Nowcasting with Indistinguishable States

Hailiang Du, Kevin Judd and Leonard A. Smith
Centre for the Analysis of Time Series
London School of Economics

h.l.du@lse.ac.uk

Outline

- Uncertainty can be quantified by ensemble
- Perfect model scenario (PMS)
  - Indistinguishable States (IS)
  - Nowcast via IS in PMS
  - (Compare with) Alternative approaches
- Imperfect model scenario (IPMS)
  - Nowcast via IS in IPMS
  - (Compare with) Alternative approaches
- Conclusion & Further discussion

Nowcast by ensemble

- In order to forecast the future evolution of a dynamical system using a model, we have to initialise the model.
- It is impossible to determine the state of the system precisely, even given a perfect model and noisy observations.
- To maintain forecast uncertainty in the initial condition, we need to launch our model with an ensemble instead of one point.
- In the same way, to maintain uncertainty in the nowcasting requires an ensemble.

Experiment Design (PMS)

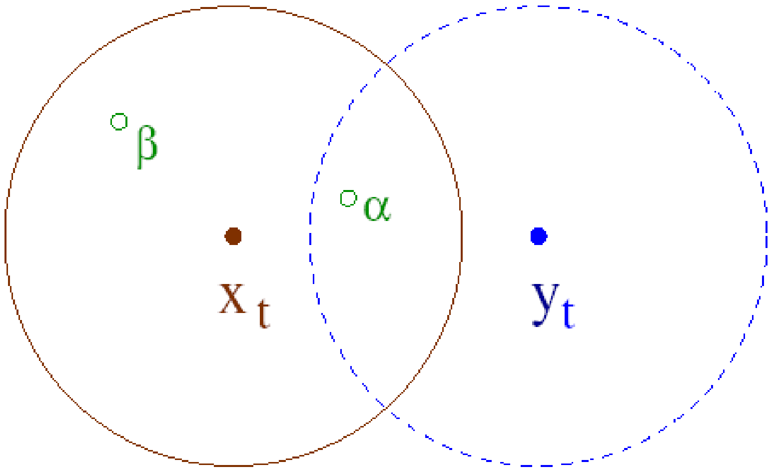
- Let $x_n \in R^d$, be the trajectory of a finite dimensional, deterministic nonlinear dynamical system:
$$x_{t+1} = F(x_t), F : R^d \to R^d.$$

- Observations: $s_t = x_t + \epsilon_t$ where ϵ is IID. In this case s_t can be thought of as in the model state space.

- Perfect model scenario: F is known, so as the noise model.

Goal: To form an ensemble near x_0, consistent with the model dynamics, given the history of previous (and current) observations $s_i, i = -n, ..., 0$.

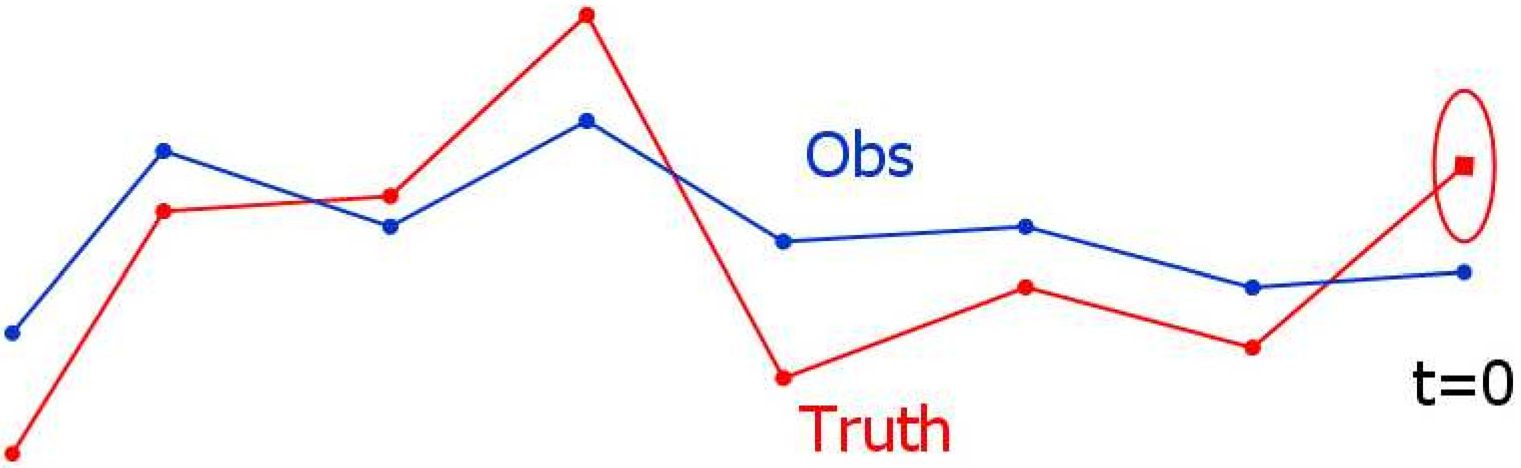
Introduction of Indistinguishable States



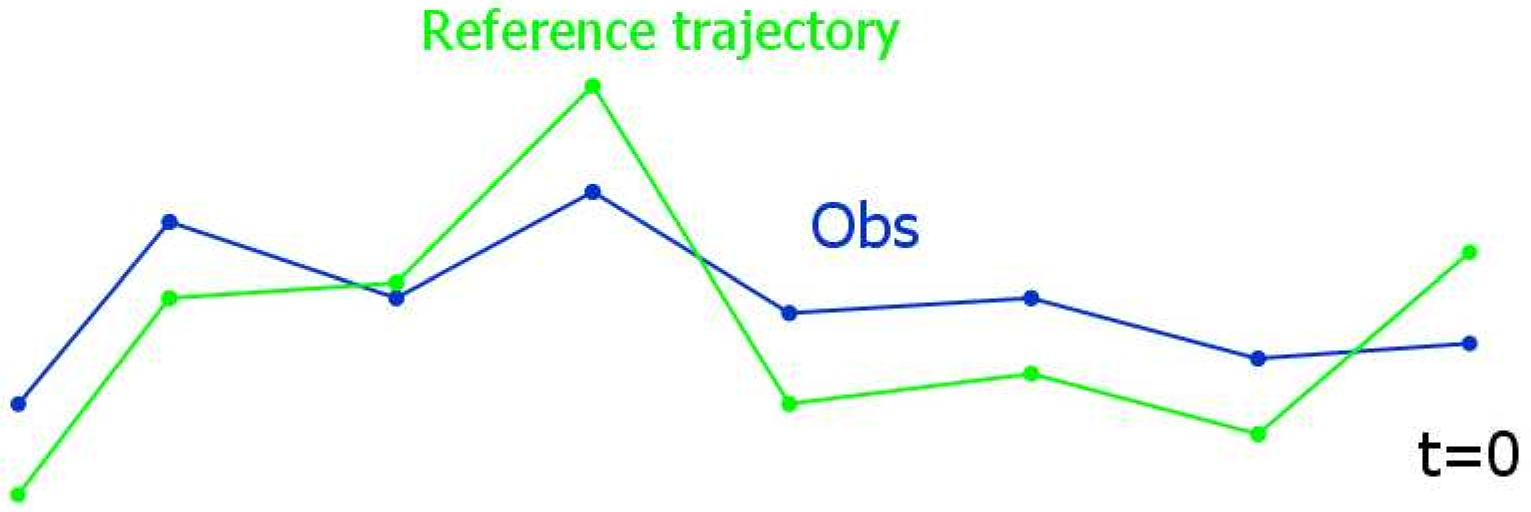
Given a noise model, the probability of $X = x_t, t = 1, ..., n$ and $Y = y_t, t = 1, ..., n$ being indistinguishable is $Q(Y \mid X)$.

K.Judd and L.A.Smith. Indistinguishable states I: perfect model scenario. Physica D, 151:125-141, 2001.

Methodology



Methodology



How to find a reference trajectory?

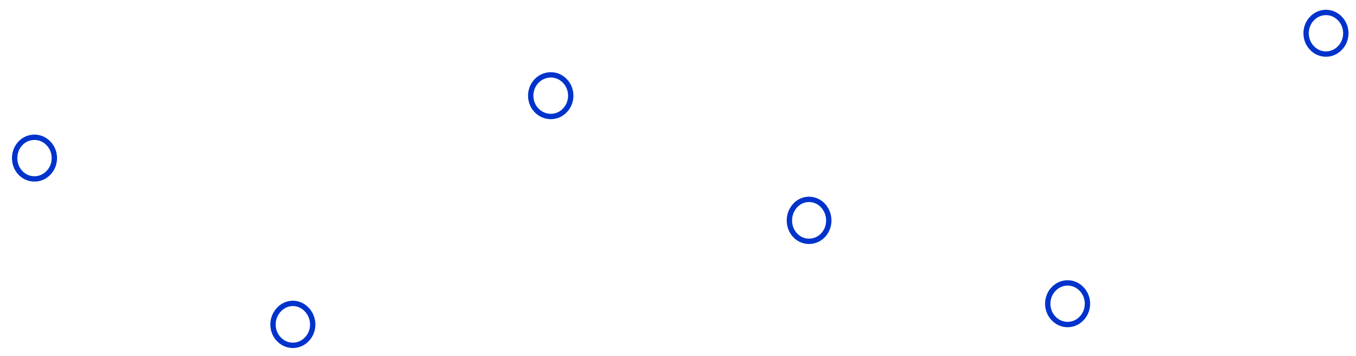
Finding reference trajectory

Given a sequence of N observations of M dimension system, we define a sequence space a $M \times N$ dimensional space, which contains any series of N model states.

Define the mismatch error:

$$\delta = \mid f(x_i) - x_{i+1} \mid$$

Applying a Gradient Descent algorithm, starting at the observations and evolving so as to minimise the sum of the squared δ.



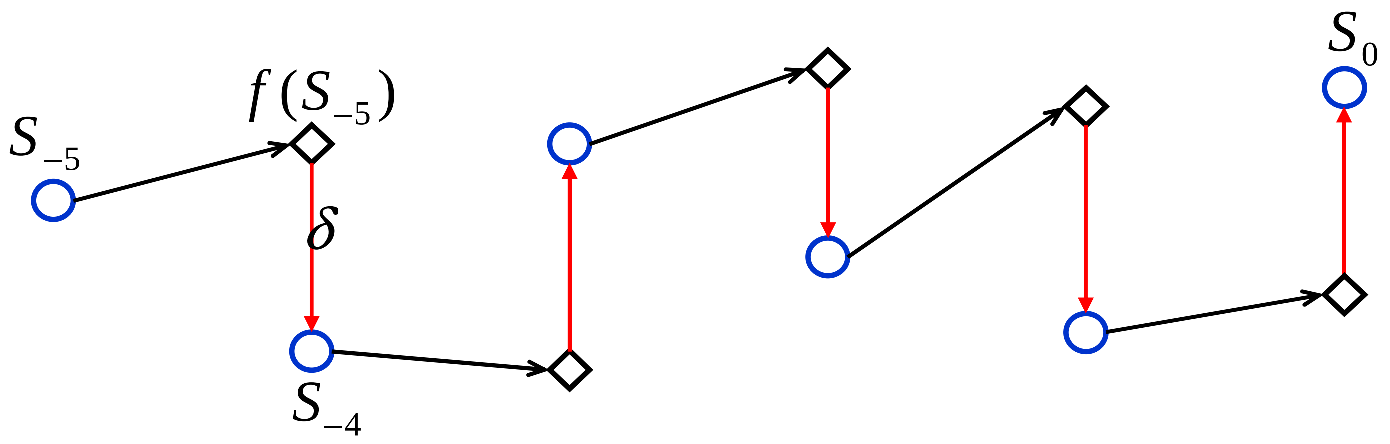
Finding reference trajectory

Given a sequence of N observations of M dimension system, we define a sequence space a $M \times N$ dimensional space, which contains any series of N model states.

Define the mismatch error:

$$\delta = | f(x_i) - x_{i+1} |$$

Applying a Gradient Descent algorithm, starting at the observations and evolving so as to minimise the sum of the squared δ.



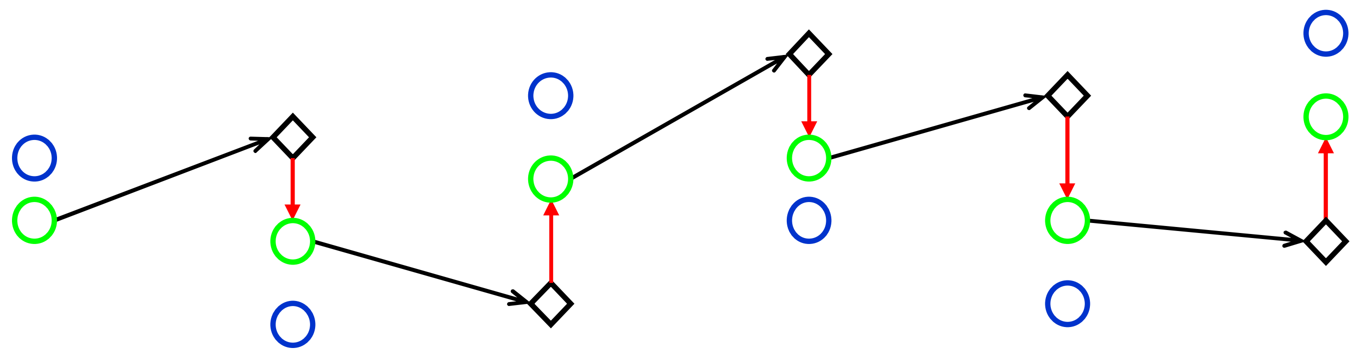
Finding reference trajectory

Given a sequence of N observations of M dimension system, we define a sequence space a $M \times N$ dimensional space, which contains any series of N model states.

Define the mismatch error:

$$\delta = \mid f(x_i) - x_{i+1} \mid$$

Applying a Gradient Descent algorithm, starting at the observations and evolving so as to minimise the sum of the squared δ.



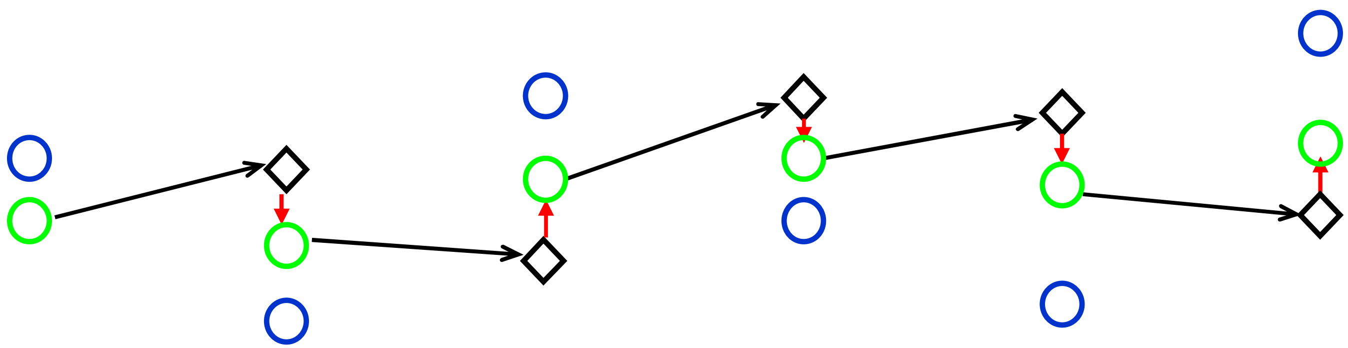
Finding reference trajectory

Given a sequence of N observations of M dimension system, we define a sequence space a $M \times N$ dimensional space, which contains any series of N model states.

Define the mismatch error:

$$\delta = | f(x_i) - x_{i+1} |$$

Applying a Gradient Descent algorithm, starting at the observations and evolving so as to minimise the sum of the squared δ.



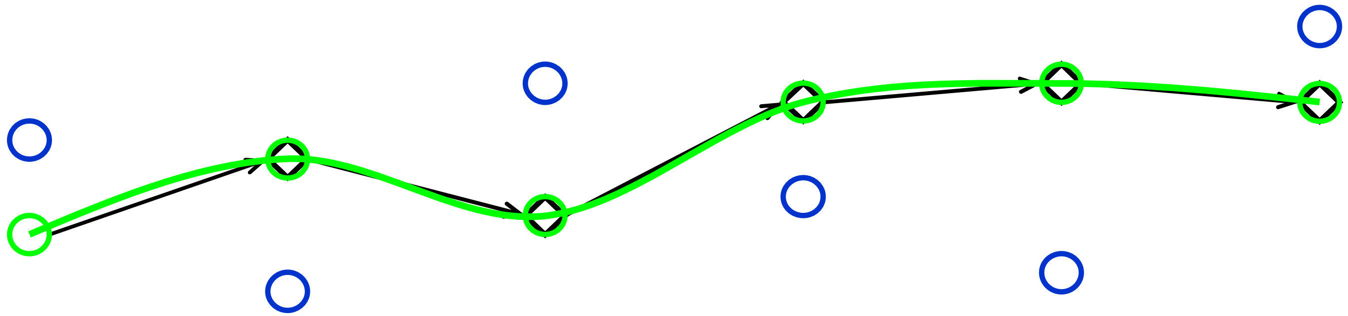
Finding reference trajectory

Given a sequence of N observations of M dimension system, we define a sequence space a $M \times N$ dimensional space, which contains any series of N model states.

Define the mismatch error:

$$\delta = \mid f(x_i) - x_{i+1} \mid$$

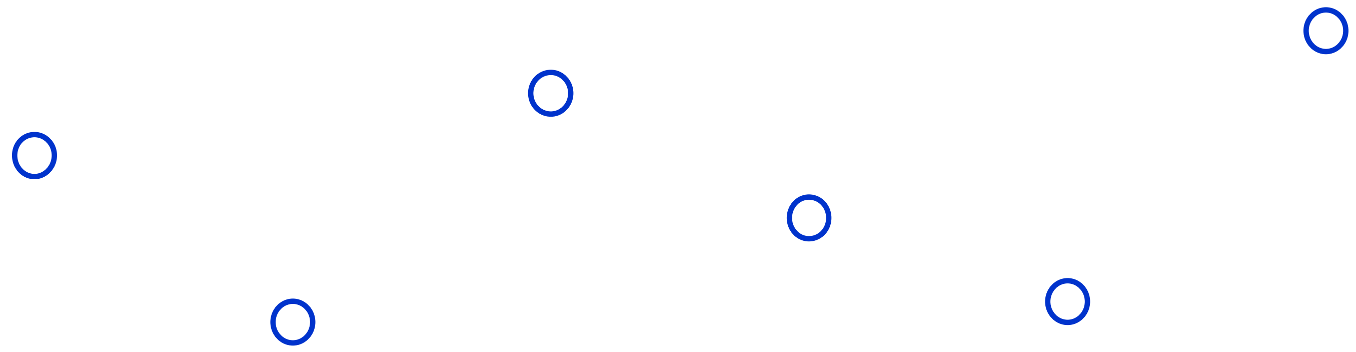
Applying a Gradient Descent algorithm, starting at the observations and evolving so as to minimise the sum of the squared δ.



4DVAR

Let $x_t = F(x_{t-1})$, the 4DVAR cost function is:

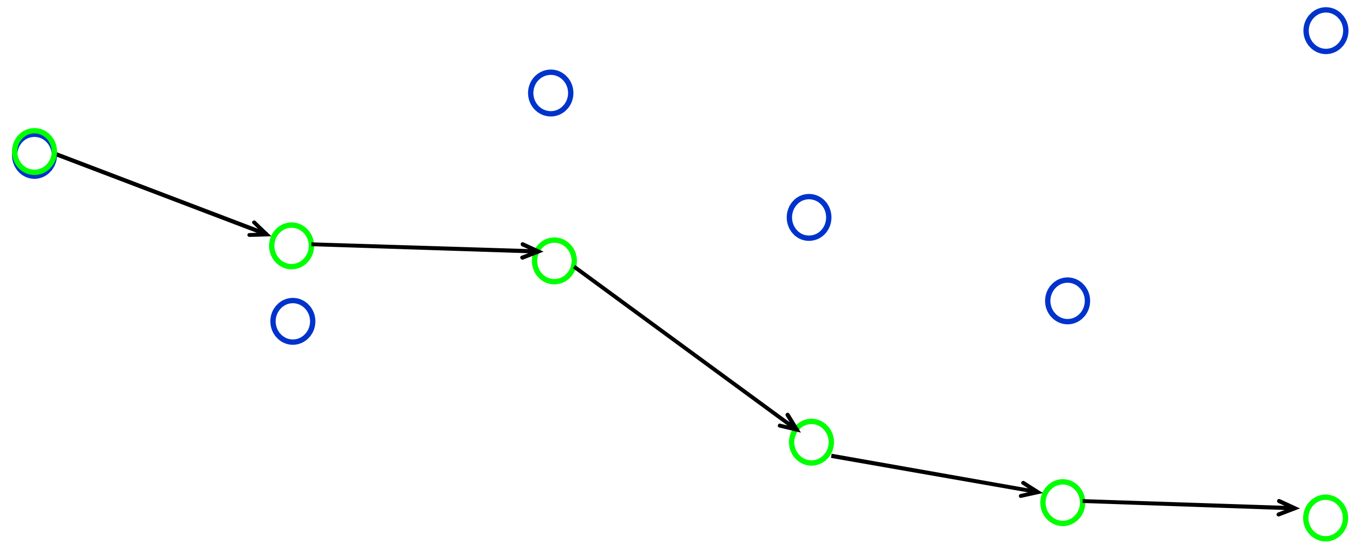
$$C_{4d\,\mathrm{var}}(x) = \frac{1}{2}(x_{-n} - x_{-n}^b)^T B_{-n}^{-1}(x_{-n} - x_{-n}^b) + \frac{1}{2}\sum_{t=-n}^{0}(x_t - s_t)^T \Gamma^{-1}(x_t - s_t)$$



4DVAR

Let $x_t = F(x_{t-1})$, the 4DVAR cost function is:

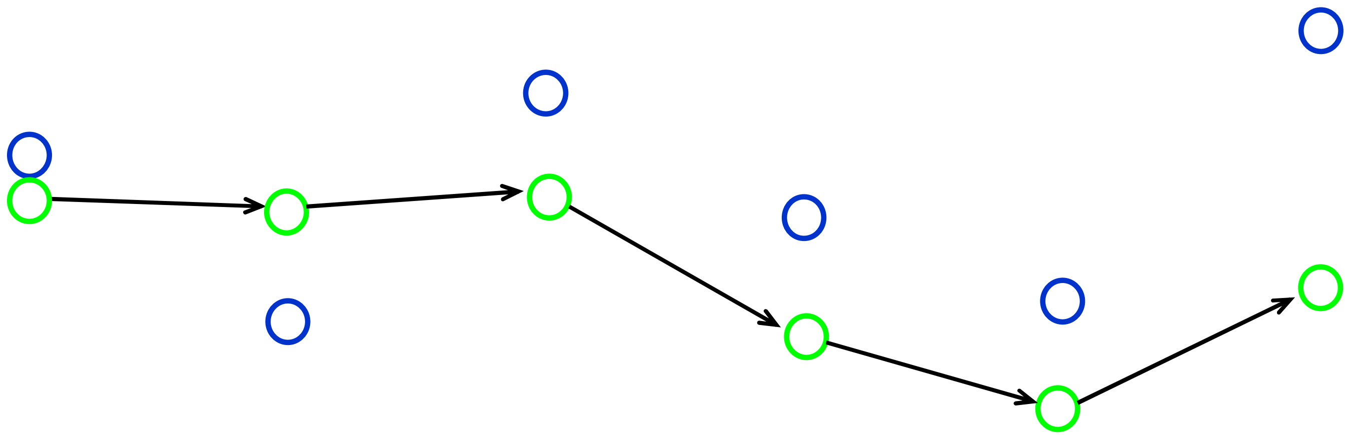
$$C_{4d\,\mathrm{var}}(x) = \frac{1}{2}(x_{-n} - x_{-n}^b)^T B_{-n}^{-1}(x_{-n} - x_{-n}^b) + \frac{1}{2}\sum_{t=-n}^{0}(x_t - s_t)^T \Gamma^{-1}(x_t - s_t)$$



4DVAR

Let $x_t = F(x_{t-1})$, the 4DVAR cost function is:

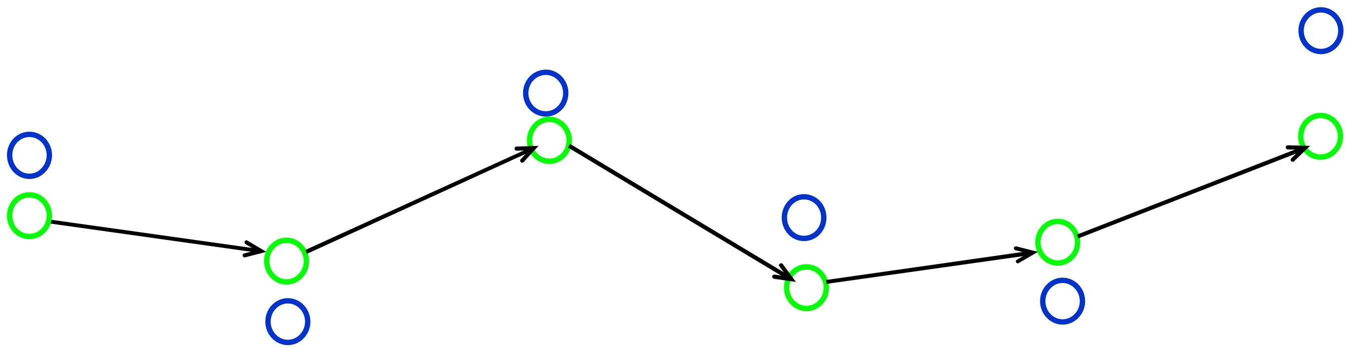
$$C_{4d\,\mathrm{var}}(x) = \frac{1}{2}(x_{-n} - x_{-n}^{b})^T B_{-n}^{-1}(x_{-n} - x_{-n}^{b}) + \frac{1}{2}\sum_{t=-n}^{0}(x_t - s_t)^T \Gamma^{-1}(x_t - s_t)$$



4DVAR

Let $x_t = F(x_{t-1})$, the 4DVAR cost function is:

$$C_{4d\,\mathrm{var}}(x) = \frac{1}{2}(x_{-n} - x_{-n}^{b})^{T} B_{-n}^{-1}(x_{-n} - x_{-n}^{b}) + \frac{1}{2}\sum_{t=-n}^{0}(x_t - s_t)^{T}\Gamma^{-1}(x_t - s_t)$$



Local minima in 4DVAR cost function

Gauthier (1992), Stensrud and Bao (1992) and Miller et al. (1994) found that performance of assimilation varies significantly depending on the length of the assimilation window and difficulties arises with the extension of assimilation window due to the occurrence of multiple minima in the cost function.

ISGD vs 4DVAR

Window length	a) Distance from observations					
	Average		Lower		Upper	
	4DVAR	ISGD	4DVAR	ISGD	4DVAR	ISGD
4 steps	1.58	1.66	1.51	1.59	1.63	1.73
6 steps	11.06	1.77	8.17	1.71	14.28	1.83
8 steps	51.84	1.85	46.16	1.80	58.54	1.90
Window length	**b) Distance from truth**					
	Average		Lower		Upper	
	4DVAR	ISGD	4DVAR	ISGD	4DVAR	ISGD
4 steps	0.52	0.61	0.48	0.55	0.55	0.67
6 steps	9.51	0.39	6.70	0.36	12.59	0.42
8 steps	50.04	0.28	43.59	0.25	55.77	0.31

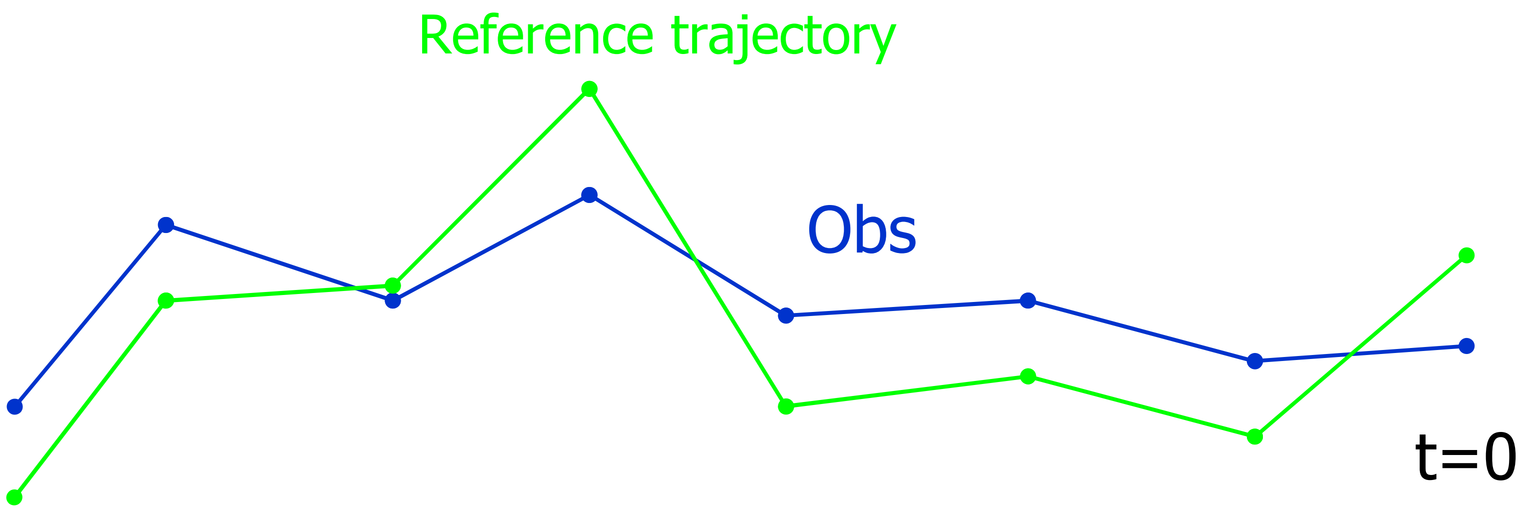
Table 1. a) Distance between the observations and the model trajectory generated by 4DVAR and ISGD for Ikeda experiment, b) Distance between the true states and the model trajectory generated by 4DVAR and ISGD for Ikeda experiment.

Form ensemble via ISIS

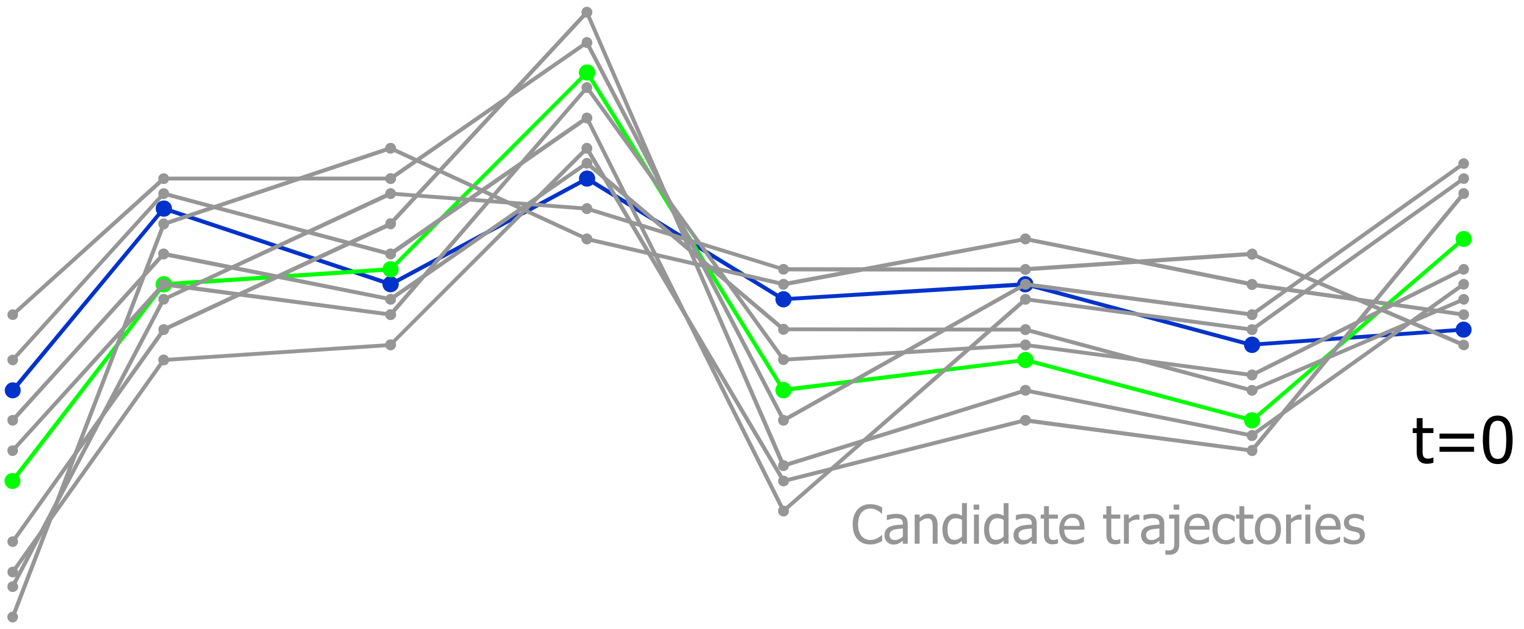
- Given the reference trajectory, there are many ways of finding candidate trajectories.
- Draw ensemble members according to Q(Y|X*)
- Weight ensemble members according to the likelihood of observations

Indistinguishable States Importance Sampler

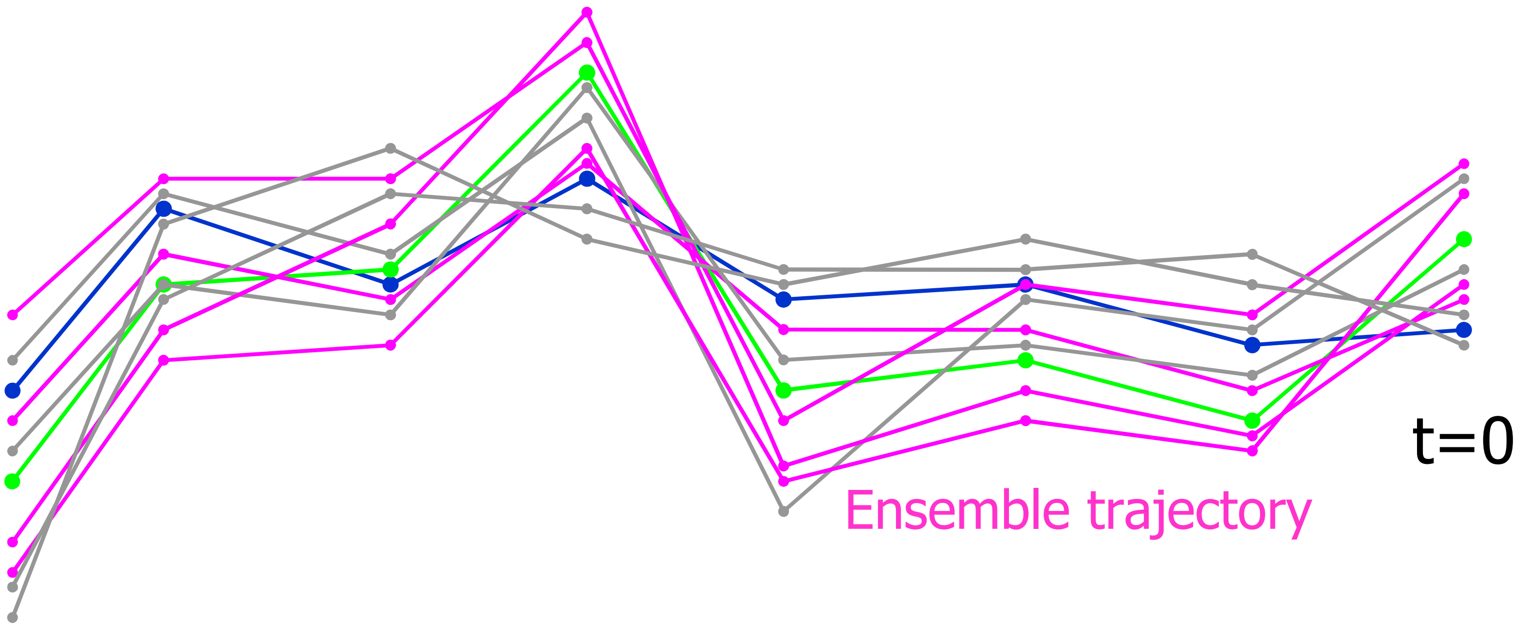
Form ensemble via ISIS



Form ensemble via ISIS

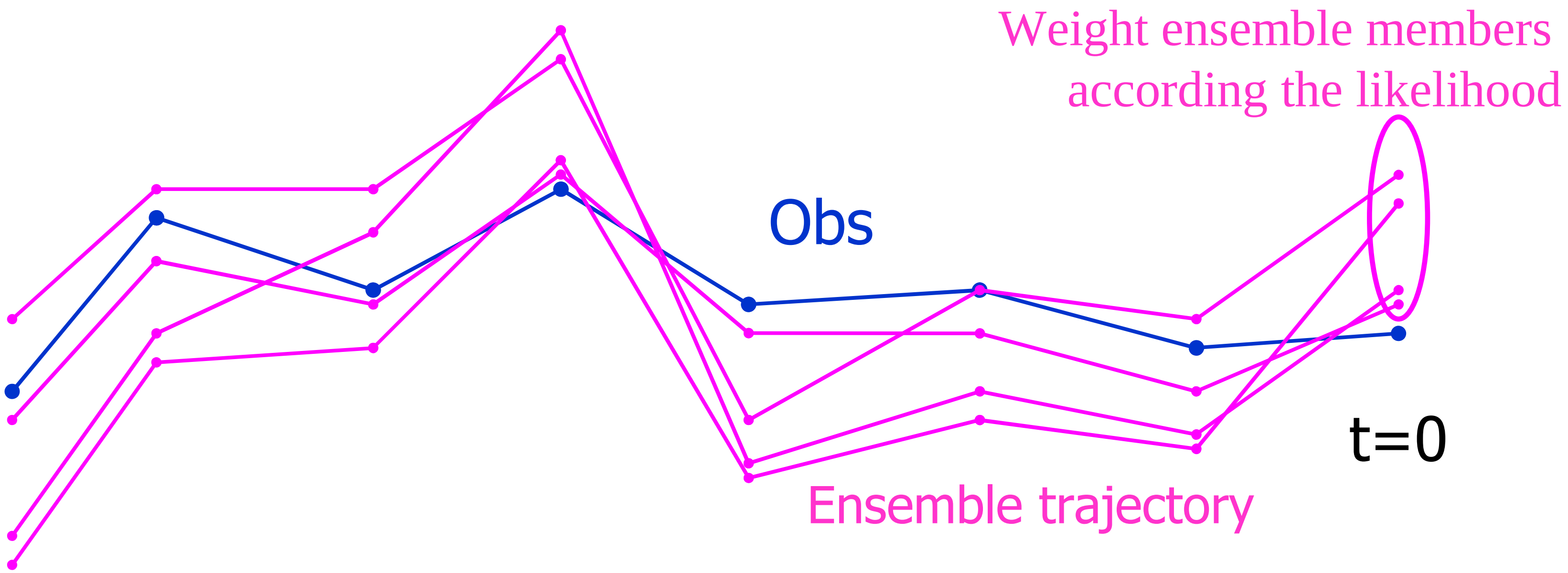


Form ensemble via ISIS



Draw ensemble members according to Q density

Form ensemble via ISIS



Ensemble Kalman Filter Method

Anderson(2001) introduced an ensemble adjustment Kalman filter method by sequentially updating equally weighted ensemble members according to the observations.

$$X_{i,k}^{u} = X_{i,k} + \gamma(\alpha - 1)y_k - \gamma\alpha\bar{y_k} + \gamma\bar{y_k^u} \qquad (1)$$

K is the number of ensemble members. $X_{i,k}$ is the prior ensemble for state variable X_i. As our model states stay at exactly the same space of the observations, $y_k = X_k$. $\gamma = \sigma_{x,y}^2/\sigma_{y,y}^2$ where $\sigma_{x,y}^2$ is the prior covariance between the ensembles y_k and $X_{i,k}$, and $\sigma_{y,y}^2$ is the variance of y_k. $\alpha = \sqrt{\sigma_{obs}^2(\sigma_{obs}^2 + \sigma_{y,y}^2)^{-1}}$, where σ_{obs}^2 is the observational error variance. The updated mean for the observation variable, $\bar{y_k^u} = (1 - \frac{\sigma_{y,y}^2}{\sigma_{obs}^2+\sigma_{y,y}^2})(\bar{y_k} + \sigma_{obs}^{-2}y^o\sigma_{y,y}^2)$.

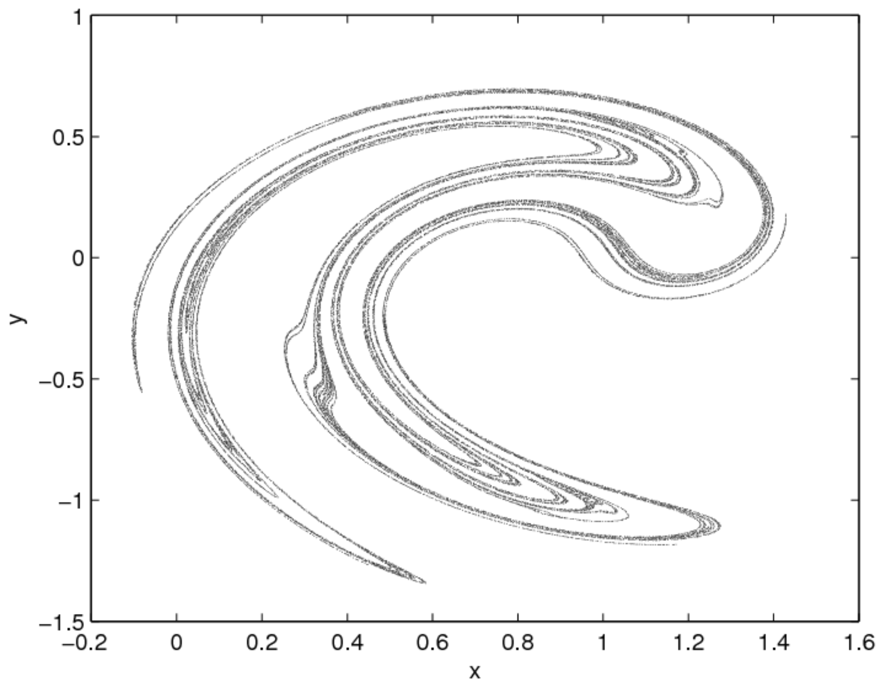
Ensemble Kalman Filter Method

Anderson(2001) introduced an ensemble adjustment Kalman filter method by sequentially updating equally weighted ensemble members according to the observations.

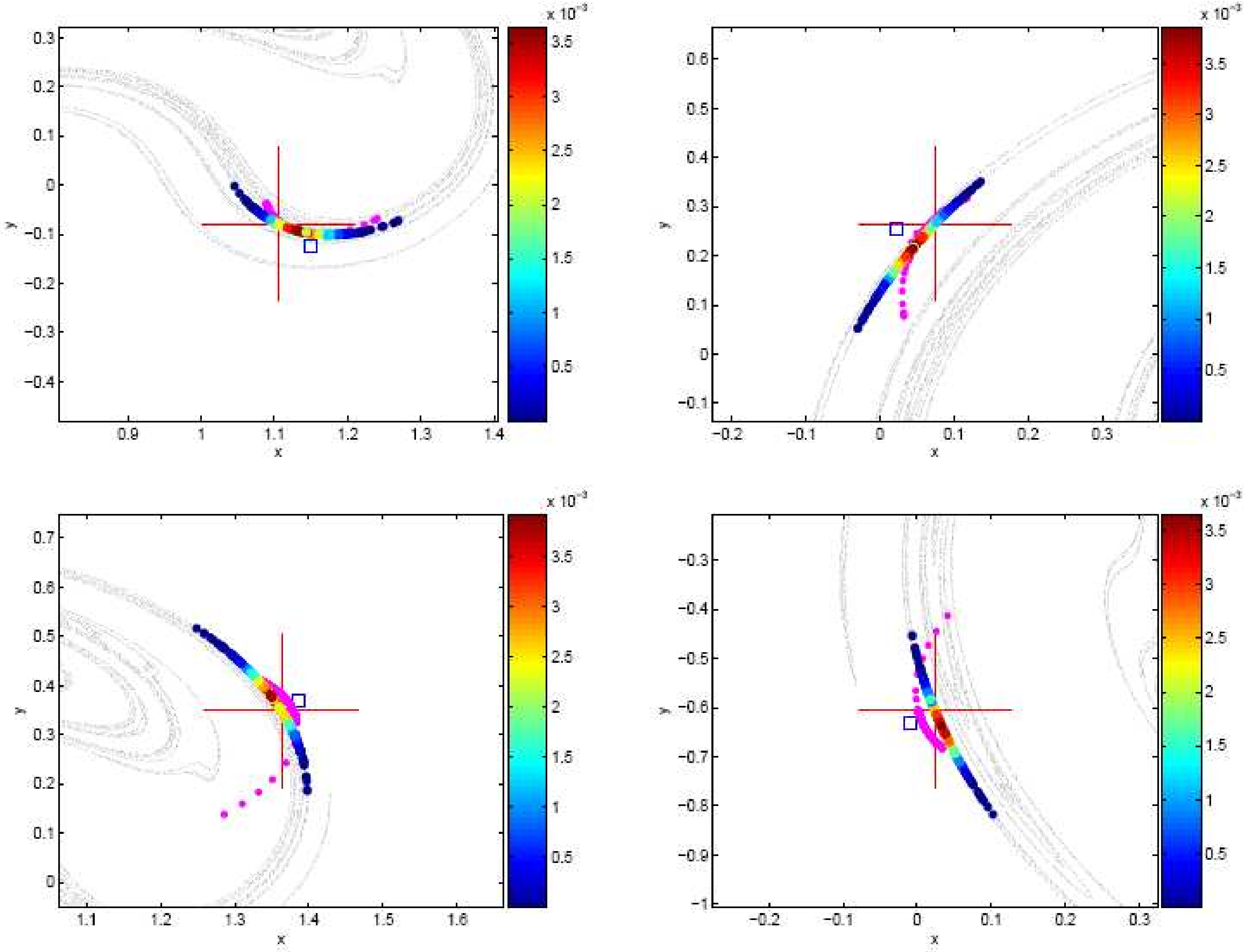
Anderson,J.L.,2001: An ensemble adjustment Kalman filter for data assimilation. Mon Wea Rev, 129, 2884-2903

Ensemble members in the state space

Compare ensemble members generated by Indistinguishable states method and Ensemble Kalman Filter method in the state space.



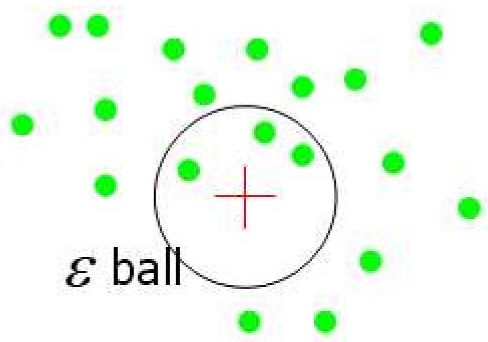
Low dimensional example to visualize, higher dimensional results later.



Ikeda Map, Std of observational noise 0.05, 512 ensemble members

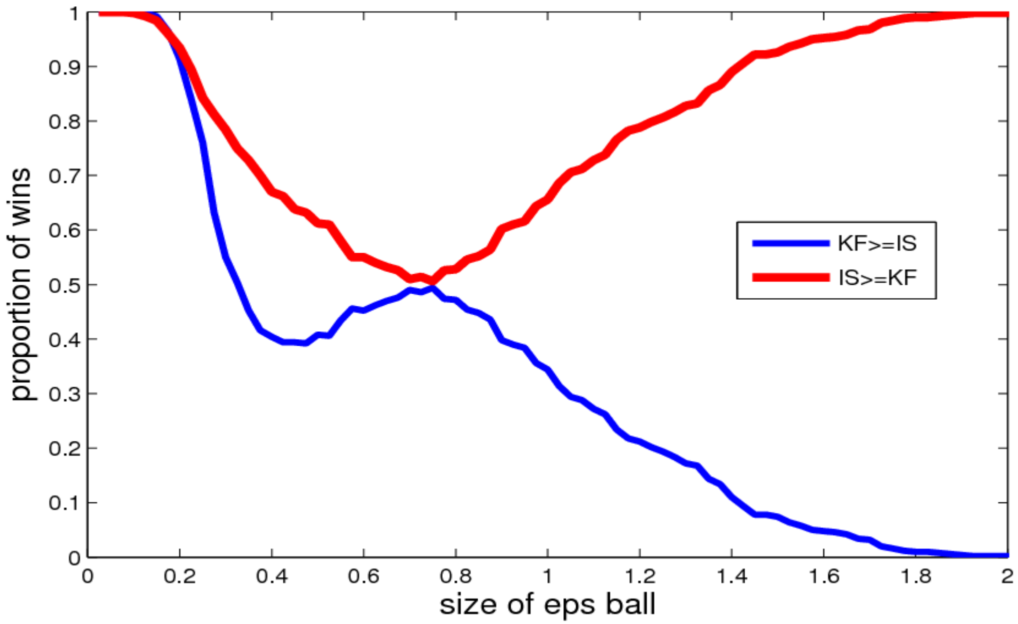
Simple evaluation method

We evaluate these two methods by looking at the probability mass that stay inside different sizes of ϵ ball and counting the proportion of times one methods beats the other (if tie, both win).



Simple evaluation method

We evaluate these two methods by looking at the probability mass that stay inside different sizes of ϵ ball and counting the proportion of times one methods beats the other (if tie, both win).



Lorenz96 12-D, $\sigma_{obs} = 1$, $Nens_{IS} = 64$; $Nens_{KF} = 1024$

Imperfect Model Scenario

- In the IPMS, model state and system state are living in the different state space.
- Let x_t be a projection of system trajectory into model state space R^d.
- The chaotic model has dynamics $y_{t+1} = f(y_t)$, $y_t \in R^d$.
- Let $f(.)$ be the best model we have.
- Observations: $s_t = x_t + \epsilon_t$ where ϵ is IID.
- Define the model error, $\omega_t^* = x_t - f(x_{t-1}), \omega_t^* \in R^d$

Imperfect Model Scenario

- No model trajectories are able to be consistent with the infinite observations.

- There are pseudo-orbits, with non-zero mismatch error, that are consistent with the observations. We define pseudo-orbit $z_t, t = 0, -1, -2, \ldots$

$$z_{i+1} = f(z_i) + \omega_i, \ \omega_i \text{ is not IID}$$

- Confounding of observational noise and model error prevents one identifying either of them.

- Data assimilation can explore the model dynamics by employing pseudo-orbits.

Toy model-system pairs

Ikeda system:

$$x_{n+1} = \gamma + u(x_n \cos\theta - y_n \sin\theta)$$

$$y_{n+1} = u(x_n \sin\theta + y_n \cos\theta),$$

where $\theta = \beta - \alpha/(1 + x_n^2 + y_n^2)$

Imperfect model is obtained by using the truncated polynomial, i.e.

$$\cos\theta = \cos(\omega + \pi) \mapsto -\omega + \omega^3/6 - \omega^5/120$$

$$\sin\theta = \sin(\omega + \pi) \mapsto -1 + \omega^2/2 - \omega^4/24$$

Toy model-system pairs

Lorenz96 system:

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F - \frac{h_x c}{b}\sum_{j=1}^{n} y_{i,j}$$

$$\frac{dy_{j,i}}{dt} = cby_{j+1,i}(y_{j-1,i} - y_{j+2,i}) - cy_{j,i} + -\frac{h_y c}{b}x_i$$

Imperfect model:

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F$$

Insight of Gradient Descent

Given a sequence of N observations of M dimension system, we define a sequence space a $M \times N$ dimensional space, which contains any series of N model states.

Define the mismatch error:

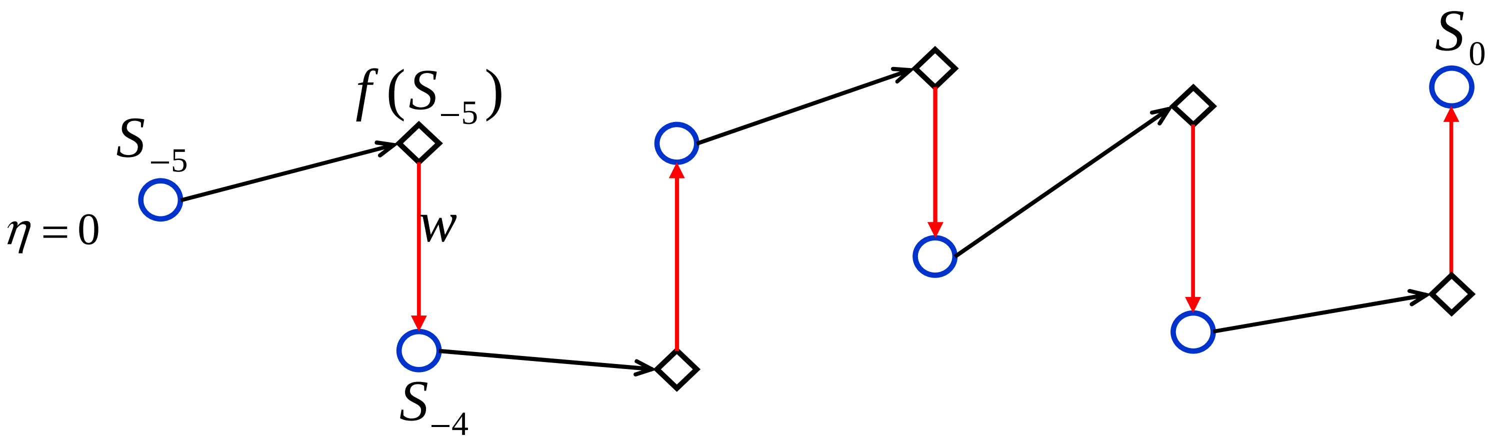
$$\delta = | f(x_i) - x_{i+1} |$$

Applying a Gradient Descent algorithm, starting at the observations and evolving so as to minimise the sum of the squared δ.

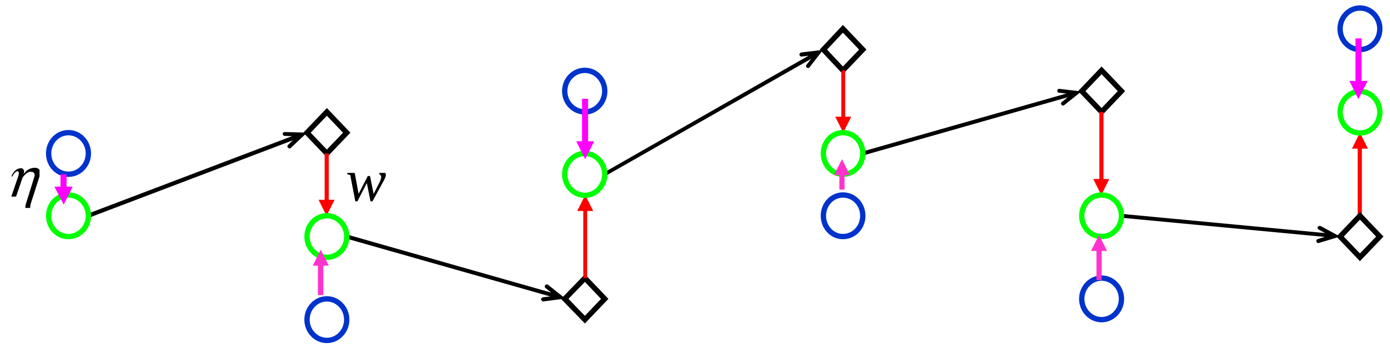
Define the implied noise to be $\eta_i = s_i - z_i$

and the imperfection error to be $\omega_i = z_i - f(z_{i-1})$

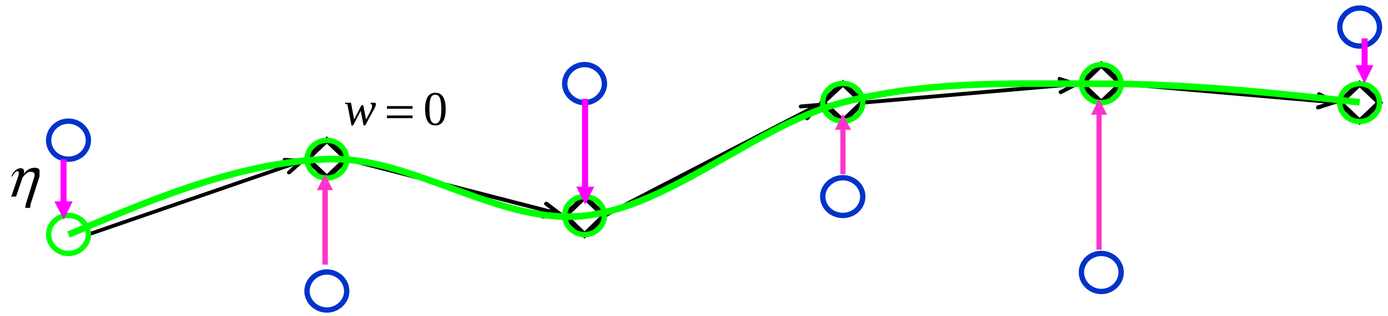
Insight of Gradient Descent



Insight of Gradient Descent



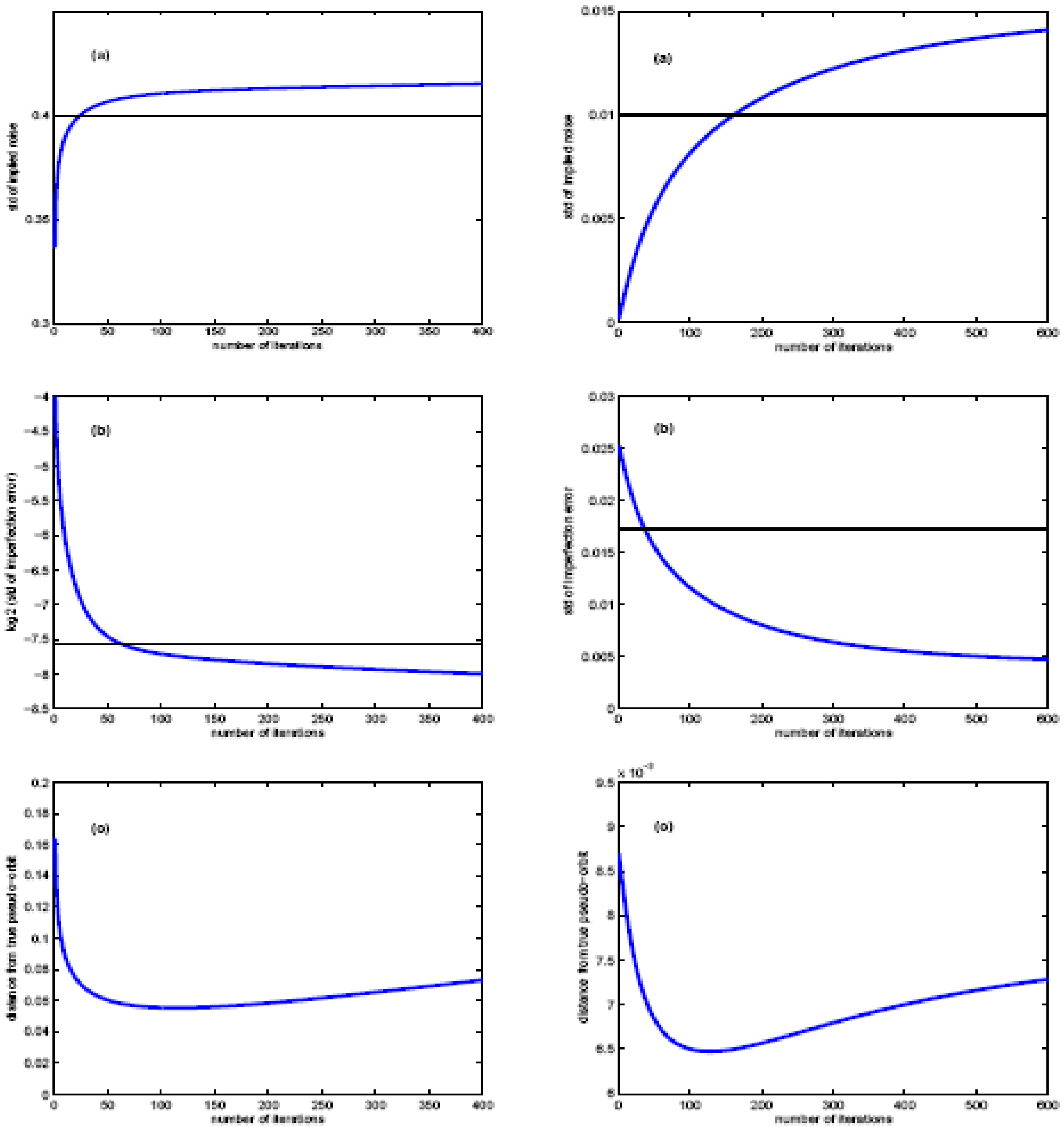
Insight of Gradient Descent



Implied noise

Imperfection error

Distance from the “truth”



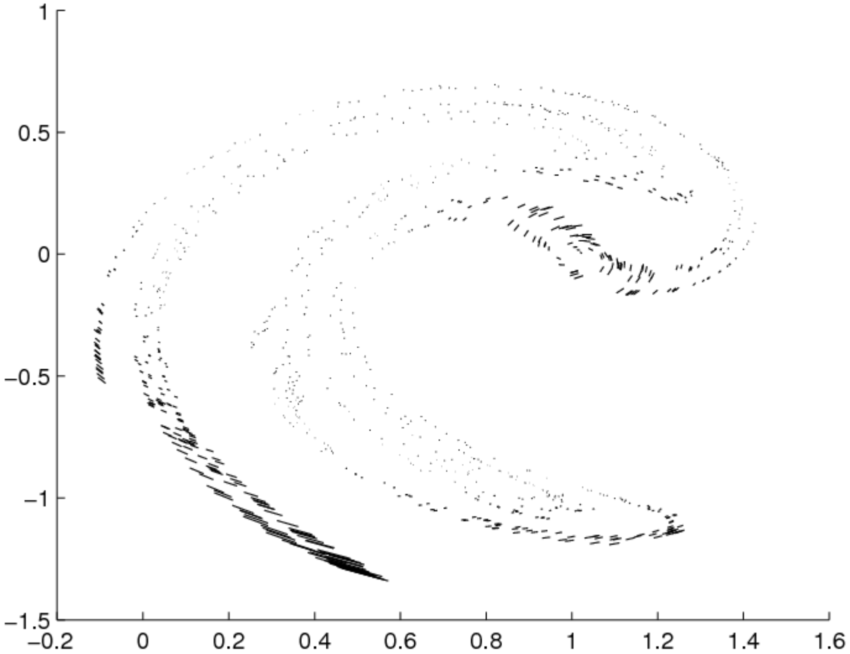
Statistics of the pseudo-orbit as a function of the number of Gradient Descent iterations for both higher dimension Lorenz96 system-model pair experiment (left) and low dimension Ikeda system-model pair experiment (right).

ISGD with stopping criteria

- ISGD minimization with “intermediate” runs produces more consistent pseudo-orbits
- Certain criteria need to be defined in advance to decide when to stop.
- The stopping criteria can be built by testing the consistency between implied noise and the noise model
- or by minimizing some forecast utility function

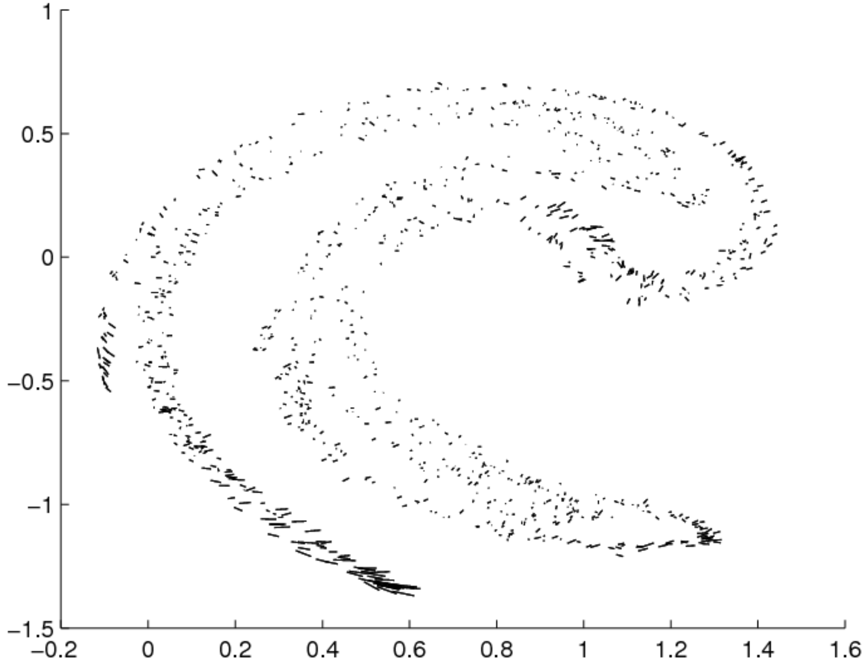
Imperfection error vs model error

Noise level 0.01



Model error

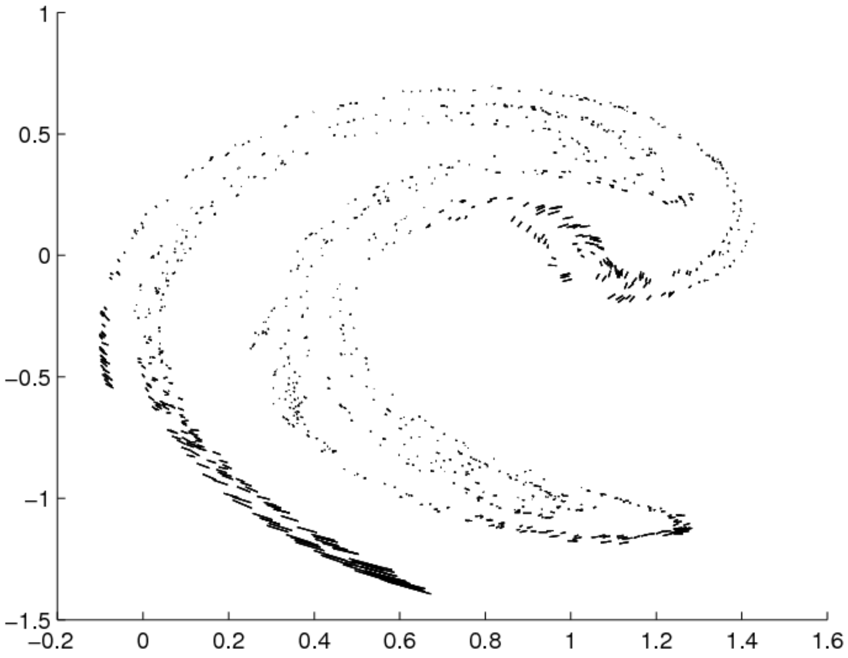
Not accessible!



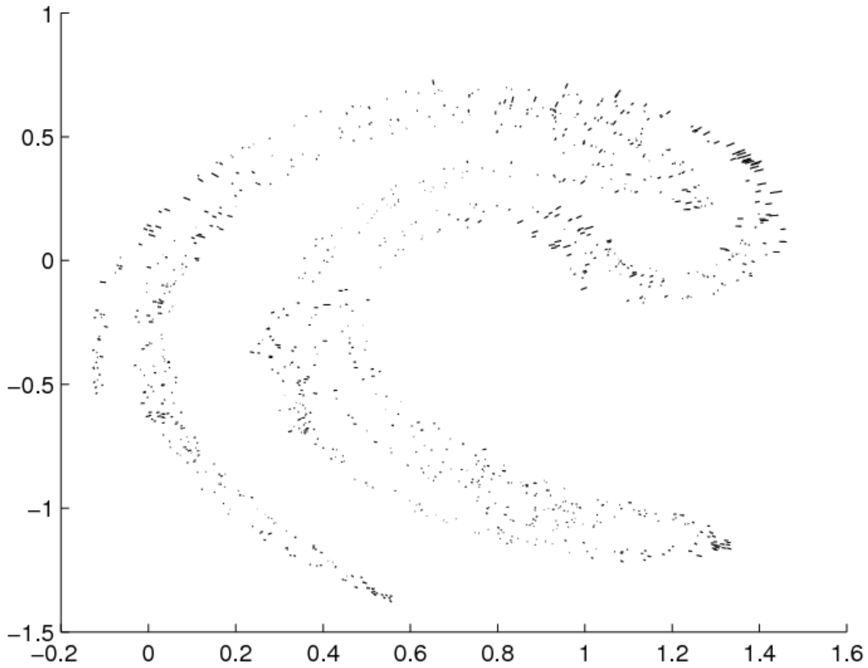
Imperfection error

Imperfection error vs model error

Noise level 0.002



Noise level 0.05



Imperfection error

WC4DVAR

WC4DVAR cost function:

$$C_{wc4d\,\mathrm{var}} = \frac{1}{2}(x_0 - x_0^b)^T B_0^{-1}(x_0 - x_0^b) + \frac{1}{2}\sum_{t=0}^{N}(x_t - s_t)^T \Gamma^{-1}(x_t - s_t)$$

$$+\frac{1}{2}\sum_{t=1}^{N}(x_t - F(x_{t-1}))^T Q^{-1}(x_t - F(x_{t-1}))$$

ISGD vs WC4DVAR

Window length	Distance from observations					
	Average		Lower		Upper	
	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$
6 hours	16.42	14.00	16.24	13.85	16.59	14.14
12 hours	20.60	14.40	20.41	14.30	20.78	14.50
24 hours	81.11	14.52	78.17	14.45	84.17	14.59
Window length	Distance from true states					
	Average		Lower		Upper	
	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$
6 hours	5.87	4.15	5.76	4.08	5.98	4.23
12 hours	7.92	3.06	7.77	3.01	8.10	3.10
24 hours	74.29	2.45	71.04	2.42	77.61	2.47

Table 2: Lorenz96 system-model pair experiment. a) Distance between the observations and the pseudo-orbits generated by WC4DVAR and ISGD, b) Distance between the true states and the pseudo-orbits generated by WC4DVAR and ISGD.

WC4DVAR fails in long assimilation window

Window length	STD of the middle point of the pseudo-orbit					
	Median		10th percentile		90th percentile	
	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$
6 hours	0.0489	0.0402	0.0391	0.0295	0.0815	0.0697
12 hours	0.0540	0.0314	0.0411	0.0236	0.1045	0.0674
24 hours	0.2132	0.0309	0.1642	0.0227	0.3505	0.0662

Window length	STD of the end point of the pseudo-orbit					
	Median		10th percentile		90th percentile	
	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$	WC4DVAR	$ISGD^c$
6 hours	0.0563	0.0480	0.0429	0.0243	0.0934	0.0744
12 hours	0.0743	0.0477	0.0573	0.0238	0.1332	0.0741
24 hours	0.2444	0.0477	0.1859	0.0236	0.3949	0.0740

Table 3: Lorenz96 system-model pair experiment, Statistics of the standard deviation of pseudo-orbits' components for different lengths of assimilation window, for each assimilation window. a) Standard deviation of the middle point of the pseudo-orbit, b) Standard deviation of the end point of the pseudo-orbit.

Forming ensemble

- Perturbations.
- Apply the ISGD method on perturbed pseudo-orbit.
- Apply the ISGD method on the results of other data assimilation methods. Particle filter?

How do we evaluate nowcasts??

Conclusion

- Sensitivity to initial conditions limits the ability to identify the current state of nonlinear dynamical system.
- Given the noise model and perfect model, there exists a set of indistinguishable states which can not be distinguished from each other.
- Form the ensemble by draw samples from the set of indistinguishable states beats the Ensemble Kalman Filter method as IS ensemble contains the information from both model dynamics and observations.
- Outside PMS, there are no model trajectories but pseudo-orbits are consistent with the observations
- Applying the ISGD method with a stopping criteria produces more “relevant” pseudo-orbits and informative estimation of model error.

Thank you!

H.L.Du@lse.ac.uk

Centre for the Analysis of Time Series:

http://www2.lse.ac.uk/CATS/home.aspx