^2]\right) dt + \sqrt{\Delta t}\, \hat{\sigma} h(X) \circ d\tilde{W}_t$$
$$F(X) = \int_\Lambda f(X,y)\, d\mu$$
$$\hat{\sigma}^2 = \mathbb{E}[f_0^2] + 2 \sum_{n=1}^\infty \mathbb{E}[f_0(y) f_0(\Phi^n\, y)]$$

Discretisation

$$\dot{x} = \varepsilon\, h(x) f_0(y) + \varepsilon^2\, f(x,y)$$
$$\dot{y} = g(y)$$

$$x_{n+1} = x_n + \Delta t \varepsilon\, h(x_n) f_0(y_n) + \Delta t \varepsilon^2\, f(x_n, y_n)$$

Homogenisation

Homogenisation

$$dX = F(X)\, dt + \sigma h(X) \circ dW_t$$
$$F(X) = \int_\Lambda f(X,y)\, d\mu$$
$$\frac{1}{2}\sigma^2 = \int_0^\infty \mathbb{E}[f_0(y) f_0(\varphi^t y)]\, dt$$

$$dX = \Big(F(X) - \frac{1}{2}\Delta t\, h(X) h'(X)\, \mathbb{E}[f_0^2]\Big)\, dt + \sqrt{\Delta t}\, \hat{\sigma} h(X) \circ d\tilde{W}_t$$
$$F(X) = \int_\Lambda f(X,y)\, d\mu$$
$$\hat{\sigma}^2 = \mathbb{E}[f_0^2] + 2 \sum_{n=1}^{\infty} \mathbb{E}[f_0(y) f_0(\Phi^n\, y)]$$

Remarks: $\hat{\sigma}^2\, \Delta t \to \sigma^2$ for $\Delta t \to 0$

Discretisation

$$\dot{x} = \varepsilon\, h(x) f_0(y) + \varepsilon^2\, f(x,y)$$
$$\dot{y} = g(y)$$

$$x_{n+1} = x_n + \Delta t \varepsilon\, h(x_n) f_0(y_n) + \Delta t \varepsilon^2\, f(x_n, y_n)$$

Homogenisation

$$dX = F(X)\, dt + \sigma h(X) \circ dW_t$$
$$F(X) = \int_\Lambda f(X,y)\, d\mu$$
$$\frac{1}{2}\sigma^2 = \int_0^\infty \mathbb{E}[f_0(y) f_0(\varphi^t y)]\, dt$$

Homogenisation

$$dX = \left(F(X) - \frac{1}{2}\Delta t\, h(X) h'(X)\, \mathbb{E}[f_0^2]\right) dt + \sqrt{\Delta t}\, \hat{\sigma} h(X) \circ d\tilde{W}_t$$
$$F(X) = \int_\Lambda f(X,y)\, d\mu$$
$$\hat{\sigma}^2 = \mathbb{E}[f_0^2] + 2 \sum_{n=1}^{\infty} \mathbb{E}[f_0(y) f_0(\Phi^n\, y)]$$

Remarks: $\hat{\sigma}^2\, \Delta t \to \sigma^2$ for $\Delta t \to 0$

noise is neither Stratonovich nor Itô

$$\mathbf{E} := -\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\Delta}\mathbf{t}\; \mathbf{h}(\mathbf{X})\mathbf{h}'(\mathbf{X})\, \mathbb{E}[\mathbf{f_0^2}]$$

for *i.i.d.* fast dynamics, i.e. $\hat{\sigma}^2 = \mathbb{E}[f_0^2]$, the noise is Itô
(dynamics is already rough on time scale of $\mathcal{O}(\Delta t)$)

but it is never Stratonovich!

IV - Numerical integration of multi-scale systems

The only difference between the two homogenised equations is

$$\mathbf{E} := -\frac{1}{2}\Delta t\, h(X) h'(X)\, \mathbb{E}[f_0^2]$$

How can we interpret this extra drift term in the homogenised equation of the discretisation?

Backward error analysis:

appears in first-order schemes, but not in higher order schemes

(Frank & GAG, SIAM MMS (2018))

Can the extra term be significant? It is only $\mathcal{O}(\Delta t)$

$$\dot{x} = \varepsilon\sqrt{x}y + \varepsilon^2 b(c-x)y^2$$

$$y = \eta + \zeta$$

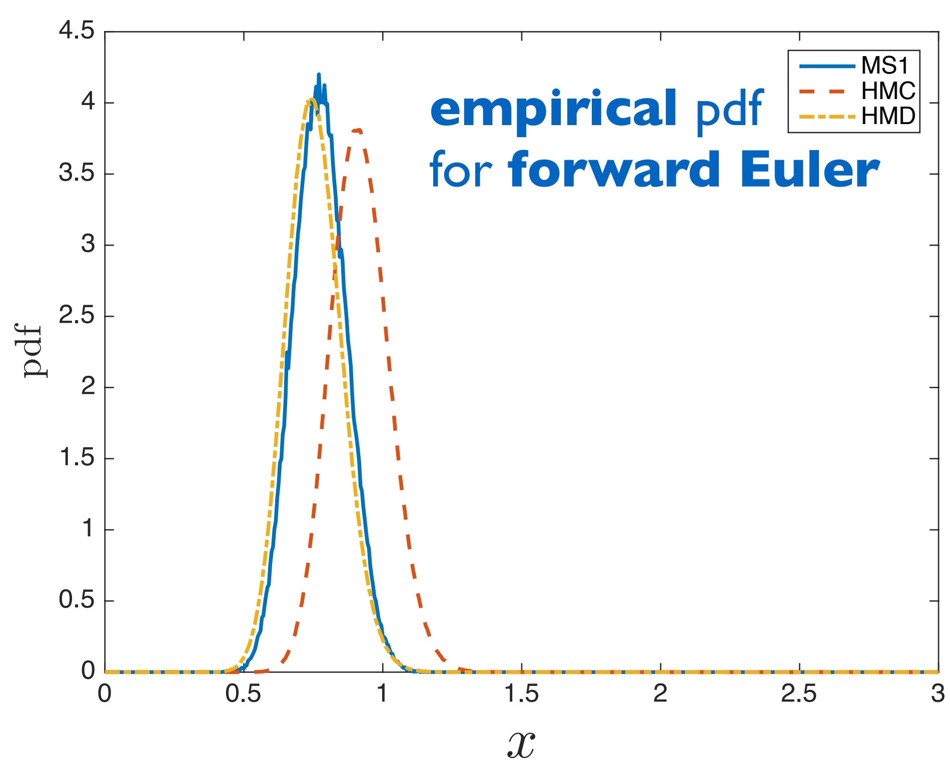
$$\begin{cases} \dot{\xi} = -\eta - \zeta \\ \dot{\eta} = \xi + r\eta \\ \dot{\zeta} = s + (\xi - u)\zeta \end{cases}$$

(fast Rössler system)

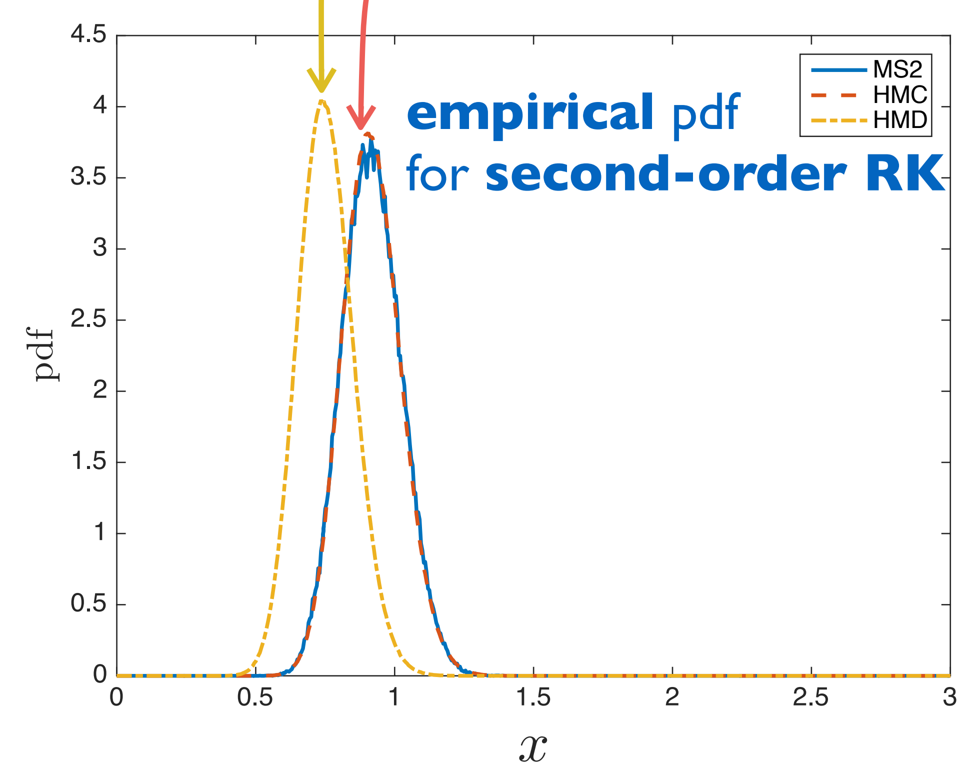
$r = s = 0.25$
$u = 7$

pdf of homogenised equation
for full **continuous**-time multi-scale system

pdf of homogenised equation
for full **discrete**-time multi-scale system



$\varepsilon = 0.1$



15.6% error in mean!

V - The problem of parametrising small-scale convection

The inadequate representation of atmospheric convection in GCMs leads to

- considerable uncertainty in estimating climate sensitivity
- ambiguities in the numerical simulation of the Earth's climate, for example when comparing the inter-model mean and spread of hydrological-cycle related variables of the CMIP5 ensemble to observations.

Deterministic convective parametrisation:

- assumes single possible response of the small-scale convective state for given large-scale atmosphere-ocean state
- capable of only representing a mean effect of convective processes
- lack of variability at small scales (can propagate upscale)
- increase in spatial resolution does not allow for sufficient number of convective events to justify an average

Stochastic convective parametrisation:

physics-based

Buizza et al (1999), Lin & Neelin (2000, 2003), Berner et al (2005), Khouider et al (2002, 2003), Stechmann & Neelin (2011) and many others

data-driven

Horenko (2011): data-based Markov chain
Dorrestijn et al (2013, 2015): data-driven multi-cloud model; and not so many others

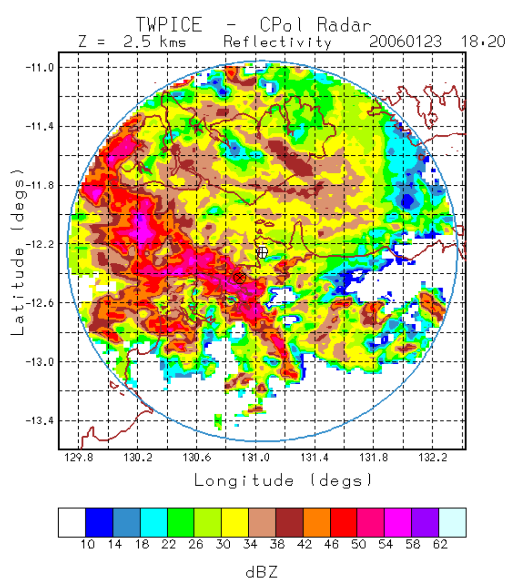
transparency ⟷ accuracy

Observational Data

Two data sets at Darwin and Kwajalein

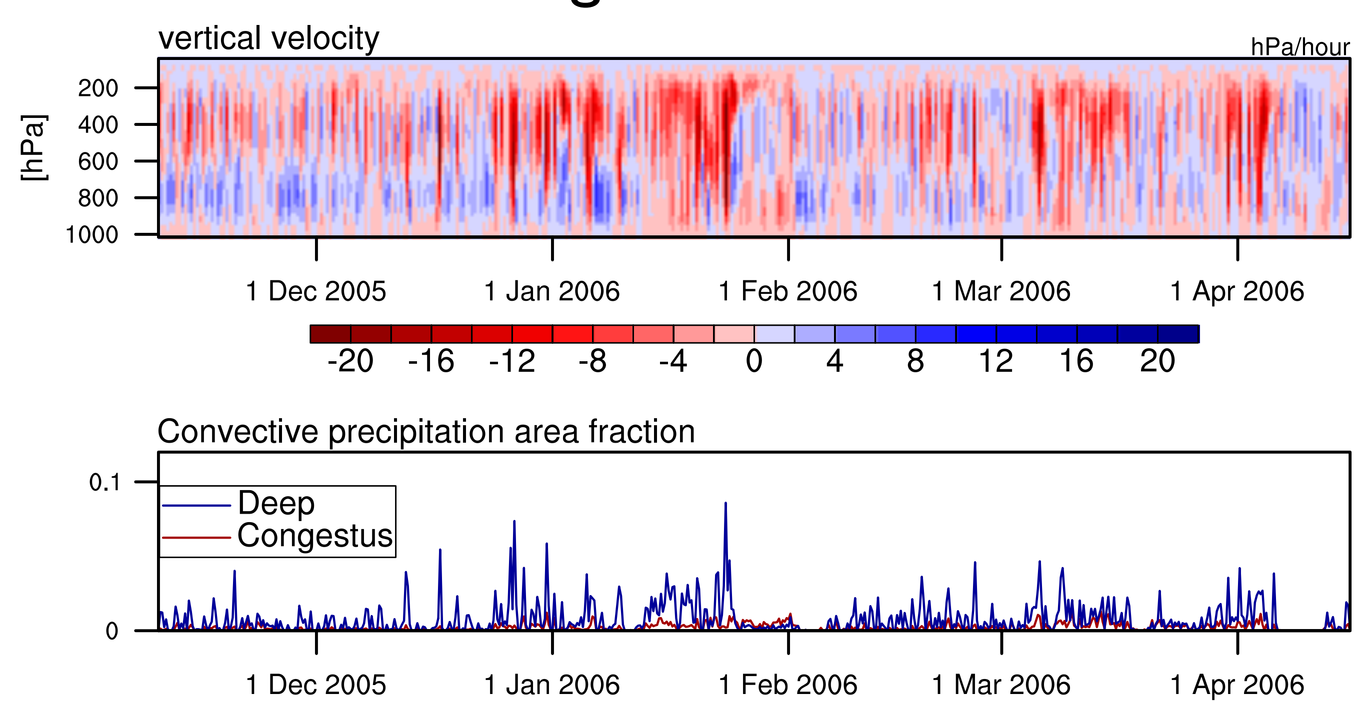
- large scale vertical velocity ω
  small scale convective activity (convective area fraction)
- 6-hourly time resolution
- 190 x 190 km^2 (typical size of GCM grid box)
- Darwin has 1890 and Kwajalein has 1095 data points
- At Darwin observations from consecutive wet seasons (2004/2005, 2005/2006, 2006/2007), with a total of 1890 6-hour means. Over Kwajalein, the analysis is applied to the time period of May 2008 - Jan 2009 yielding 1095 6-hour means

Precipitation radar observations combined with ECMWF analysis



Darwin, radar reflectivity @2.5km. 23 Jan 2006
Image courtesy: P. May

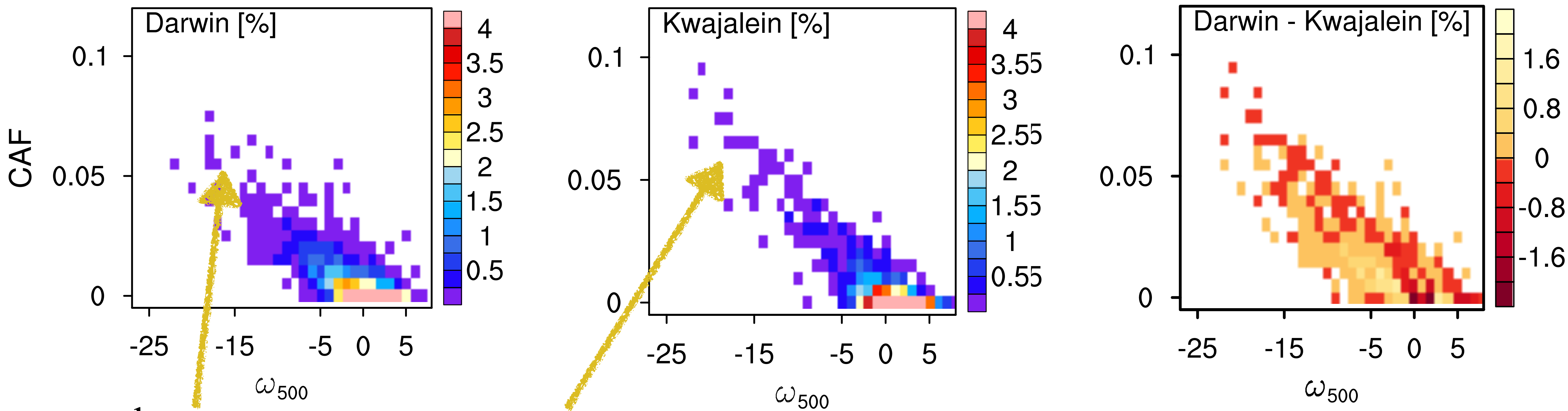
Large scale field: ω



Small scale field: CAF

Davies et al (2013)

The Differences

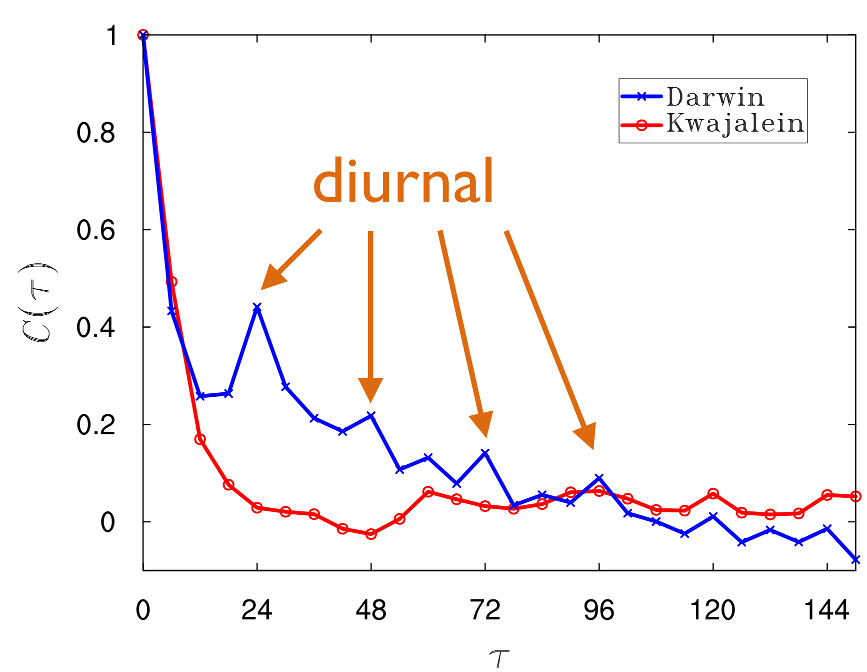


$\frac{\text{std}}{\text{mean}}$ of CAF decreases for sufficiently negative ω_{500}

→ heavy rain events behave deterministically with approximate linear behaviour

Kwajalein has a purely oceanic weather regime

Darwin features land-sea breeze induced convection (diurnal cycles)



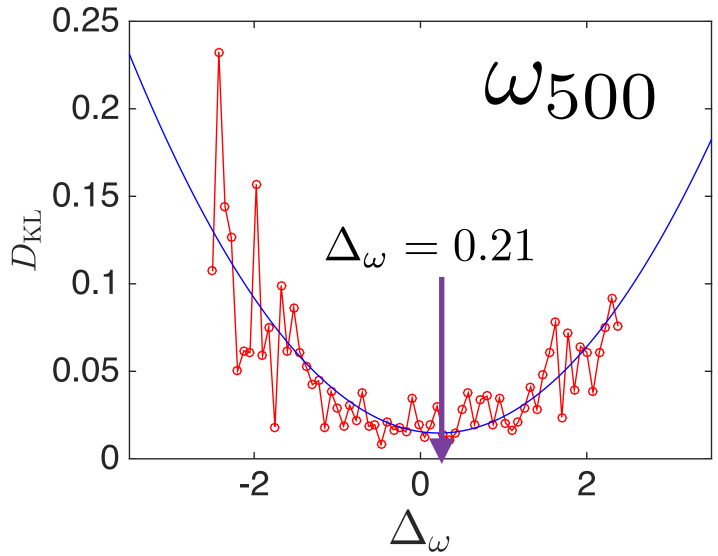
The Similarities

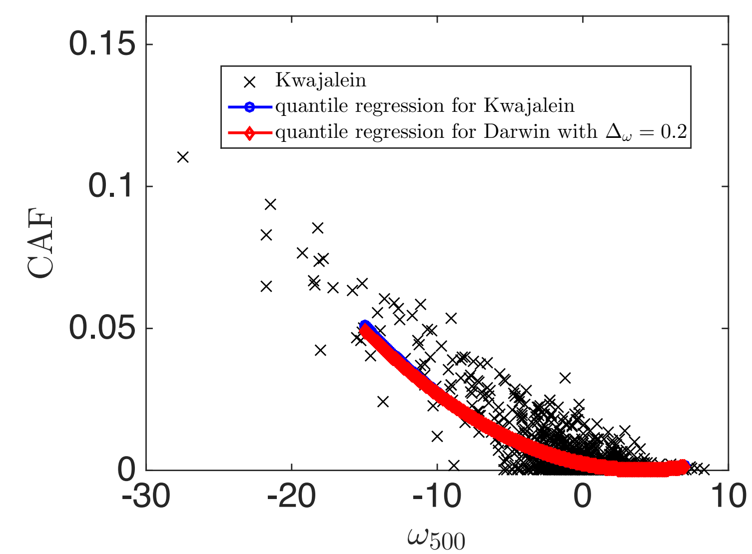
$$p^{\text{Kwajalein}}(\text{CAF}(t)|\omega_{500}(t)) \approx p^{\text{Darwin}}(\text{CAF}(t)|\omega_{500}(t) - \Delta_\omega)$$

or analogously

$$p^{\text{Darwin}}(\text{CAF}(t)|\omega_{500}(t)) \approx p^{\text{Kwajalein}}(\text{CAF}(t)|\omega_{500}(t) + \Delta_\omega)$$

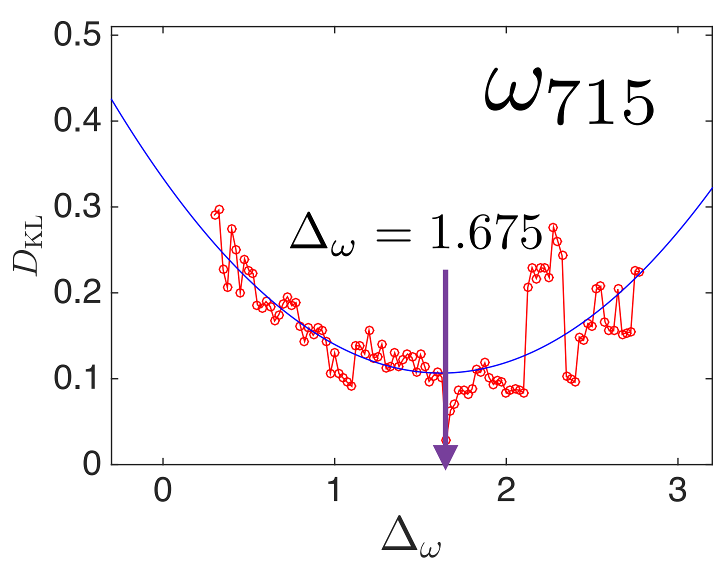
linear shift

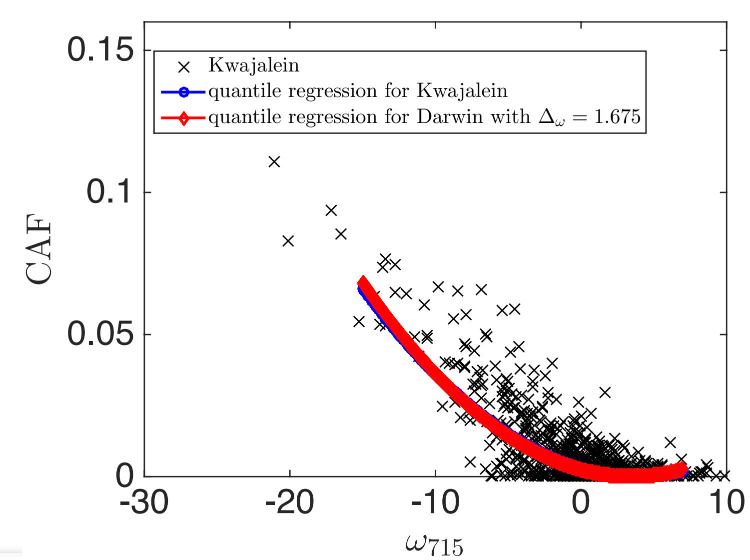




$$D_{\text{KL}}(P||Q) = \int \log\left(\frac{P(x)}{Q(x)}\right) P(x)\,dx$$

second-order median regression



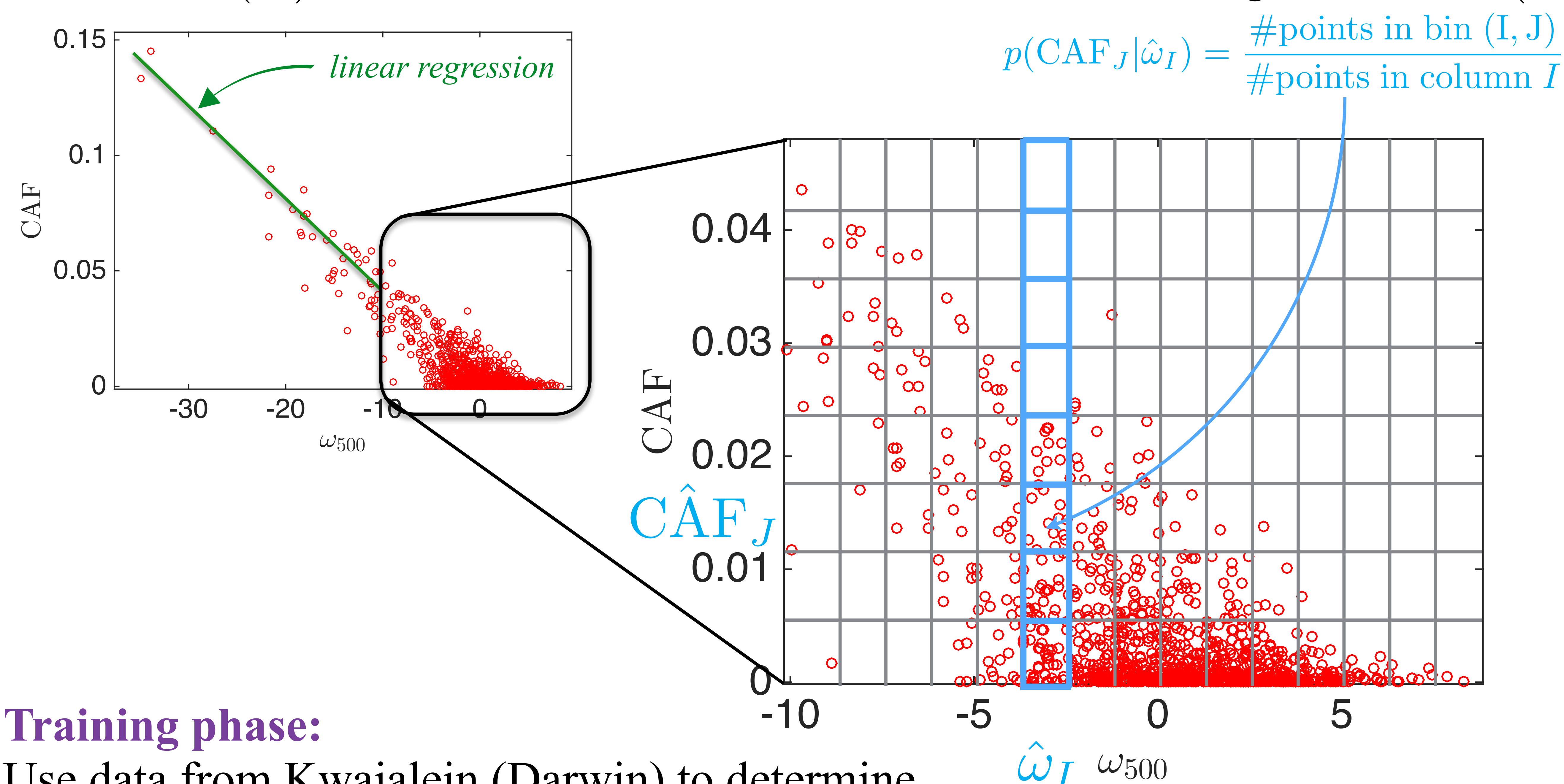


Despite the different prevalent atmospheric and oceanic regimes at the two locations, the empirical measure for the convective variables conditioned on large-scale mid-level vertical velocities for the two locations are close

This allows us to train the stochastic models at one location and then apply it to the other!

Instantaneous Random Variables

Treat $\mathrm{CAF}(t_k)$ as a random variable conditioned on the large-scale $\omega_{500}(t_k)$

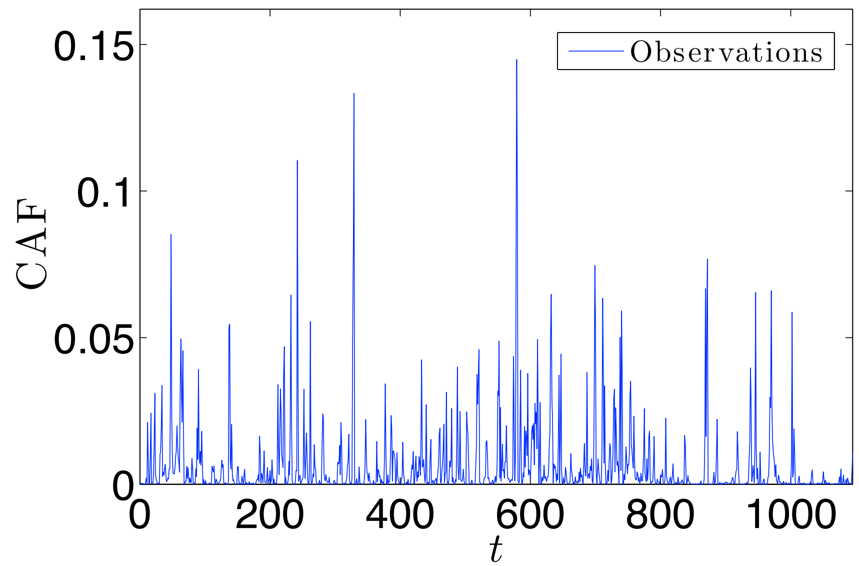


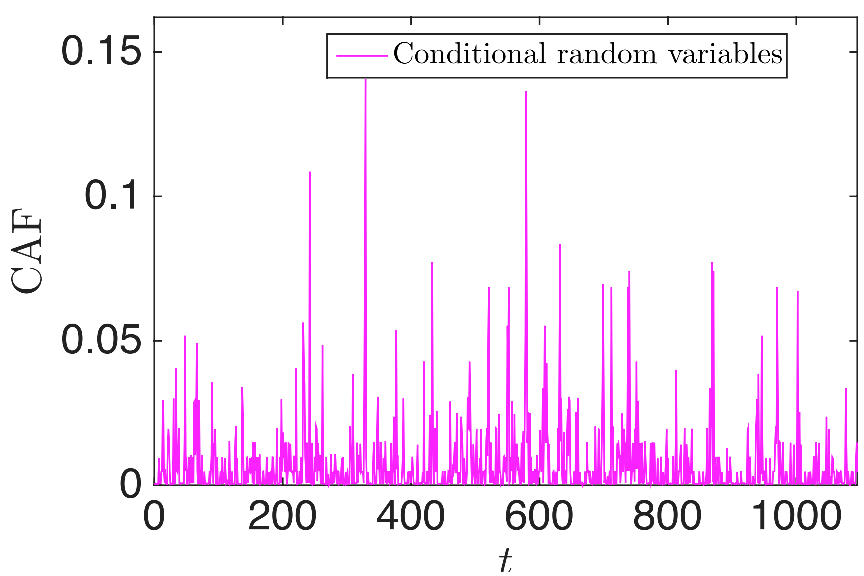
Training phase:

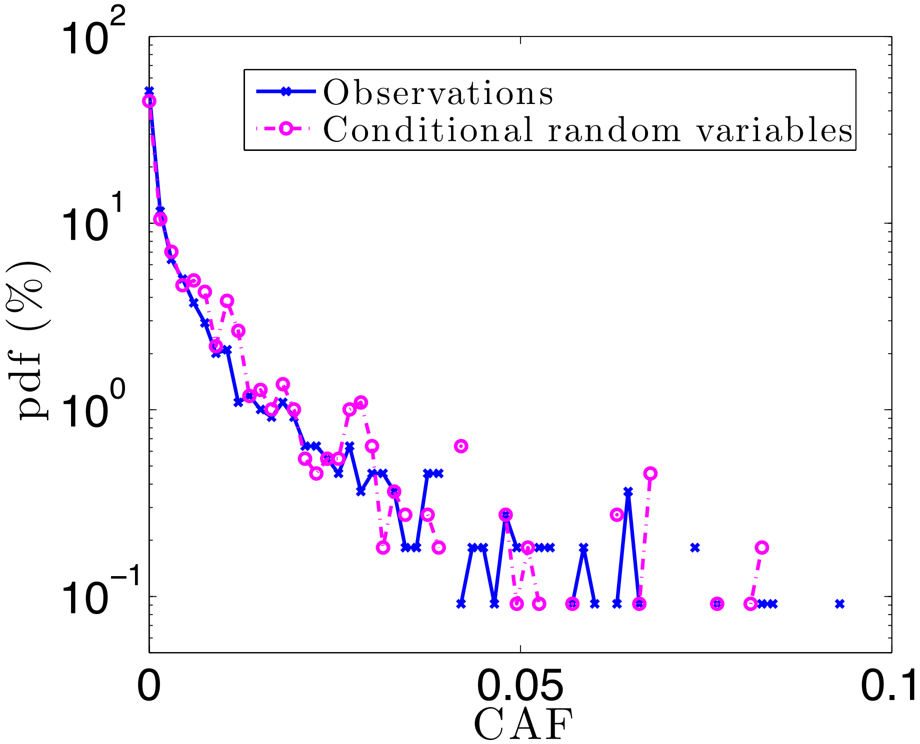
Use data from Kwajalein (Darwin) to determine the conditional probability $p(\mathrm{CAF}_J|\hat{\omega}_I)$

Application phase:

Draw CAF as random variables conditioned on observations of ω_{500} at Kwajalein (Darwin)







Mean μ, variance σ^2 and skewness ξ of CAF conditioned on ω_{500}

	μ	σ^2	ξ
observations	0.0066	$1.89\,10^{-4}$	4.27
random variable	0.0073	$1.80\,10^{-4}$	4.29

Trained at Darwin and applied to Kwajalein

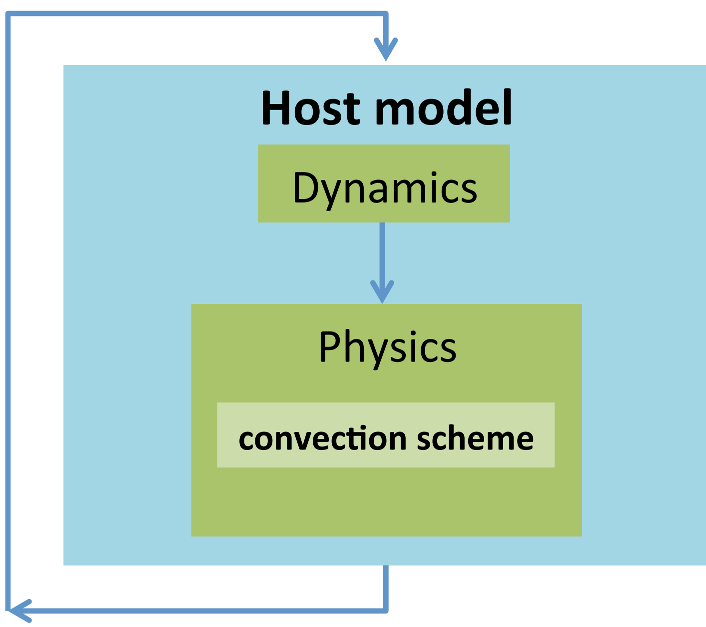
	μ	σ^2	ξ
observations	0.0080	$1.29\,10^{-4}$	2.38
random variable	0.0075	$1.45\,10^{-4}$	2.46

Trained at Kwajalein and applied to Darwin

- similarly good results for conditioning on ω_{715} or on rain rates
- cannot resolve periods of sustained non-convection near t=900

(GAG, Peters and Davies, QJRMS (2016))

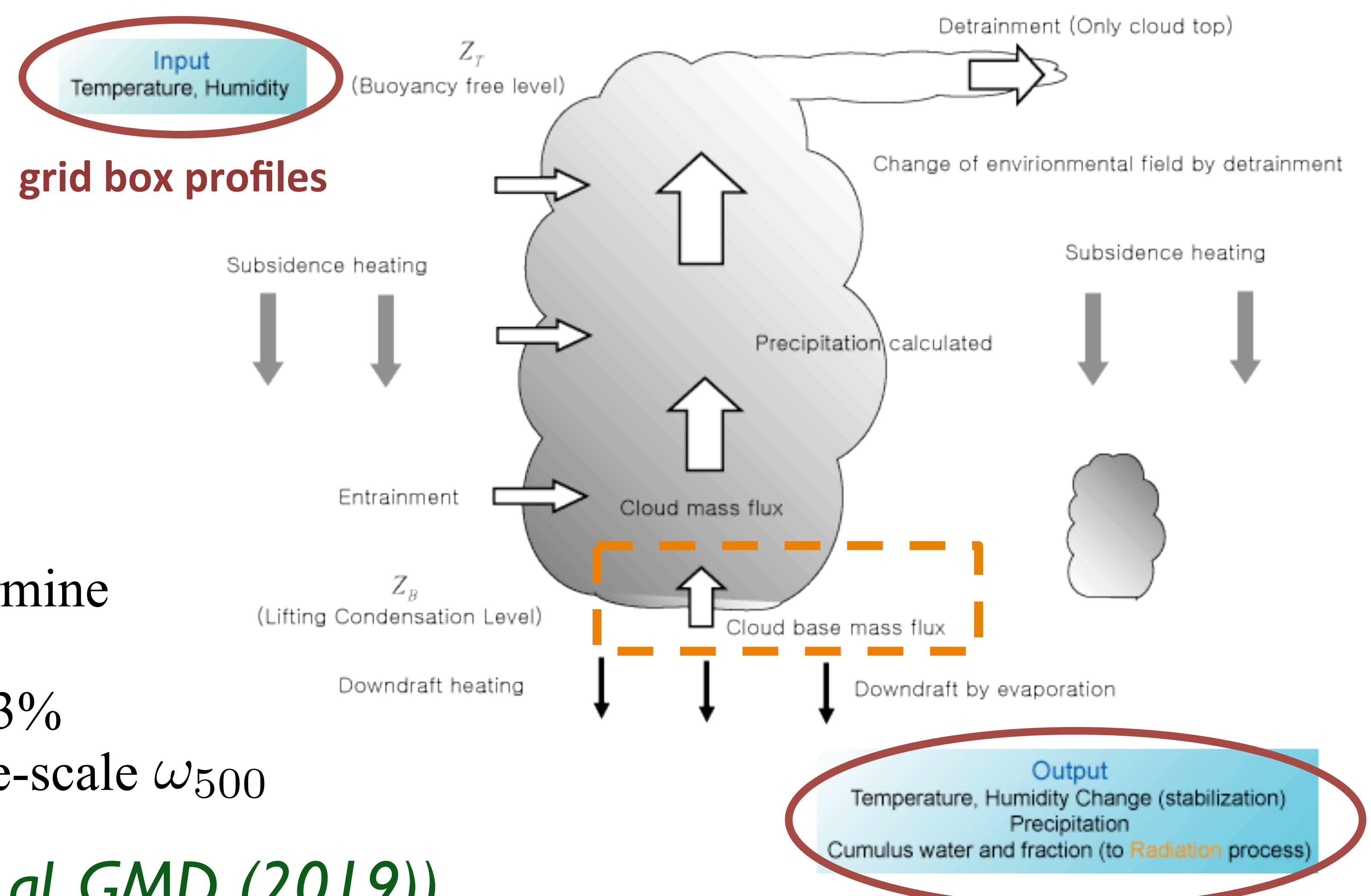
How can these stochastic parametrizations be used?



The convection scheme

- receives large-scale atmospheric state per grid box (temperature, velocities, humidity,…)
- computes vertical transport of heat, moisture
- provides tendencies to update large-scale fields

The highly challenging problem of triggering convection is performed by the convection scheme



adapted from
http://climate.snu.ac.kr/gcmdocu/Phy_Cum.htm

Mass-flux parametrizations

$$M_{\mathrm{cb}} = \rho_{\mathrm{air}}\ \omega_{\mathrm{cb}} \times \mathrm{CAF}$$

proper estimation paramount to determine overall strength of convection
Deterministic: assume fixed CAF at 3%
Stochastic: CAF conditioned on large-scale ω_{500}

(Wohltmann, Lehmann, GAG et al, GMD (2019))