An introduction to the EnKF

Pavel Sakov

Bureau of Meteorology

Student Workshop on Data Assimilation

1–2 December 2016, Melbourne

Outline

DA: the minimisation problem

Kalman filter

EnKF

Asynchronous DA

DA: the minimisation problem

$$\{\mathbf{x}_i^a\}_{i=1}^K = \arg\min_{\{\mathbf{x}_i\}_{i=1}^K} \mathcal{L}_K(\mathbf{x}_1, \ldots, \mathbf{x}_K),$$

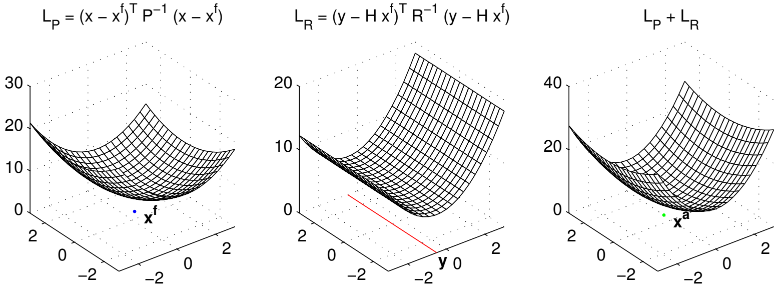
$$\mathcal{L}_K(\mathbf{x}_1, \ldots, \mathbf{x}_K) = (\mathbf{x}_1 - \mathbf{x}_1^f)^{\mathrm{T}}(\mathbf{P}_1^f)^{-1}(\mathbf{x}_1 - \mathbf{x}_1^f)$$

$$+ \sum_{i=1}^K [\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)]^{\mathrm{T}} (\mathbf{R}_i)^{-1} [\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)]$$

$$+ \sum_{i=2}^K [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1})]^{\mathrm{T}} (\mathbf{Q}_i)^{-1} [\mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1})]$$

$\mathcal{L}_i$ – cost function
i – cycle number
$\mathbf{x}_i$ – state estimate at $t = t_i$
$\mathbf{x}_1^f$ – initial state estimate
$\mathbf{P}_1^f$ – initial state error covariance estimate
$\mathbf{y}_i$ – observations
$\mathcal{H}_i$ – observation operator
$\mathcal{H}_i(\mathbf{x}_i)$ – estimated observations
$\mathbf{R}_i$ – observation error covariance
$\mathcal{M}_i$ – model operator
$\mathcal{M}_i(\mathbf{x}_{i-1})$ – state estimate propagated from t_{i-1} to t_i
$\mathbf{Q}_i$ - model error covariance

Cost function example



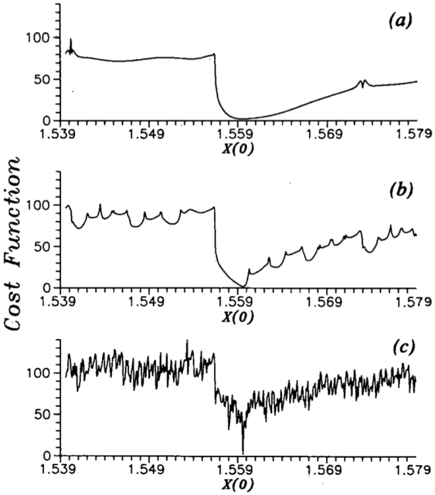
$$\mathbf{x}^f = (0.5, 0.5)^{\mathrm{T}}, \quad \mathbf{P}^f = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\mathbf{y} = -0.5, \quad \mathbf{R} = 1, \quad \mathbf{H} = (1\ 0)$$

$$\mathbf{x}^a = (0, 0.5)^{\mathrm{T}}$$

Cost function example

From Miller et al. (1994):



Lorenz (1963) model:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = \rho x - y - xz \\ \dot{z} = xy - \beta z \end{cases}$$

$$\sigma = 10, \ \rho = 28, \ \beta = 8/3$$

$$\Delta t = 0.25, \mathbf{R} = 2\mathbf{I}$$

FIG. 6. Values of cost function as a function of initial X, with initial Y and Z held constant, in the neighborhood of the initial values used in calculating the reference solution. (a) Cost function, i.e., mean-square deviation of model solution with given initial data from "observed" values, where observations up to $t = 8$ are considered. (b) As in (a) but for observations up to $t = 10$. (c) As in (a) but for observations up to $t = 15$.

The Kalman filter: linear recursive solution

- Consider the minimisation problem in the linear case

$$\mathcal{M}_i(\mathbf{x}^{(1)}) - \mathcal{M}_i(\mathbf{x}^{(2)}) = \mathbf{M}_i\,(\mathbf{x}^{(1)} - \mathbf{x}^{(2)})$$

$$\mathcal{H}_i(\mathbf{x}^{(1)}) - \mathcal{H}_i(\mathbf{x}^{(2)}) = \mathbf{H}_i\,(\mathbf{x}^{(1)} - \mathbf{x}^{(2)})$$

- Let us assume that

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k-1}} \mathcal{L}_k(\mathbf{x}_1, \ldots, \mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^a)^{\mathrm{T}} (\mathbf{P}_k^a)^{-1} (\mathbf{x}_k - \mathbf{x}_k^a) + \mathrm{Const}$$

- Then

$$\min_{\{\mathbf{x}_i\}_{i=1}^{k}} \mathcal{L}_{k+1}(\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{x}_{k+1}) = (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^a)^{\mathrm{T}} (\mathbf{P}_{k+1}^a)^{-1} (\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^a) + \mathrm{Const},$$

  where

$$\left.\begin{aligned} \mathbf{x}_{k+1}^a &= \mathbf{x}_{k+1}^f + \mathbf{K}_{k+1}[\mathbf{y}_{k+1} - \mathcal{H}_{k+1}(\mathbf{x}_{k+1}^f)] \\ \mathbf{K}_{k+1} &= \mathbf{P}_{k+1}^f (\mathbf{H}_{k+1})^{\mathrm{T}} [\mathbf{H}_{k+1} \mathbf{P}_{k+1}^f (\mathbf{H}_{k+1})^{\mathrm{T}} + \mathbf{R}_{k+1}]^{-1} \\ \mathbf{P}_{k+1}^a &= (\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}_{k+1})\mathbf{P}_{k+1}^f \end{aligned}\right\} \quad \text{"analysis"}$$

$$\left.\begin{aligned} \mathbf{x}_{k+1}^f &= \mathcal{M}_k(\mathbf{x}_k^a) \\ \mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a (\mathbf{M}_k)^{\mathrm{T}} + \mathbf{Q}_k \end{aligned}\right\} \quad \text{"propagation"}$$

A few notes about the KF solution

- The state $\mathbf{X}$ of the DA system is carried by the estimated model state vector and state error covariance:

$$\mathbf{X} = \{\mathbf{x}_i, \mathbf{P}_i\}$$

- The KF provides solution for the *last* analysis $\mathbf{x}_k^a$. As soon as the last observations are assimilated, the previous analyses $\mathbf{x}_1^a, \dots, \mathbf{x}_{k-1}^a$ are no longer optimal and can be improved by applying a *smoother*:

$$\{\mathbf{x}_1^s, \mathbf{x}_2^s, \dots, \mathbf{x}_{k-1}^s, \mathbf{x}_k^a\} = \arg \min_{\{\mathbf{x}_i\}_{i=1}^k} \mathcal{L}_k(\mathbf{x}_1, \dots, \mathbf{x}_k)$$

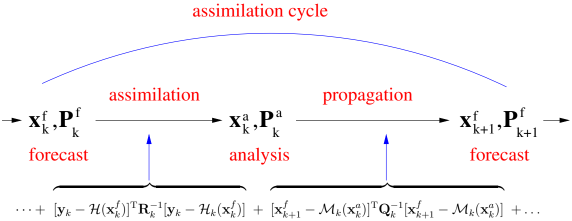
- Because the SDAS represents a (part of a) solution of the global least squares problem, it does not depend on the order in which observations are assimilated or on their grouping
- The SDAS does not depend on a linear non-singular transform of the model state
- The KF solution can be *used* in a nonlinear case by approximating

$$\mathbf{M}_i \leftarrow \nabla \mathcal{M}_i(\mathbf{x}_{i-1}^a)$$

$$\mathbf{H}_i \leftarrow \nabla \mathcal{H}_i(\mathbf{x}_i^f)$$

("extended Kalman filter", or EKF)

Terminology



- State estimate after assimilation $\mathbf{x}^a$ – analysis
- Quantity $\mathbf{X}$ after assimilation $\mathbf{X}^a$ – analysed $\mathbf{X}$ (or "analysis X")
- Advancing the analysis in time $\mathcal{M}(\mathbf{x}^a)$ – propagation
- State estimate after propagation $\mathbf{x}^f_{k+1} = \mathcal{M}_k(\mathbf{x}^a_k)$ – forecast
- Quantity $\mathbf{X}$ after propagation $\mathbf{X}^f$ – forecast $\mathbf{X}$
- One cycle of propagation and analysis – assimilation cycle
- Time interval between analyses – cycle length
- Time interval of assimilated observations – observation window
- The difference $\mathbf{y} - \mathcal{H}(\mathbf{x})$ – innovation
- The sensitivity $\mathbf{K}$ of analysis to innovation – Kalman gain, $\mathbf{x}^a - \mathbf{x}^f = \mathbf{K}\,[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)]$.

Problems with EKF

- Non-scalable: $\mathbf{P}$ is a $n \times n$ matrix
  (e.g. ocean model OFAM3: $\sim 10^9$ state elements)
- Complex: requires tangent linear model (TLM) $\mathbf{M} = \nabla\mathcal{M}(\mathbf{x})$ and adjoint model (AM) $\mathbf{M}^{\mathrm{T}}$
- Purely sequential; can not assimilate asynchronous observations
- In canonical form – can become numerically inconsistent
  ($\mathbf{P}$ looses positive definiteness)

Ensemble Kalman filter

KF: $\mathrm{SDAS} = \{\mathbf{x}, \mathbf{P}\}$

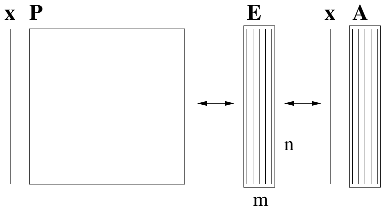
EnKF: $\mathrm{SDAS} = \{\mathbf{E}\} = \{\mathbf{x}, \mathbf{A}\}$

where:

$$\mathbf{x} = \frac{1}{m}\mathbf{E}\,\mathbf{1}$$

$$\mathbf{P} = \frac{1}{m-1}\mathbf{A}\mathbf{A}^{\mathrm{T}}$$

$$\mathbf{A} \equiv \mathbf{E} - \mathbf{x}\mathbf{1}^{\mathrm{T}}$$



	KF	EnKF
propagation	$\mathbf{x}^f_{k+1} = \mathcal{M}_k(\mathbf{x}^a_k)$	$\mathbf{E}^a_{k+1} = \mathcal{M}_k(\mathbf{E}^f_k)$
	$\mathbf{P}^f_{k+1} = \mathbf{M}_k\mathbf{P}^a_k\mathbf{M}^{\mathrm{T}}_k + \mathbf{Q}_k$	
analysis	$\mathbf{x}^a_k = \mathbf{x}^f_k + \mathbf{K}_k[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^f_k)]$	$\mathbf{x}^a_k = \mathbf{x}^f_k + \mathbf{A}^f_k\mathbf{w}_k$
	$\mathbf{P}^a_k = (\mathbf{I} - \mathbf{K}\mathbf{H}_k)\mathbf{P}^f_k$	$\mathbf{A}^a_k = \mathbf{A}^f_k\mathbf{T}_k$

A solution for $\mathbf{w}, \mathbf{T}$

Lemma Let $\mathbf{P}^f = \mathbf{A}^f(\mathbf{A}^f)^{\mathrm{T}}/(m-1)$, $\mathbf{A}^f\mathbf{1} = \mathbf{0}$.
Then any solution of

$$\mathbf{A}^a(\mathbf{A}^a)^{\mathrm{T}}/(m-1) \equiv \mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f,$$
$$\mathbf{A}^a\mathbf{1} = \mathbf{0}$$

can be written as

$$\mathbf{A}^a = \mathbf{A}^f\mathbf{T}\mathbf{U},$$

where

$$\mathbf{T} = \left[\mathbf{I} + (\mathbf{H}\mathbf{A}^{\mathbf{f}})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{A}^f\right]^{-1/2}$$
$$\mathbf{U}: \ \mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}, \ \mathbf{U}\mathbf{1} = 0.$$

Also,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{A}^f\mathbf{w}, \quad \mathbf{w} = \left[\mathbf{I} + (\mathbf{HA})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{HA}\right]^{-1}(\mathbf{HA})^{\mathrm{T}}\left[\mathbf{y} - \mathcal{H}(\mathbf{x}^f)\right].$$

Observation Both $\mathbf{T}$ and $\mathbf{w}$ depend only on observed elements of the state: $\mathbf{T} = \mathbf{T}(\mathcal{H}(\mathbf{E}^f), \mathbf{R})$, $\mathbf{w} = \mathbf{w}(\mathcal{H}(\mathbf{E}^f), \mathbf{y}, \mathbf{R})$.

EnKF: an algorithm example

$$\textbf{function } [\mathbf{E}_2] = \textbf{enkf\_cycle}(\mathbf{E}_1, \mathbf{y}_2, \mathbf{R}_2, \mathcal{M}_{12}, \mathcal{H}_2)$$

$$\mathbf{E}_2 = \mathcal{M}_{12}(\mathbf{E}_1)$$

$$\mathbf{HE}_2 = \mathcal{H}_2(\mathbf{E}_2)$$

$$\mathbf{Hx}_2 = \mathbf{HE}_2\,\mathbf{1}/m \qquad \mathcal{H}(\mathbf{x}_2^f) \leftarrow \mathcal{H}(\mathbf{E}_1^f)\,\mathbf{1}/m$$

$$\mathbf{HA}_2 = \mathbf{HE}_2 - \mathbf{Hx}_2\mathbf{1}^{\mathrm{T}} \qquad \mathbf{HA}_2^f \leftarrow \mathcal{H}(\mathbf{E}_2^f)(\mathbf{I} - \mathbf{11}^{\mathrm{T}}/m)$$

$$\mathbf{D} = [\mathbf{I} + (\mathbf{HA_2})^{\mathrm{T}}\mathbf{R}_2^{-1}\mathbf{HA_2}/(m-1)]^{-1}$$

$$\mathbf{w} = \mathbf{D}\,(\mathbf{HA})^{\mathrm{T}}\mathbf{R}_2^{-1}[\mathbf{y}_2 - \mathbf{Hx}_2]$$

$$\mathbf{T} = \mathbf{D}^{1/2}$$

$$\mathbf{E}_2 = \mathbf{E}_2\,(\mathbf{w1}^{\mathrm{T}} + \mathbf{T}) \qquad \mathbf{x}_2^a = \mathbf{x}_2^f + \mathbf{A}_2^f\mathbf{w}, \quad \mathbf{A}_2^a = \mathbf{A}_2^f\mathbf{T}$$

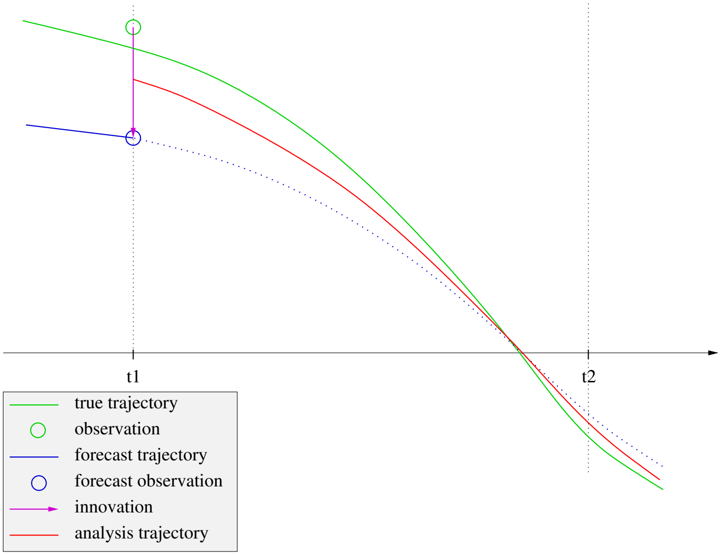
Synchronous and asynchronous data assimilation

Synchronous, or “3D” assimilation = observations are assumed to be taken at the assimilation time

Asynchronous, or “4D” assimilation = observations can be taken at time different than the assimilation time

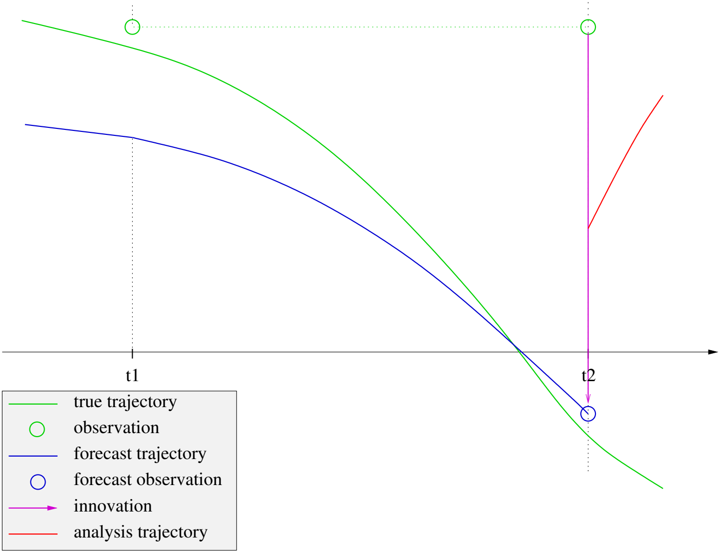
Asynchronous DA

Assimilation at the time of observation



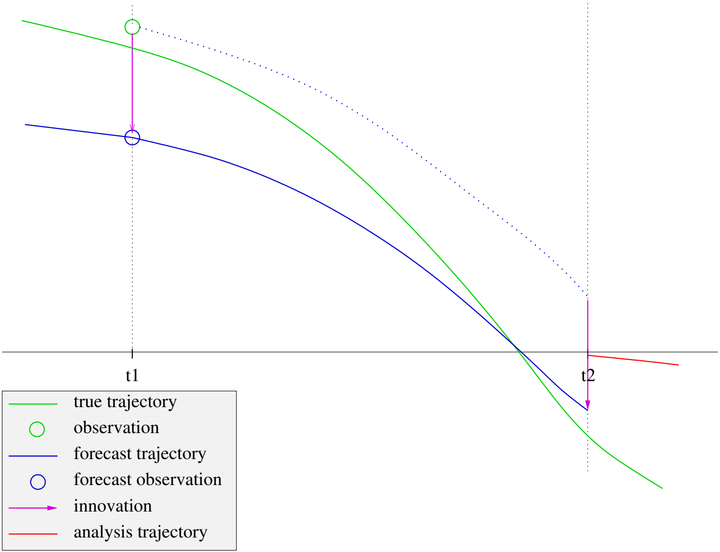
Asynchronous DA

Ignoring the time of observation ("3D")



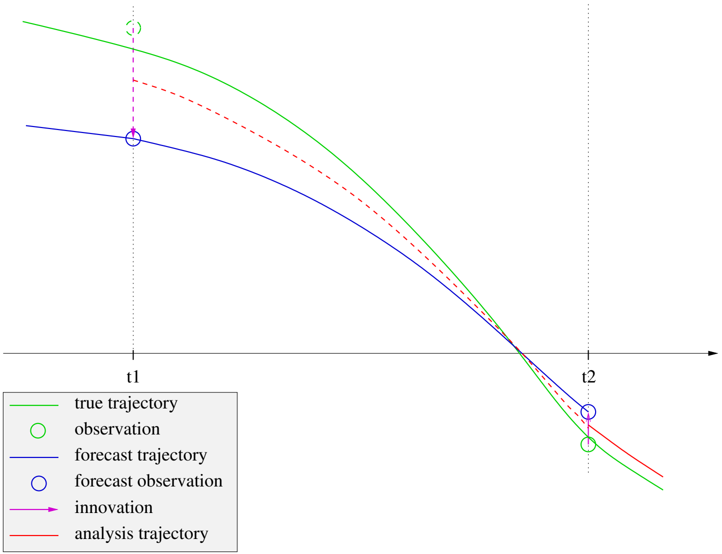
Asynchronous DA

Using the actual innovation ("FGAT")



Asynchronous DA

Asunchronous DA ("4D")



Evolution of corrections

Let us assimilate observations at t_1

$$\delta \mathbf{x}_1 \equiv \mathbf{x}_1^a - \mathbf{x}_1^f = \mathbf{A}_1^f \boxed{\mathbf{w}_1}$$

$$\delta \mathbf{A}_1 \equiv \mathbf{A}_1^a - \mathbf{A}_1^f = \mathbf{A}_1^f \boxed{\mathbf{T}_1}$$

Let $\mathbf{M}_{12}$ be the tangent linear propagator along the forecast system trajectory between t_1 and t_2:

$$\delta \mathbf{x}_2 = \mathbf{M}_{12}\, \delta \mathbf{x}_1 + O\left(|\delta \mathbf{x}_1|^2\right)$$

At t_2 the corrections become:

$$\delta \mathbf{x}_2 \sim \mathbf{M}_{12}\, \delta \mathbf{x}_1 = \mathbf{M}_{12}(\mathbf{A}_1^f \mathbf{w}_1) = (\mathbf{M}_{12}\mathbf{A}_1^f)\mathbf{w}_1 \sim \mathbf{A}_2^f \boxed{\mathbf{w}_1}$$

$$\delta \mathbf{A}_2 \sim \mathbf{M}_{12}\, \delta \mathbf{A}_1 = \mathbf{M}_{12}(\mathbf{A}_1^f \mathbf{T}_1) = (\mathbf{M}_{12}\mathbf{A}_1^f)\mathbf{T}_1 \sim \mathbf{A}_2^f \boxed{\mathbf{T}_1}$$

Applying ensemble transforms calculated at observation time to the forecast ensemble at any other timer yields correctly evolved increments

But $\mathbf{w}, \mathbf{T}$ are only functions of forecast ensemble observations and observations.

$$\boxed{\mathbf{HE} = \left([(\mathbf{HE})_1]^{\mathrm{T}}, [(\mathbf{HE})_2]^{\mathrm{T}}, \ldots, [(\mathbf{HE})_k]^{\mathrm{T}}\right)^{\mathrm{T}}}$$

EnKF: some basics that we have not got covered

- Rank deficiency and localisation
- Nonlinearity: what is a nonlinear system?
- Suboptimality and inflation
- Diagnosing a DA system
- Advantages and disadvatanges versus variational methods
- Iterative schemes

Summary

- Kalman filter is a *linear recursive* solution of the (generally) nonlinear global in time least squares problem
- Ensemble Kalman filter is a
  - simple
  - scalable
  - deterministic
  - derivative-less

  state space formulation of the KF
- The EnKF ensemble anomalies are just a *factorisation* of the state error covariance
- which means that the EnKF "ensemble" is *not* a statistical ensemble,
- and that it is *not* supposed to carry the PDF of the state error