Ergodicity in data assimilation methods

David Kelly

Andy Majda Xin Tong

Courant Institute
New York University
New York NY
www.dtbkelly.com

March 16, 2016

Applied math seminar, UC Berkeley / LBNL

What is data assimilation?

Suppose u satisfies

$$\frac{du}{dt} = F(u)$$

with some **unknown** initial condition u_0. We are most interested in geophysical models, so think high dimensional, nonlinear, possibly stochastic.

Suppose we make *partial, noisy* observations at times $t = h, 2h, \ldots, nh, \ldots$

$$y_n = Hu_n + \xi_n$$

where H is a linear operator (think low rank projection), $u_n = u(nh)$, and $\xi_n \sim N(0, \Gamma)$ iid.

The aim of **data assimilation** is to say something about the conditional distribution of u_n given the observations $\{y_1, \ldots, y_n\}$

Illustration (Initialization)

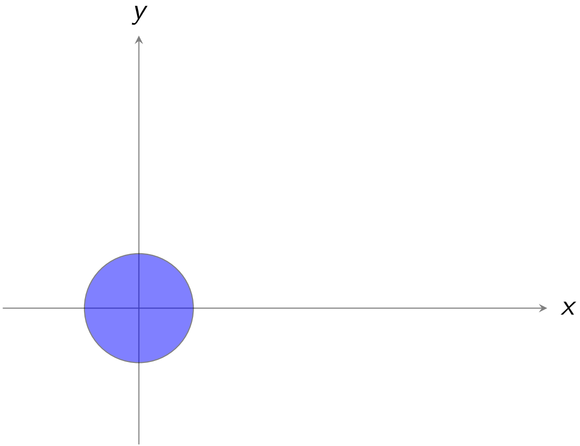


Figure: The blue circle represents our guess of u_0. Due to the uncertainty in u_0, this is a probability measure.

Illustration (Forecast step)

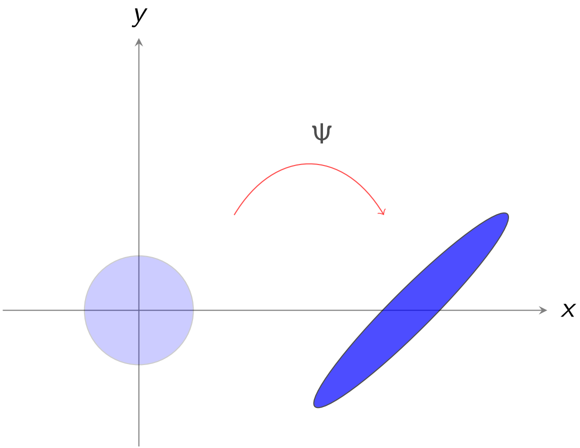


Figure: Apply the time h flow map Ψ. This produces a new probability measure which is our forecasted estimate of u_1. This is called the forecast step.

Illustration (Make an observation)

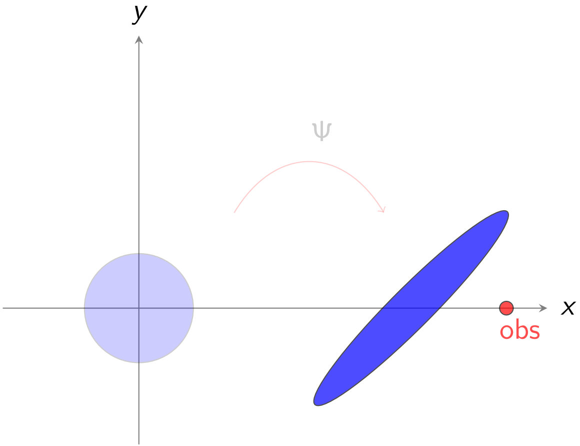


Figure: We make an observation $y_1 = Hu_1 + \xi_1$. In the picture, we only observe the x variable.

Illustration (Analysis step)

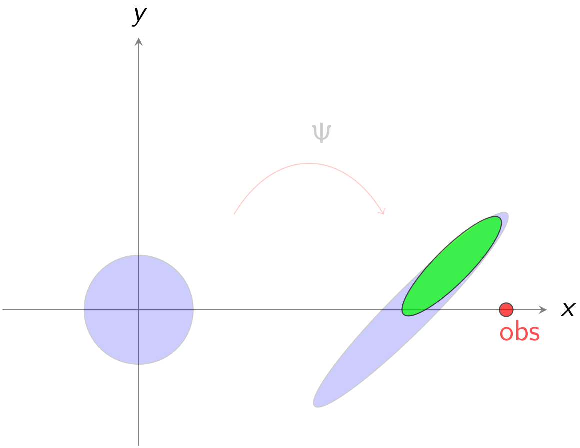


Figure: Using Bayes formula we compute the conditional distribution of $u_1|y_1$. This new measure (called the posterior) is the new estimate of u_1. The uncertainty of the estimate is reduced by incorporating the observation. The forecast distribution steers the update from the observation.

Bayes' formula filtering update

Let $Y_n = \{y_0, y_1, \dots, y_n\}$. We want to compute the conditional density $\mathbf{P}(u_{n+1}|Y_{n+1})$, using $\mathbf{P}(u_n|Y_n)$ and y_{n+1}.

By Bayes' formula, we have

$$\mathbf{P}(u_{n+1}|Y_{n+1}) = \mathbf{P}(u_{n+1}|Y_n, y_{n+1}) \propto \mathbf{P}(y_{n+1}|u_{n+1})\mathbf{P}(u_{n+1}|Y_n)$$

But we need to compute the integral

$$\mathbf{P}(u_{n+1}|Y_n) = \int \mathbf{P}(u_{n+1}|Y_n, u_n)\mathbf{P}(u_n|Y_n)du_n \; .$$

In geophysical models, we can have $u \in \mathbb{R}^N$ where $N = O(10^8)$. The rigorous Bayesian approach is computationally infeasible.

The Kalman Filter

For a linear model $u_{n+1} = M u_n + \eta_{n+1}$, the Bayesian integral is Gaussian and can be computed explicitly. The conditional density is characterized by its mean and covariance

$$m_{n+1} = (1 - K_{n+1} H)\widehat{m}_n + K_{n+1} H y_{n+1}$$
$$C_{n+1} = (I - K_{n+1} H)\widehat{C}_{n+1} \ ,$$

where

- $(\widehat{m}_{n+1}, \widehat{C}_{n+1})$ is the **forecast** mean and covariance.
- $K_{n+1} = \widehat{C}_{n+1} H^T (\Gamma + H \widehat{C}_{n+1} H^T)^{-1}$ is the **Kalman gain**.

The procedure of updating $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$ is known as the **Kalman filter**.

Ensemble Kalman filter (Evensen 94)

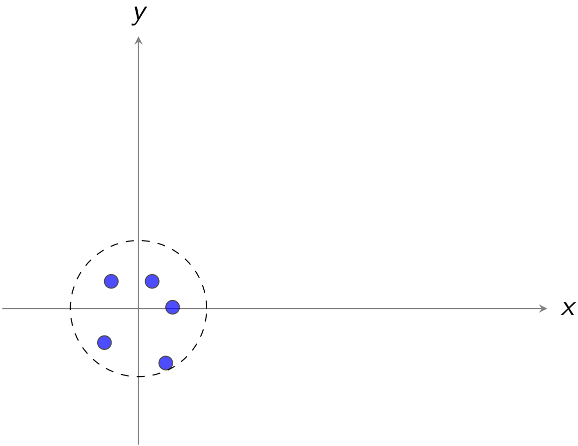


Figure: Start with K ensemble members drawn from some distribution. Empirical representation of u_0. The ensemble members are denoted $u_0^{(k)}$.

Only KN numbers are stored. Better than Kalman if $K < N$.

Ensemble Kalman filter (Forecast step)

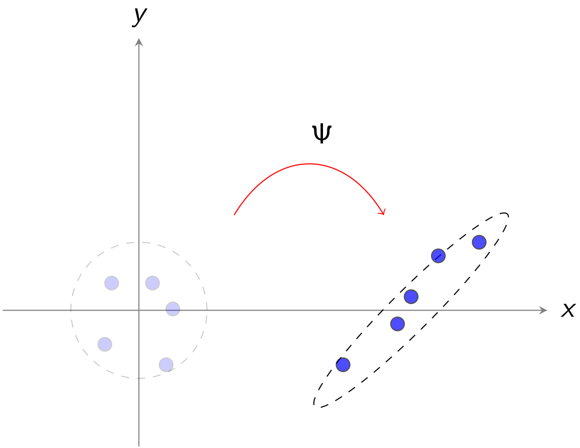


Figure: Apply the dynamics Ψ to each ensemble member.

Ensemble Kalman filter (Make obs)

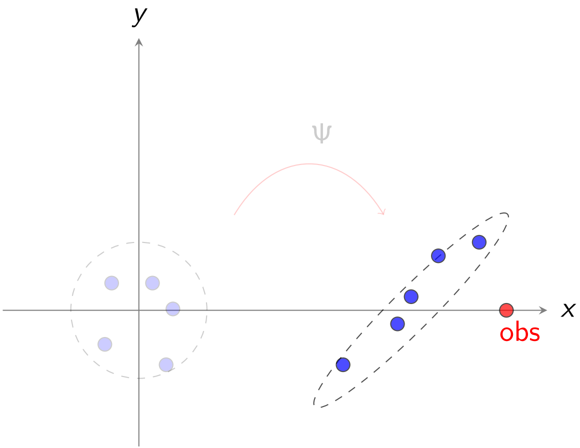


Figure: Make an observation.

Ensemble Kalman filter (Analysis step)

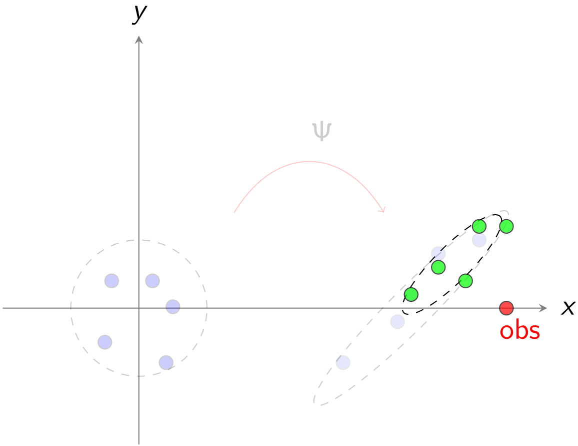


Figure: Approximate the forecast distribution with a Gaussian. Fit the Gaussian using the empirical statistics of the ensemble.

How to implement the Gaussian approximation

The naive method is to simply write:

$$\mathbf{P}(y_1|u_1)\mathbf{P}(u_1) \propto \exp(-\frac{1}{2}|\Gamma^{-1/2}(y_1 - Hu_1)|^2)\exp(-\frac{1}{2}|\widehat{C}^{-1/2}(u_1 - \widehat{m}_1)|^2)$$

with the empirical statistics

$$\widehat{m}_1 = \frac{1}{K}\sum_{k=1}^{K}\Psi^{(k)}(u_0^{(k)})$$

$$\widehat{C}_1 = \frac{1}{K-1}\sum_{k=1}^{K}\left(\Psi^{(k)}(u_0^{(k)}) - \widehat{m}_1\right)\left(\Psi^{(k)}(u_0^{(k)}) - \widehat{m}_1\right)^T .$$

In the linear model case $\Psi(u_n) = Mu_n + \eta_n$, this produces an unbiased estimate of the posterior mean, but a biased estimate of the covariance.

How to implement the Gaussian approximation

A better approach is to sample using **Randomized Maximum Likelihood** (RML) method: Draw the sample $u_1^{(k)}$ by minimizing the functional

$$\frac{1}{2}|\Gamma^{-1/2}(y_1^{(k)} - Hu)|^2 + \frac{1}{2}|\widehat{C}_1^{-1/2}(u - \Psi(u_0^{(k)}))|^2$$

where $y_1^{(k)} = y_1 + \xi_1^{(k)}$ is a perturbed observation.

In the linear case $\Psi(u_n) = Mu_n + \eta_n$, this produces iid Gaussian samples with mean and covariance satisfying the Kalman update equations, with $\widehat{C}$ in place of the true forecast covariance.

We end up with

$$u_1^{(k)} = (1 - K_1 H)\Psi(u_0^{(k)}) + K_1 H y_1^{(k)}$$

Ensemble Kalman filter (Perturb obs)

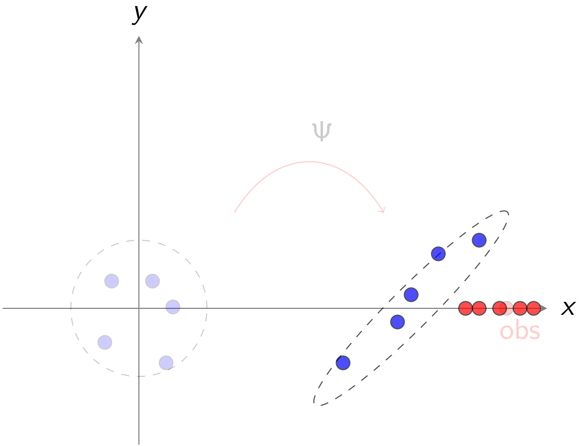


Figure: Turn the observation into K artificial observations by perturbing by the same source of observational noise.

$$y_1^{(k)} = y_1 + \xi_1^{(k)}$$

Ensemble Kalman filter (Analysis step)

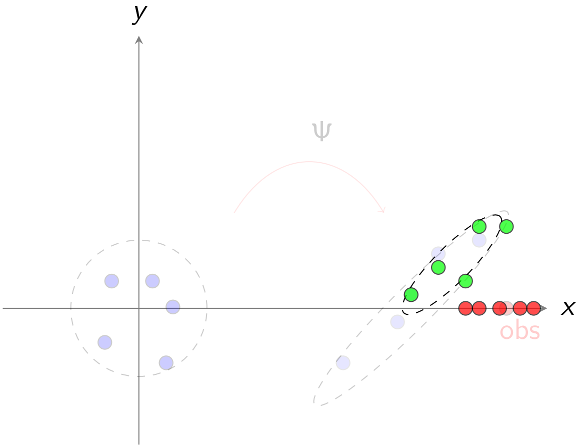


Figure: Update each member using the Kalman update formula. The Kalman gain K_1 is computed using the ensemble covariance.

$$u_1^{(k)} = (1 - K_1 H)\Psi(u_0^{(k)}) + K_1 H y_1^{(k)} \quad K_1 = \widehat{C}_1 H^T (\Gamma + H \widehat{C}_1 H^T)^{-1}$$

$$\widehat{C}_1 = \frac{1}{K-1} \sum_{k=1}^{K} (\Psi(u_0^{(k)}) - \widehat{m}_{n+1})(\Psi(u_0^{(k)}) - \widehat{m}_{n+1})^T$$

Stability / ergodicity of filters

We ask whether the filter **inherits** important physical properties from the underlying model. For instance, if the model is known to be **ergodic**, can the same be said of the filter?

The *truth-filter* process $(u_n, u_n^{(1)}, \dots, u_n^{(K)})$ is a homogeneous Markov chain. We will seek ergodicity results for the pair rather than the filter alone.

The theoretical framework

A Markov chain $\{X_n\}_{n\in\mathbb{N}}$ on a state space $\mathcal{X}$ is called **geometrically ergodic** if it has a unique invariant measure π and for any initialization X_0 we have

$$\left|\mathbf{E}_{X_0} f(X_n) - \int f(x)\pi(dx)\right| \leq C(X_0) r^n$$

for some $r \in (0,1)$ and any m-ble bdd f.

The Meyn-Tweedie approach is to verify two assumptions that guarantee geometric ergodicity:

1- Lyapunov function / Energy dissipation: $\mathbf{E}_n |X_{n+1}|^2 \leq \alpha |X_n|^2 + \beta$ with $\alpha \in (0,1)$, $\beta > 0$.

2- Minorization: Find compact $C \subset \mathcal{X}$, measure ν supported on C, $\kappa > 0$ such that $P(x, A) \geq \kappa \nu(A)$ for all $x \in C$, $A \subset \mathcal{X}$.

Inheriting an energy principle

Suppose we know the model satisfies an energy principle

$$\mathbf{E}_n|\Psi(u)|^2 \leq \alpha|u|^2 + \beta$$

for $\alpha \in (0,1), \beta > 0$. Does the filter inherit the energy principle?

$$\mathbf{E}_n|u_{n+1}^{(k)}|^2 \leq \alpha'|u_n^{(k)}|^2 + \beta'$$

Observable energy (Tong, Majda, K. 15)

We have

$$u_{n+1}^{(k)} = (I - K_{n+1}H)\Psi(u_n^{(k)}) + K_{n+1}Hy_{n+1}^{(k)}$$

Start by looking at the observed part:

$$Hu_{n+1}^{(k)} = (H - HK_{n+1}H)\Psi(u_n^{(k)}) + HK_{n+1}Hy_{n+1}^{(k)} \ .$$

But notice that

$$\begin{aligned}(H - HK_{n+1}H) &= (H - H\widehat{C}_{n+1}H^T(I + H\widehat{C}_{n+1}H^T)^{-1}H) \\ &= (I + H\widehat{C}_{n+1}H^T)^{-1}H\end{aligned}$$

Hence

$$|(H - HK_{n+1}H)\Psi(u_n^{(k)})| \leq |H\Psi(u_n^{(k)})|$$

Observable energy (Tong, Majda, K. 15)

We have the energy estimate

$$\mathbf{E}_n |H u_{n+1}^{(k)}|^2 \leq (1+\delta)|H\Psi(u_n^{(k)})|^2 + \beta'$$

for arb small δ. Unfortunately, the same trick doesn't work for the unobserved variables ... However, if we assume an observable energy criterion instead:

$$|H\Psi(u_n^{(k)})|^2 \leq \alpha |H u_n^{(k)}|^2 + \beta \quad (\star)$$

Then we obtain a Lyapunov function for the observed components of the filter:

$$|H u_n^{(k)}|^2 \leq \alpha' |H u_n^{(k)}|^2 + \beta' \; .$$

eg. $(\star)$ is true for linear dynamics if there is no interaction between observed and unobserved variables at infinity.

Can we get around the problem by tweaking the algorithm?

Covariance inflation (Tong, Majda, K. 15)

We modify algorithm by introducing a **covariance inflation** :

$$\widehat{C}_{n+1} \mapsto \widehat{C}_{n+1} + \lambda_{n+1} I$$

where

$$\lambda_{n+1} \propto \Theta_{n+1} \mathbf{1}(\Theta_{n+1} > \Lambda)$$

$$\Theta_{n+1} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} |y_{n+1}^{(k)} - H\Psi(u_n^{(k)})|^2}$$

and Λ is some constant threshold. If the predictions are near the observations, then there is no inflation.

Thm. The modified EnKF inherits an energy principle from the model.

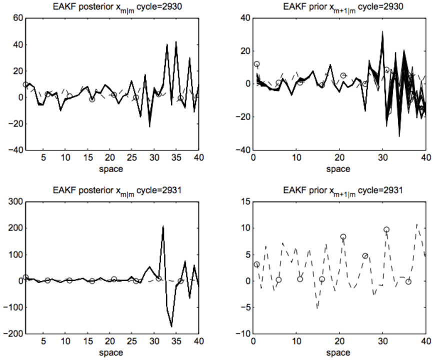
$$|\Psi(x)|^2 \leq \alpha |x|^2 + \beta \Rightarrow \mathbf{E}_n |u_{n+1}^{(k)}|^2 \leq \alpha' |u_n^{(k)}|^2 + \beta'$$

Consequently, the modified EnKF is stable (ergodic).

Stability should not be taken for granted!

Catastrophic filter divergence

Lorenz-96: $\dot{u}_j = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F$ with $j = 1, \dots, 40$. Periodic BCs. Observe every fifth node. *(Harlim-Majda 10, Gottwald-Majda 12)*



True solution in a bounded set, but filter **blows up** to machine infinity in finite time!

For complicated models, only heuristic arguments offered as explanation.

Can we **prove** *it for a simpler constructive model?*

The rotate-and-lock map (K., Majda, Tong. PNAS 15.)

The model $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ is a composition of two maps $\Psi(x, y) = \Psi_{lock}(\Psi_{rot}(x, y))$ where

$$\Psi_{rot}(x, y) = \begin{pmatrix} \rho\cos\theta & -\rho\sin\theta \\ \rho\sin\theta & \rho\cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and Ψ_{lock} rounds the input to the nearest point in the grid

$$\mathcal{G} = \{(m, (2n+1)\varepsilon) \in \mathbb{R}^2 : m, n \in \mathbb{Z}\} \ .$$

It is easy to show that this model has an **energy dissipation principle**:

$$|\Psi(x, y)|^2 \leq \alpha |(x, y)|^2 + \beta$$

for $\alpha \in (0, 1)$ and $\beta > 0$.

(a)

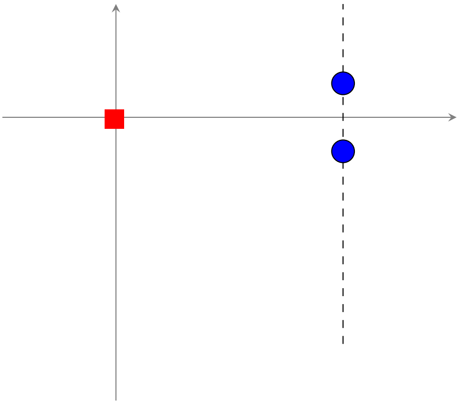


Figure: The red square is the trajectory $u_n = 0$. The blue dots are the positions of the forecast ensemble $\Psi(u_0^+)$, $\Psi(u_0^-)$. Given the locking mechanism in Ψ, this is a natural configuration.

(b)

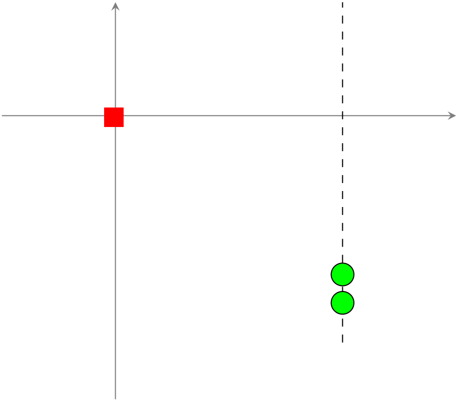


Figure: We make an observation (H shown below) and perform the analysis step. The green dots are u_1^+, u_1^-.

$$H = \begin{pmatrix} 1 & 0 \\ \varepsilon^{-2} & 1 \end{pmatrix} \qquad y_1 = (\xi_{1,x}, \xi_{1,y} + \varepsilon^{-2}\xi_{1,x})$$

$$u_1^\pm \approx (\hat{x}, \pm\varepsilon - 2\hat{x}/(1 + 2\varepsilon^2))$$

(c)

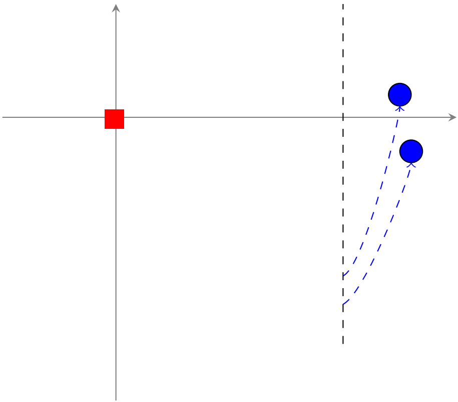


Figure: Beginning the next assimilation step. Apply Ψ_{rot} to the ensemble (blue dots)

(d)

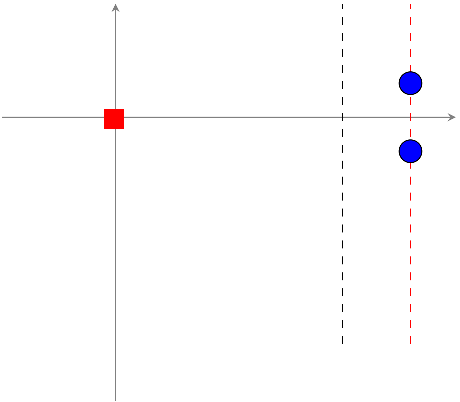


Figure: Apply Ψ_{lock}. The blue dots are the forecast ensemble $\Psi(u_1^+)$, $\Psi(u_1^-)$. Exact same as frame 1, but higher energy orbit. The cycle repeats leading to **exponential growth**.

Theorem (K.-Majda-Tong 15 PNAS)

For any $N > 0$ and any $p \in (0,1)$ there exists a choice of parameters such that

$$\mathbf{P}\left(|u_n^{(k)}| \geq M_n \ \ \text{for all } n \leq N\right) \geq 1 - p$$

where M_n is an exponentially growing sequence.

ie - The filter can be made to grow exponentially for an arbitrarily long time with an arbitrarily high probability.

Next: Conditional ergodicity

The above notion of ergodicity tells us that the filter is behaving in a statistical sense like a real **physical** model.

Another useful notion of ergodicity concerns the long-time behaviour of the measure $\mathbf{P}(u_n|Y_n)$ for a **fixed** sequence of observations Y_n.

If we initialize two filters differently, forecast with independent models, but feed in the same observations, do the filters converge to each other?

Use ideas from ergodicity for *Markov chains in random environments* (Ongoing project w/ J. Mattingly, A. Stuart.)

References

1 - D. Kelly, K. Law & A. Stuart. *Well-Posedness And Accuracy Of The Ensemble Kalman Filter In Discrete And Continuous Time.* **Nonlinearity** (2014).

2 - D. Kelly, A. Majda & X. Tong. *Concrete ensemble Kalman filters with rigorous catastrophic filter divergence*. **Proc. Nat. Acad. Sci.** (2015).

3 - X. Tong, A. Majda & D. Kelly. *Nonlinear stability and ergodicity of ensemble based Kalman filters*. **Nonlinearity** (2016).

4 - X. Tong, A. Majda & D. Kelly. *Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation*. To appear in **Comm. Math. Sci.** (2016).

All my slides are on my website (www.dtbkelly.com) **Thank you**!