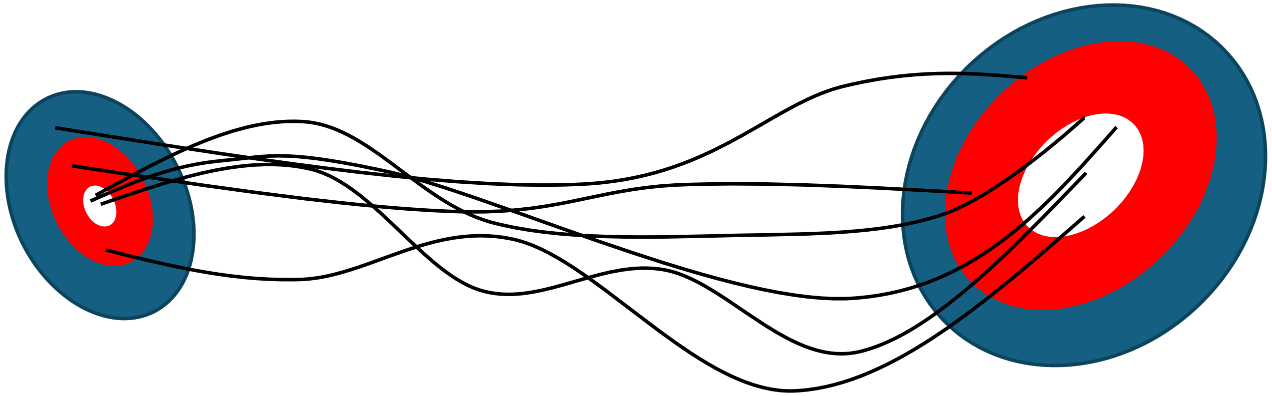
The Ensemble Kalman filter



Part I: Theory

Guannan Hu

Data-assimilation training course. 9th-13th June 2025, University of Reading

Recap of data assimilation problem

- Given prior knowledge (background) and observations, we estimate the system state at a given time
- This posterior estimate is known as analysis
- **Bayes' theorem** allows us pose this problem in terms of the respective PDFs:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$$

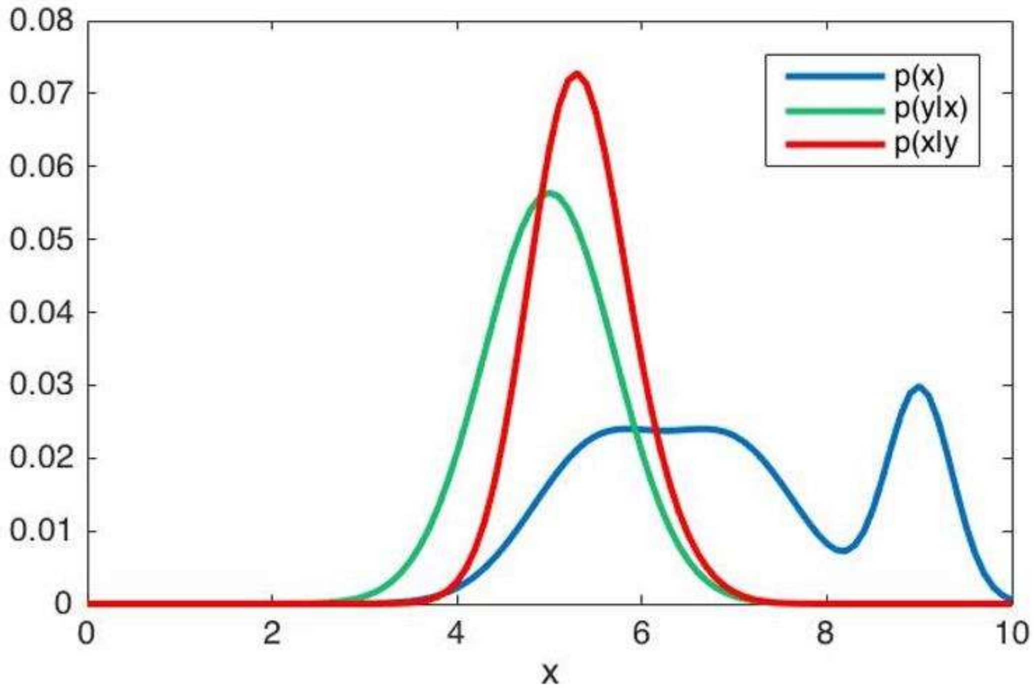
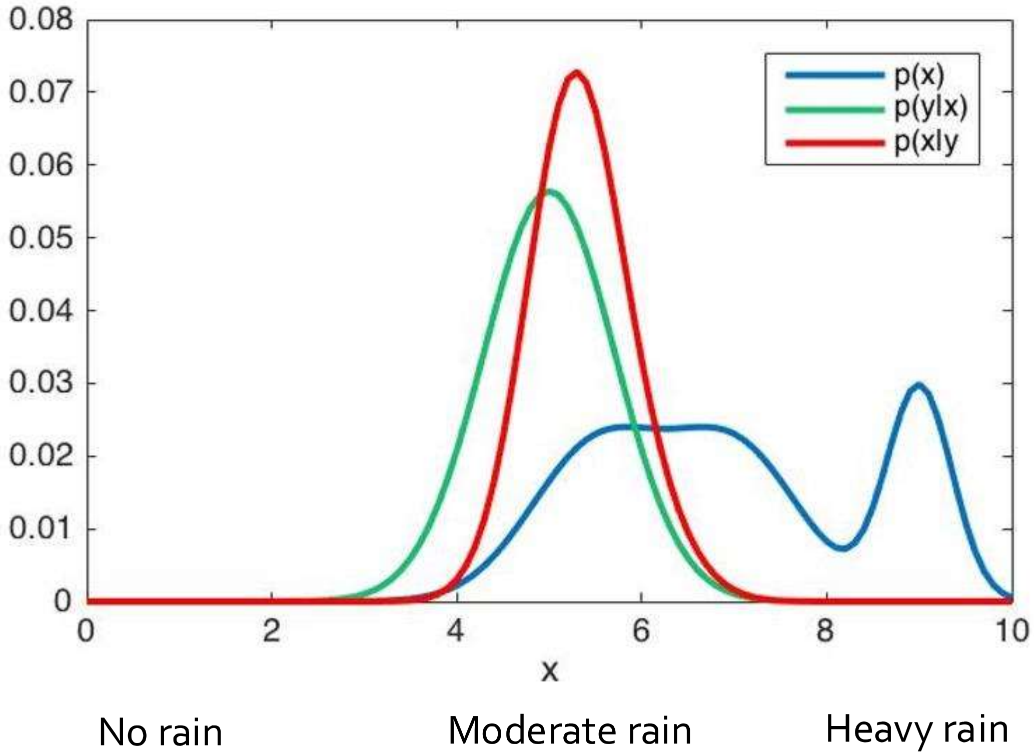


Figure: 1D example of Bayes' theorem.

An example: rainfall in a grid box

- **Background:** Uncertain whether rainfall was moderate or heavy
- **Observation:** Suggests moderate rainfall was more probable
- **Analysis:** Applying Bayes' theorem → increased confidence in moderate rainfall, with reduced uncertainty compared to either the background or observation alone



Recap of cost function

Under Gaussian assumptions

Maximise the posterior probability $p(\mathbf{x} \mid \mathbf{y})$

=

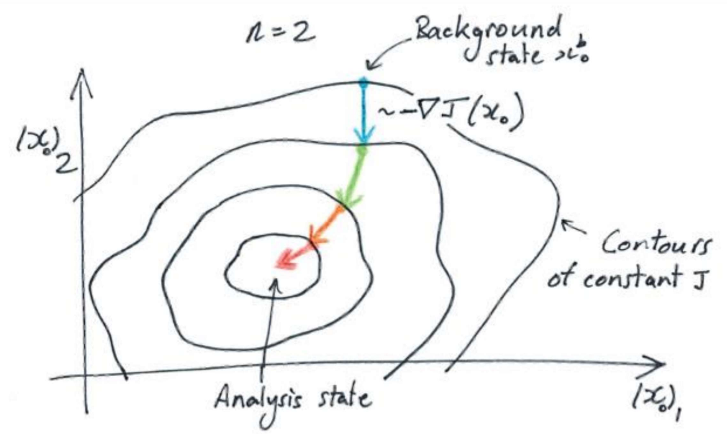
Minimise $-\log[p(\mathbf{x} \mid \mathbf{y})]$

=

Minimise the cost function $J(\mathbf{x})$

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^{\mathrm{b}})^{\top}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^{\mathrm{b}}) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^{\top}\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))$$

Recap of variational DA



Cost function:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^{\mathrm{b}})^{\top}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^{\mathrm{b}}) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^{\top}\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}))$$

- Minimising $J(\mathbf{x})$ is equivalent to solving $\nabla J(\mathbf{x}) = 0$
- This is solved using gradient-based methods (e.g., conjugate gradient)
- At each iteration, a small **variation** is applied to the state variable $\mathbf{x}$ to move toward the minimum

Why variational methods?

Well-posed problem

- Gaussian assumption
- Near-linear assumption
- Full rank B matrix

Extensive practical developments

- Control variable transform
- Incremental formulation
- Preconditioning
- Weak-constraint 4DVar

Operational proven

- Met Office
- European Centre for Medium-Range Weather Forecasts (ECMWF)
- Météo-France

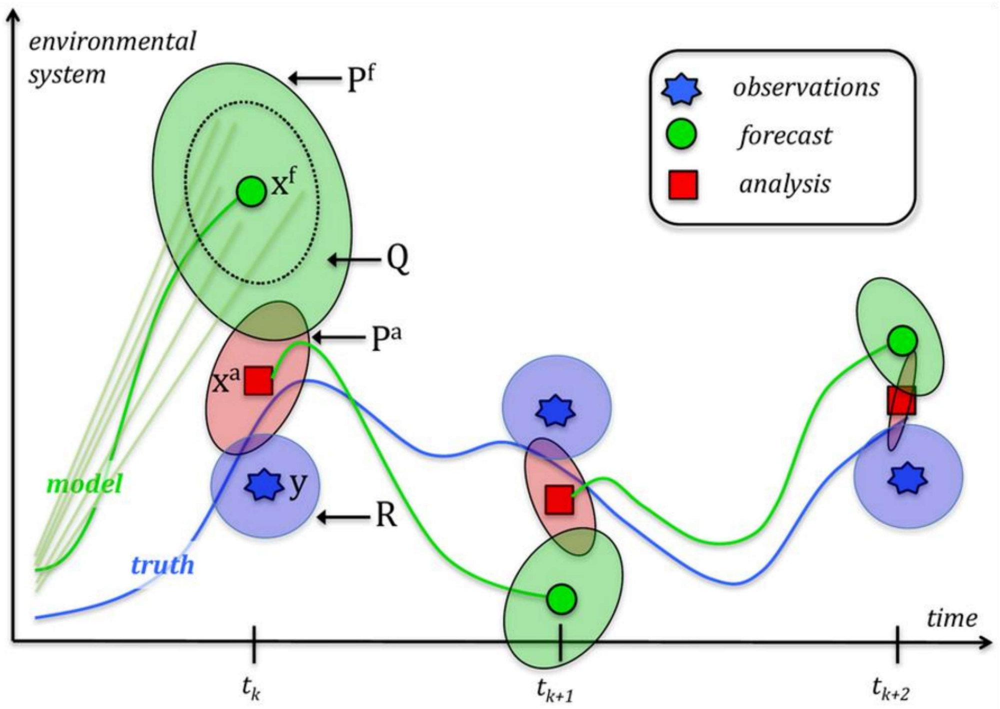
Why different methods?

Gaussian and linear assumptions not always valid	• Convection • Rainfall
Development of the Tangent Linear Model (TLM) and its adjoint	• Time-consuming • Difficult to maintain as the nonlinear model evolves
Validity of the TLM	• Restrict the length of the assimilation window (4DVar)
B matrix is predominately static	• Does not reflect flow-dependent error statistics

Ensemble Kalman filters

- Another **major class** of operational DA methods
- Do not require **iterative minimization** of a cost function
- Compute the analysis directly from an equation that approximates the **cost function solution**
- Background error covariances are estimated from a forecast **ensemble**
  - flow-dependent
  - account for model error
- Can be categorized as **stochastic** or **deterministic**
- Based on the **Kalman filter** algorithm

Kalman filter algorithm (two steps)



Tandeo et al. (2020)

Update step (t_k):

Update mean and covariance of prior using observations to obtain posterior

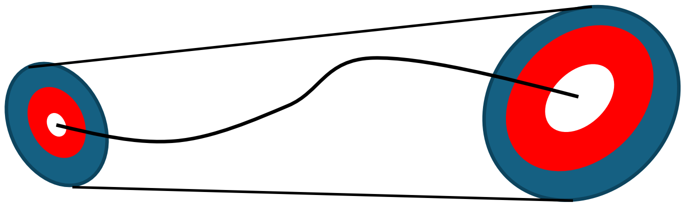
Prediction step ($t_k \rightarrow t_{k+1}$):

Evolve posterior at time t_k forward in time using a model to obtain prior at time t_{k+1}

Update step: Kalman equations

Analysis	$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{K}\left(\mathbf{y} - h\left(\mathbf{x}^{\mathrm{f}}\right)\right)$	• Analytical solution to the cost function • Updates background state using observations, a nonlinear observation operator and Kalman gain
Kalman gain	$\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}$	Depends on background and observation error covariances
Analysis error covariance	$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{f}}$	Update background error covariance, reflecting reduced uncertainty after assimilation

Prediction step

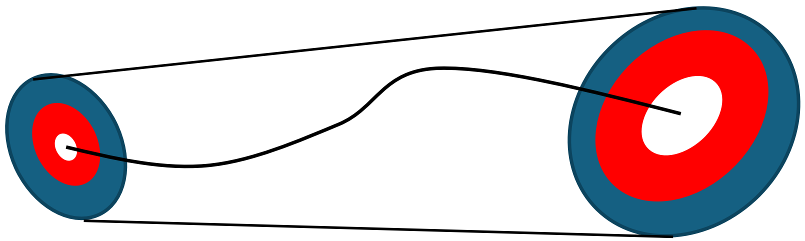


Mean	$\mathbf{x}^{\mathrm{f}}_{k+1} = M_{t_k \to t_{k+1}}(\mathbf{x}^{\mathrm{a}}_k) + \boldsymbol{\eta}_{k+1}$	• Mean state evolves in time by a forecast model (M) • The model error is represented by $\boldsymbol{\eta}_{k+1} \sim \mathcal{N}(0, \mathbf{Q}_{k+1})$
Covariance	$\mathbf{P}^{\mathrm{f}}_{k+1} = \mathbf{M}\mathbf{P}^{\mathrm{a}}_k\mathbf{M}^{\top} + \mathbf{Q}_{k+1}$	• Updating the covariance is trickier • The extended Kalman filter (EKF) does this using the TL and adjoint models ($\mathbf{M}$ and $\mathbf{M}^{\mathrm{T}}$)

Motivation for ensemble Kalman filters

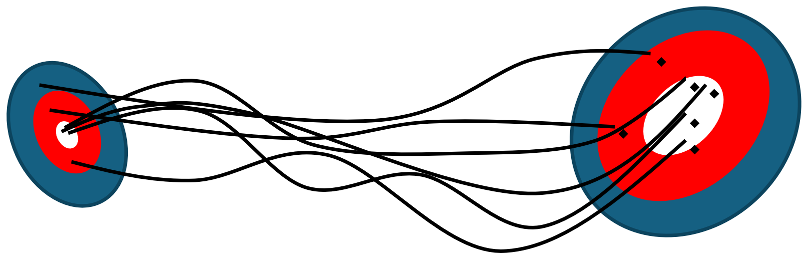
❖ Covariance propagation in the EKF requires the **TL** and **adjoint** models:

$$\mathbf{P}_{k+1}^{\mathrm{f}} = \mathbf{M}\mathbf{P}_{k}^{\mathrm{a}}\mathbf{M}^{\top} + \mathbf{Q}_{k+1}$$



*For most environmental applications, the **size of the matrices** makes it computationally expensive*

❖ **Alternative:** estimate the covariance matrix using a set of model simulations called an **ensemble**



Ensemble estimate of error covariances

$$\mathbf{P}^{\mathrm{f}} \approx \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}^{\mathrm{f}(i)} - \bar{\mathbf{x}}^{\mathrm{f}}\right)\left(\mathbf{x}^{\mathrm{f}(i)} - \bar{\mathbf{x}}^{\mathrm{f}}\right)^{\top}$$

- $\mathbf{x}^{\mathrm{f}(i)}$ = model state vector of the i-th ensemble member
- $\bar{\mathbf{x}}^{\mathrm{f}}$ = ensemble mean
- N = ensemble size
- $\mathbf{P}^{\mathrm{f}}$ = ensemble error covariance matrix

Ensemble perturbation matrix

$$\mathbf{P}^{\mathrm{f}} \approx \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}^{\mathrm{f}(i)} - \bar{\mathbf{x}}^{\mathrm{f}}\right)\left(\mathbf{x}^{\mathrm{f}(i)} - \bar{\mathbf{x}}^{\mathrm{f}}\right)^{\top} = \frac{1}{N-1}\mathbf{X}^{\mathrm{f}}\left(\mathbf{X}^{\mathrm{f}}\right)^{\top}$$

Each column of $\mathbf{X} \in \mathbb{R}^{n\times N}$ is the difference between an ensemble member and the ensemble mean

$$\mathbf{X} = \begin{bmatrix} \left(\mathbf{x}^{(1)} - \bar{\mathbf{x}}\right)_1 & \left(\mathbf{x}^{(2)} - \bar{\mathbf{x}}\right)_1 & \cdots & \left(\mathbf{x}^{(N)} - \bar{\mathbf{x}}\right)_1 \\ \left(\mathbf{x}^{(1)} - \bar{\mathbf{x}}\right)_2 & \left(\mathbf{x}^{(2)} - \bar{\mathbf{x}}\right)_2 & \cdots & \left(\mathbf{x}^{(N)} - \bar{\mathbf{x}}\right)_2 \\ \vdots & \vdots & \vdots & \vdots \\ \left(\mathbf{x}^{(1)} - \bar{\mathbf{x}}\right)_n & \left(\mathbf{x}^{(2)} - \bar{\mathbf{x}}\right)_n & \cdots & \left(\mathbf{x}^{(N)} - \bar{\mathbf{x}}\right)_n \end{bmatrix}$$

Ensemble perturbation matrix in observation space

$$h: n \to m$$

Define

$$\mathbf{y}^{\mathrm{f}} = h(\mathbf{x}^{\mathrm{f}})$$

Then

$$\mathbf{Y} = \begin{bmatrix} \left(\mathbf{y}^{\mathrm{f}(1)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_1 & \left(\mathbf{y}^{\mathrm{f}(2)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_1 & \cdots & \left(\mathbf{y}^{\mathrm{f}(N)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_1 \\ \left(\mathbf{y}^{\mathrm{f}(1)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_2 & \left(\mathbf{y}^{\mathrm{f}(2)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_2 & \cdots & \left(\mathbf{y}^{\mathrm{f}(N)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_2 \\ \vdots & \vdots & \vdots & \vdots \\ \left(\mathbf{y}^{\mathrm{f}(1)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_m & \left(\mathbf{y}^{\mathrm{f}(2)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_m & \cdots & \left(\mathbf{y}^{\mathrm{f}(N)} - \bar{\mathbf{y}}^{\mathrm{f}}\right)_m \end{bmatrix} \in \mathbb{R}^{m \times N}$$

For linear observation operators

$$\mathbf{Y} = \mathbf{H}\mathbf{X}$$

$$\frac{1}{N-1}\mathbf{Y}\mathbf{Y}^{\top} = \frac{1}{N-1}\mathbf{H}\mathbf{X}\mathbf{X}^{\top}\mathbf{H}^{\top} = \mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}$$

Classic EnKF update (Envensen 1994)

- Kalman gain can be expressed using ensemble perturbation matrices

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1} = \mathbf{X}^{\mathrm{f}}\left(\mathbf{Y}^{\mathrm{f}}\right)^{\top}\left(\mathbf{Y}^{\mathrm{f}}\left(\mathbf{Y}^{\mathrm{f}}\right)^{\top} + (N-1)\mathbf{R}\right)^{-1}$$

- **Perturbated observations** for the i-th ensemble member

$$\mathbf{y}^{(i)} = \mathbf{y} + \boldsymbol{\varepsilon}^{(i)},\ \boldsymbol{\varepsilon}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

- Analysis of each member

$$\mathbf{x}^{\mathrm{a}(i)} = \mathbf{x}^{\mathrm{f}(i)} + \mathbf{K}\left(\mathbf{y}^{(i)} - h\left(\mathbf{x}^{\mathrm{f}(i)}\right)\right)$$

Why perturb observations?

Without perturbing observations, the ensemble estimate of $\mathbf{P}^{\mathrm{a}}$ is

$$\frac{1}{N-1}\mathbf{X}^{\mathrm{a}}(\mathbf{X}^{\mathrm{a}})^{\top} = (\mathbf{I}-\mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}-\mathbf{KH})^{\top}$$

With perturbing observation, it becomes

$$\frac{1}{N-1}\mathbf{X}^{\mathrm{a}}(\mathbf{X}^{\mathrm{a}})^{\top} = (\mathbf{I}-\mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}-\mathbf{KH})^{\top} + \mathbf{KRK}^{\top} = (\mathbf{I}-\mathbf{KH})\mathbf{P}^{\mathrm{f}}$$

which matches the Kalman filter solution (Burgers et al. 1998)

Why EnKF is more efficient?

Background error covariances in the EKF is

$$\mathbf{P}_{k+1}^{\mathrm{f}} = \mathbf{M}\mathbf{P}_{k}^{\mathrm{a}}\mathbf{M}^{\top} + \mathbf{Q}_{k+1}$$

- Requires matrix-matrix multiplications of size $n \times n$
- For numerical weather prediction, typically $n = 10^8$

Background error covariances in the EnKF is

$$\mathbf{P}_{k+1}^{\mathrm{f}} \approx \frac{1}{N-1}\mathbf{X}_{k+1}^{\mathrm{f}}\left(\mathbf{X}_{k+1}^{\mathrm{f}}\right)^{\top}$$

- Requires matrix-matrix multiplications of size $n \times N$
- For numerical weather prediction, typically $N = 10^2$

Stochastic vs deterministic filters

The classic EnKF is **stochastic**

→ requires perturbing observations

$$\mathbf{y}^{(i)} = \mathbf{y} + \boldsymbol{\varepsilon}^{(i)}$$

- Ensure ensemble correctly samples the analysis error covariance

$$\frac{1}{N-1}\mathbf{X}^{\mathrm{a}}(\mathbf{X}^{\mathrm{a}})^{\mathsf{T}} \approx (\mathbf{I}-\mathbf{KH})\mathbf{P}^{\mathrm{f}}$$

- Introduce additional sampling noise

Deterministic (square root) filters:

- EnSRF (Whitaker and Hamill, 2002)
- **ETKF (Bishop et al., 2001)**
- EAKF (Anderson, 2001)
- LETKF (Hunt et al., 2007)

- Avoid the need to perturb observations
- While still obtain the correct analysis error covariance

Square root filters

Idea: do not update each ensemble member separately, but update ensemble mean and perturbation simultaneously

$$\bar{\mathbf{x}}^{\mathrm{a}} = \bar{\mathbf{x}}^{\mathrm{f}} + \mathbf{K}\left(\mathbf{y} - \overline{h(\mathbf{x}^{\mathrm{f}})}\right)$$
$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}}\mathbf{T}$$

The **transformation** matrix $\mathbf{T}$ is chosen such that

$$\frac{1}{N-1}\mathbf{X}^{\mathrm{f}}\mathbf{T}\left(\mathbf{X}^{\mathrm{f}}\mathbf{T}\right)^{\top} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{f}}$$

- Matrix $\mathbf{T}$ is not uniquely defined $\rightarrow$ different deterministic filters
- Although different filters lead to different ensembles, they all span the same subspace (Tippet et al., 2003)

Transformation matrix

Using the Kalman gain

$$\mathbf{K} = \mathbf{X}^{\mathrm{f}}(\mathbf{Y}^{\mathrm{f}})^{\top}\left(\mathbf{Y}^{\mathrm{f}}(\mathbf{Y}^{\mathrm{f}})^{\top} + (N-1)\mathbf{R}\right)^{-1} = \mathbf{X}^{\mathrm{f}}\mathbf{Z}$$

we have

$$(\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{f}} = \frac{1}{N-1}\left(\mathbf{I} - \mathbf{X}^{\mathrm{f}}\mathbf{Z}\,\mathbf{H}\right)\mathbf{X}^{\mathrm{f}}\left(\mathbf{X}^{\mathrm{f}}\right)^{\top}$$

$$= \frac{1}{N-1}\mathbf{X}^{\mathrm{f}}\left(\mathbf{I} - \mathbf{Z}\,\mathbf{Y}^{\mathrm{f}}\right)\left(\mathbf{X}^{\mathrm{f}}\right)^{\top}$$

Thus, we seek the **square root**

$$\mathbf{T}\mathbf{T}^{\top} = \left(\mathbf{I} - \mathbf{Z}\,\mathbf{Y}^{\mathrm{f}}\right)$$

$$\mathbf{P}^{\mathrm{a}} = \frac{1}{N-1}\mathbf{X}^{\mathrm{f}}\mathbf{T}(\mathbf{X}^{\mathrm{f}}\mathbf{T})^{\mathrm{T}}$$

Ensemble transform Kalman filter

Using the **Sherman-Morrison-Woodbury** formula (Equation 2.1.4 of Golub & Van Loan, 1996), we obtain

$$\mathbf{T}\mathbf{T}^{\top} = \left(\mathbf{I} + \frac{1}{N-1}\left(\mathbf{Y}^{\mathrm{f}}\right)^{\top}\mathbf{R}^{-1}\mathbf{Y}^{\mathrm{f}}\right)^{-1} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\top}\right)^{-1}$$

Then, possible **transformation matrices** are

$$\mathbf{T} = \mathbf{U}\boldsymbol{\Sigma}^{-1/2}$$
$$\mathbf{T} = \mathbf{U}\boldsymbol{\Sigma}^{-1/2}\mathbf{U}^{\top}$$

Note: not all **T** satisfying the estimate of the analysis error covariance led to unbiased analysis ensembles, and a sufficient condition is $\mathbf{X}^{\mathrm{a}}\mathbf{1}_N = \mathbf{X}^{\mathrm{f}}\mathbf{T}\mathbf{1}_N = \mathbf{0}$ (Livings et al., 2008; Wang et al., 2004).

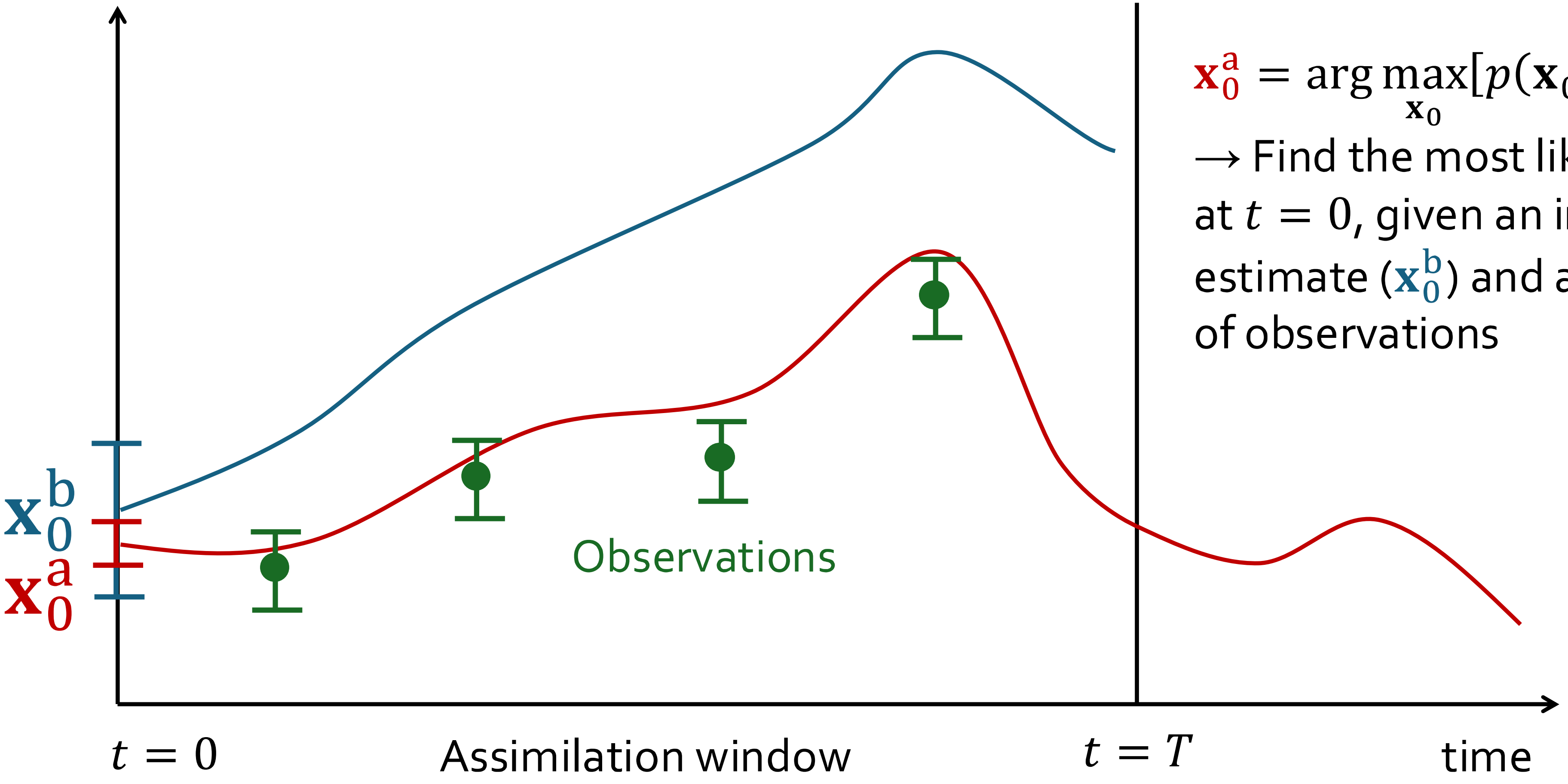
Treatment of model error

- EnKF allows for an **imperfect model** by adding noise at each time step of the model evolution

$$\mathbf{x}_k^{\mathrm{f}(i)} = M_{t_{k-1}\to t_k}(\mathbf{x}_{k-1}^{\mathrm{a}(i)}) + \boldsymbol{\eta}_k^{(i)}, \text{ where } \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

- Strategies for representing model error (depending on the assumed error source):
  - Multiphysics
  - Stochastic kinetic energy backscatter
  - Stochastically perturbed physical tendencies
  - Perturbed parameters
  - Or combinations of the above

Assimilation window in 4DVar

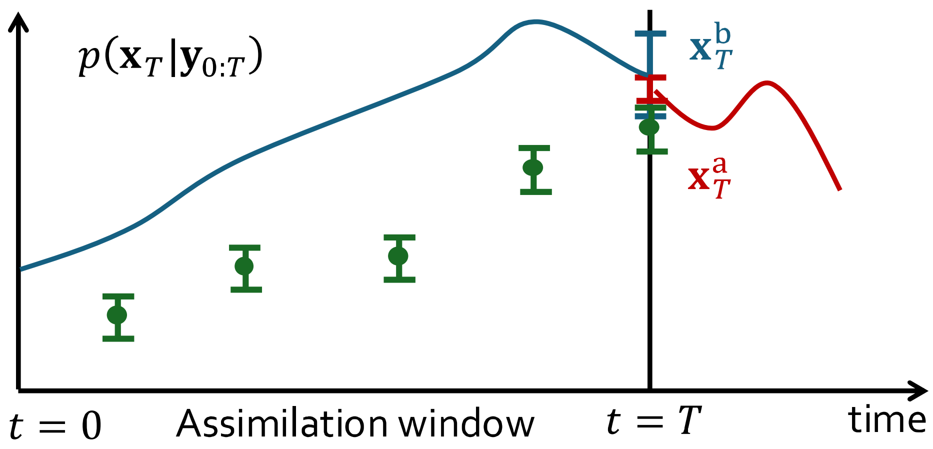


$\mathbf{x}_0^{\mathrm{a}} = \arg\max_{\mathbf{x}_0}[p(\mathbf{x}_0|\mathbf{y}_{0:T})]$

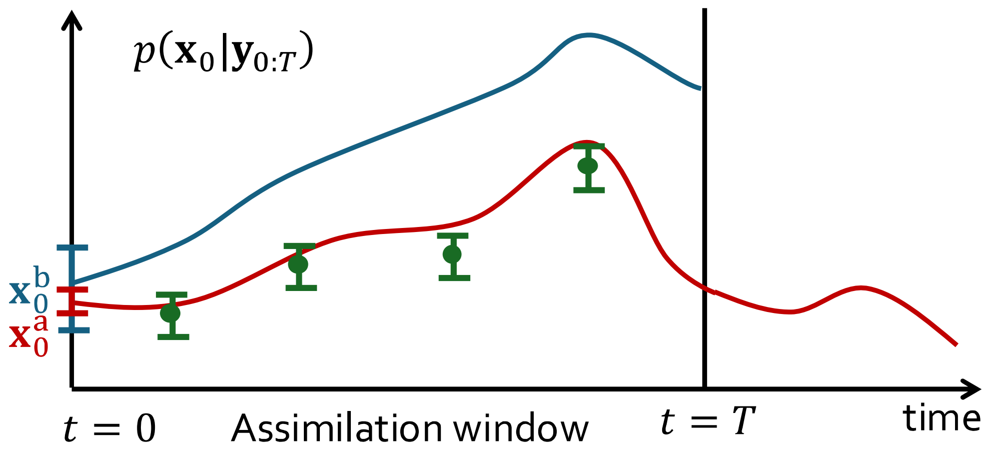
$\longrightarrow$ Find the most likely state at $t = 0$, given an initial estimate ($\mathbf{x}_0^{\mathrm{b}}$) and a window of observations

Kalman filter vs smoother

Filter uses observations **before** the analysis time

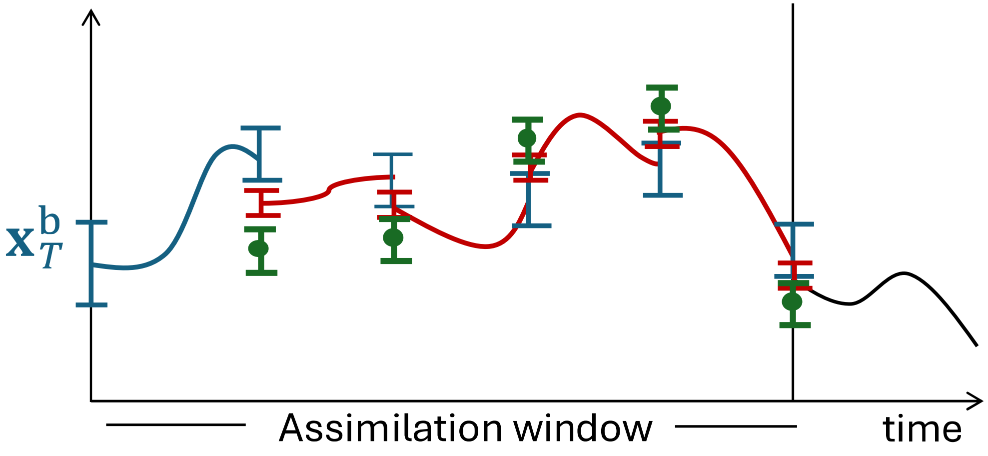


Smoother uses observations **after** (and before) the analysis time



Sequential update

- Observations can be assimilated sequentially in time, rather than assimilated at one time
- The two cases are equivalent, given
  - Linear model and observation operator
  - Gaussian errors
  - Prior error covariances specified and evolved in time exactly



Summary of ensemble Kalman filters

Advantages

- **Flow-dependent** background error statistics
- No need of the development of **TL and adjoint** models
- Easy to account for **model error**
- Easy to **parallelize**

Disadvantages

- **Sensitive to ensemble size**
  - $\rightarrow$ under sampling can lead to filter divergence and spurious correlations
  - $\rightarrow$ mitigated by **localisation** and **inflation** techniques (Tomorrow)
- **Costly to run** multiple versions of a forecast
- Assumes Gaussian statistics
  - $\rightarrow$ may be invalid in highly nonlinear systems (Tomorrow)