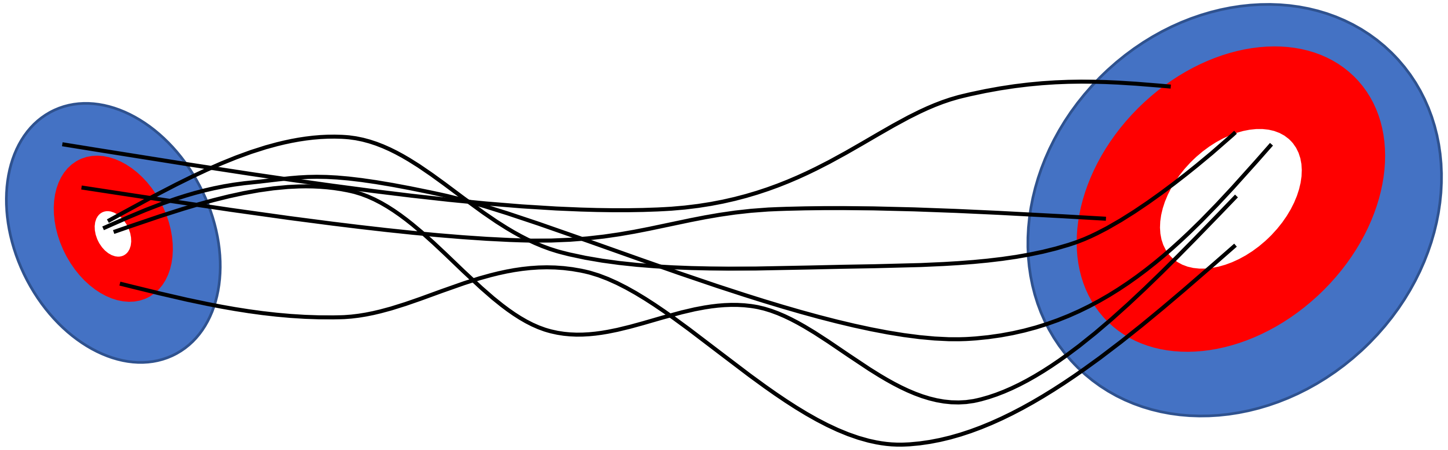
The Ensemble Kalman filter



Part I: Theory

Alison Fowler

Data-assimilation training course. 21^{st}-24^{th} June, University of Reading

Recap of Data Assimilation problem

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

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

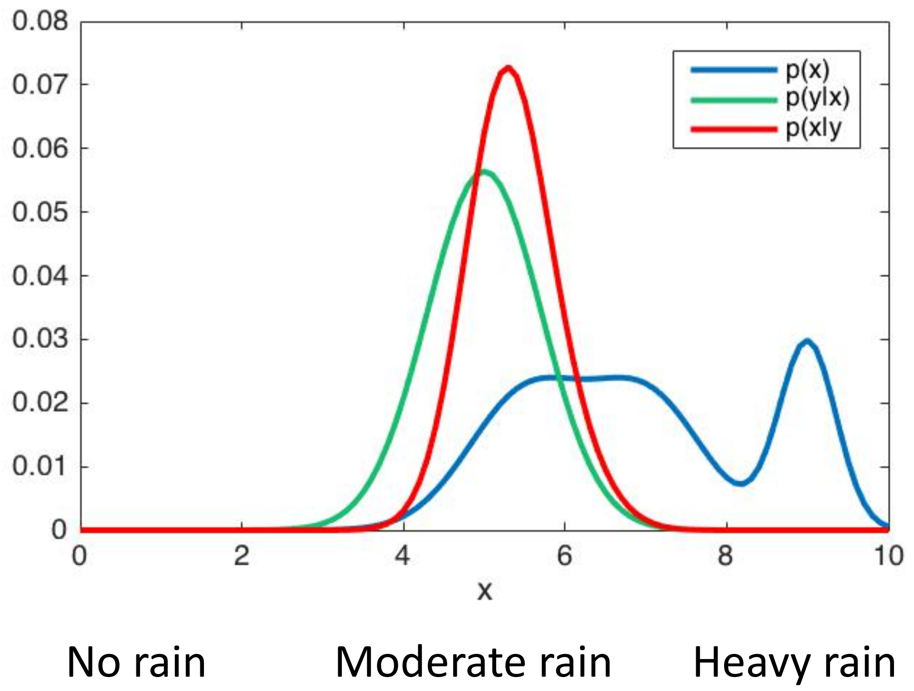


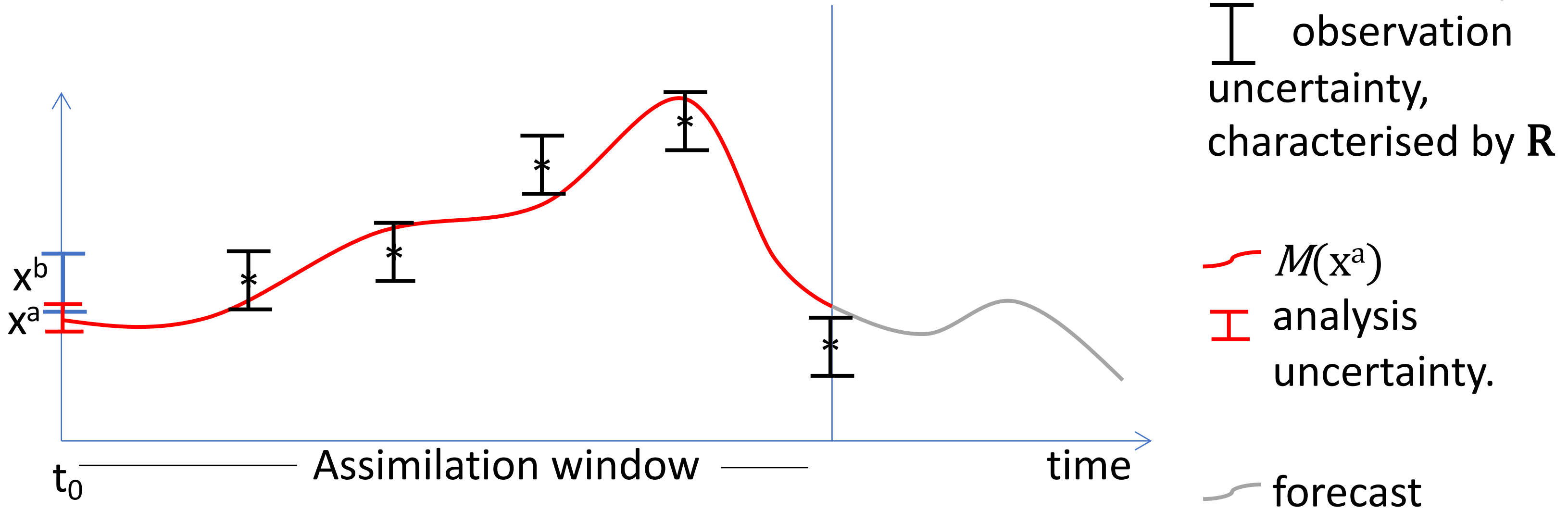
Figure: 1D example of Bayes' theorem.

For example, this could be rainfall amount in a given grid box.
A-priori we are unsure if there will be moderate or heavy rainfall. The observation only gives probability to the rainfall being moderate.
Applying Bayes' theorem, we can now be certain that the rainfall was moderate, and the uncertainty is reduced compared to both the observations and our a-priori estimate.

Recap of Data Assimilation problem

- Quiz

Recap of 4DVar



- 4DVar aims to find the most likely state at time t_0, given an initial estimate, $\mathbf{x}_0^b$, and a window of observations at p observation times.

$$
\begin{aligned}
\mathbf{x}_0^a &= \arg\max_{\mathbf{x}_0}(p(\mathbf{x}_0|\mathbf{y}_1, \ldots, \mathbf{y}_p)) \\
&= \arg\min_{\mathbf{x}_0}(-log(p(\mathbf{x}_0|\mathbf{y}_1, \ldots, \mathbf{y}_p))) \\
&= \arg\min_{\mathbf{x}_0}(J(\mathbf{x}_0))
\end{aligned}
$$

Recap of 4DVar

- J (the cost function) is derived assuming Gaussian error distributions and a perfect model.

$$J(\mathbf{x}_0) = \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)$$
$$+ \sum_{i=1}^{p} \left(\mathbf{y}_i - h(M_{t_0 \to t_i}(\mathbf{x}_0))\right)^T \mathbf{R}_i^{-1} \left(\mathbf{y}_i - h(M_{t_0 \to t_i}(\mathbf{x}_0))\right)$$

- Practical methods for minimizing J were shown in yesterday's lectures.

Recap of 4DVar: why do any different?

Advantages

- Gaussian and near-linear assumption makes this an efficient algorithm.
- Minimisation of the cost function is a well posed problem (the B-matrix is designed to be full rank).
- Analysis is consistent with the model.
- Lots of theory and techniques to modify the basic algorithm to make it a pragmatic method for various applications, e.g. incremental 4DVar, preconditioning, control variable transforms, weak constraint 4DVar...
- Met Office and ECMWF both use methods based on 4DVar for their atmospheric assimilation.

Disadvantages

- Gaussian assumption is not always valid.
- Relies on the validity of TL and perfect model assumption. This tends to restrict the length of the assimilation window.
- Development of TL model, $\mathbf{M}$, and adjoint, $\mathbf{M}^T$, is very time consuming and difficult to update as the non-linear model is developed.
- B-matrix is predominately static.

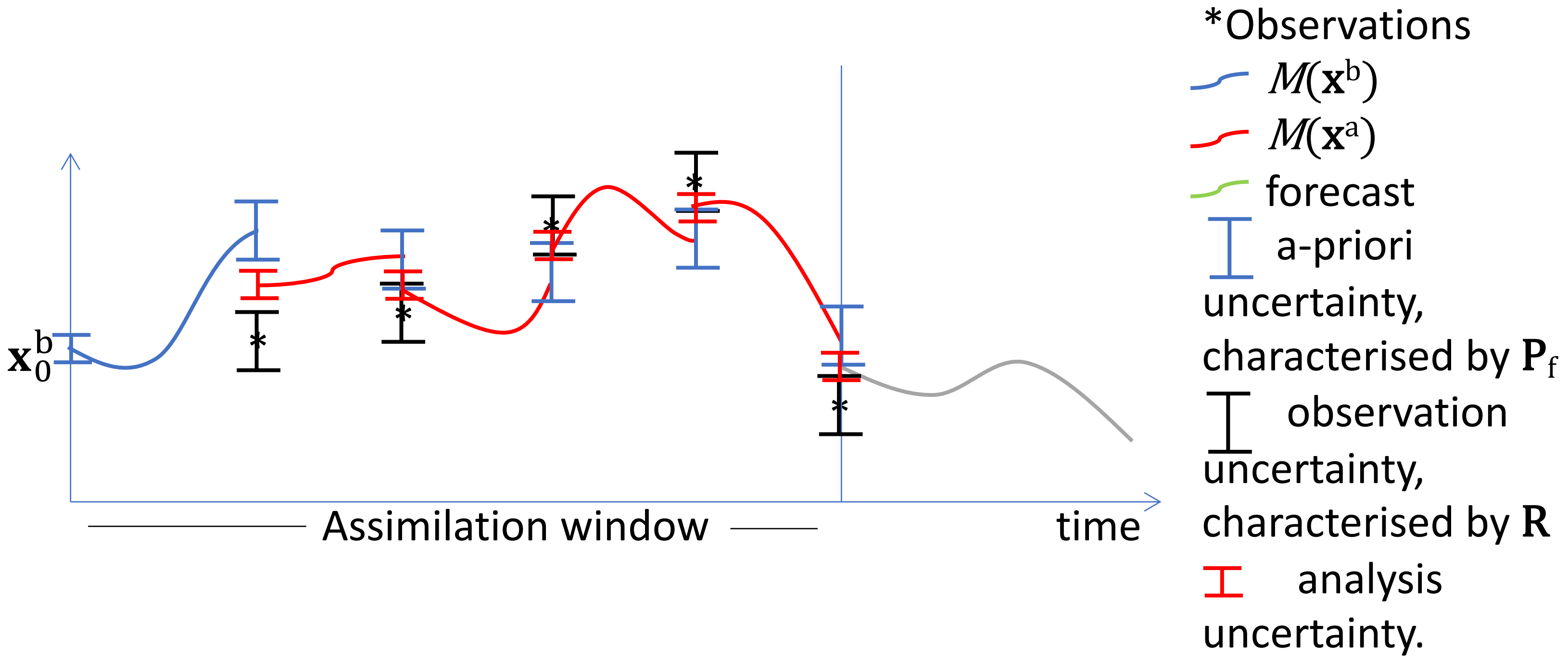
This motivates a different approach…

The Ensemble Kalman Filter

- The ensemble Kalman filter aims to overcome the following disadvantages of 4DVar:
    - The need for a tangent linear and adjoint model
    - The assumption of a perfect model
    - The static B-matrix

- It is still grounded in the Gaussian and near-linear assumptions i.e. only need to find the mean and covariance of the posterior distribution. This helps it to be feasible for large-scale problems

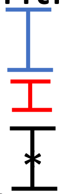
Sequential DA (or filter)

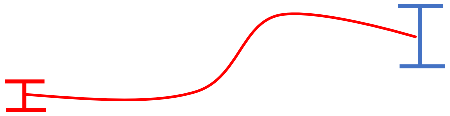


- Instead of assimilating all observations at one time, assimilate them sequentially in time.
- This can be shown to be equivalent to the variational problem, assuming a linear model and all error covariances are treated consistently. Crucially this last point means that the prior error covariances most evolve during the assimilation window.

The Kalman Filter algorithm

Th Kalman filter algorithm consists of two steps:

The update step,  where we update the mean and covariance of the prior at the observation time to become the posterior at the observation time.

The prediction step,  where we update the posterior pdf, described by the mean (assumed to be the analysis) and its covariance, to become the prior at the next observation time.

The Kalman Equations

- The analysis that gives the mean of $p(\mathbf{x}|\mathrm{y})$ assuming the prior and likelihood are Gaussian can be found analytically:

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}\left(\boldsymbol{y} - h\left(\mathbf{x}^{\mathrm{b}}\right)\right),$$

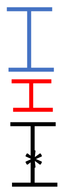
$$\text{where } \mathbf{K} = \mathbf{B}\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right)^{-1}.$$

- If we incorporate the model in the observation operator, then this is implicitly solved for when minimizing the cost function in variational DA.

- Similarly, the analysis error covariance matrix can be given analytically as

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

The update step



- The update step makes use of the same analytical equations solved implicitly in variational data assimilation (see previous slide).
- Update the mean to give the analysis mean at time *k*:

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \mathbf{K}\left(\mathbf{y}_k - h\left(\mathbf{x}_k^{\mathrm{f}}\right)\right), \qquad (1)$$

$$\text{where } \mathbf{K} = \mathbf{P}_k^{\mathrm{f}}\,\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}_k^{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_k\right)^{-1}.$$

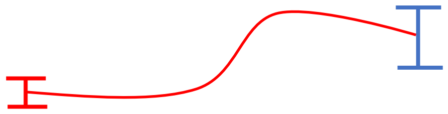
- Update the covariance to give the analysis error covariance at time *k* :

$$\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_k^{\mathrm{f}}$$

Note the change in notation!

f stands for forecast and replaces b, which stood for background in Var.

The prediction step



- The mean of the posterior at time k can be updated to give the mean of the prior at time $k+1$ by using the prediction model:

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

η_k represents uncertainty in the model at time k

- Updating the covariance is trickier.
- The **Extended Kalman filter** (EKF, Grewal and Andrews (2008)) does this using the TL and adjoint of the non-linear model.

$$\mathbf{P}^{\mathrm{f}}_{k} = \mathbf{M}\mathbf{P}^{\mathrm{a}}_{k-1}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}, \qquad \text{where } \mathbf{M} = \nabla_{\mathbf{x}_{k-1}} \mathbf{x}_k = \nabla_{\mathbf{x}_{k-1}} M_{t_{k-1} \rightarrow t_k}(\mathbf{x}_{k-1})$$

Motivation for the ensemble Kalman filter (EnKF)

- The extended Kalman Filter still needs the TL and adjoint model to propagate the covariance matrix.

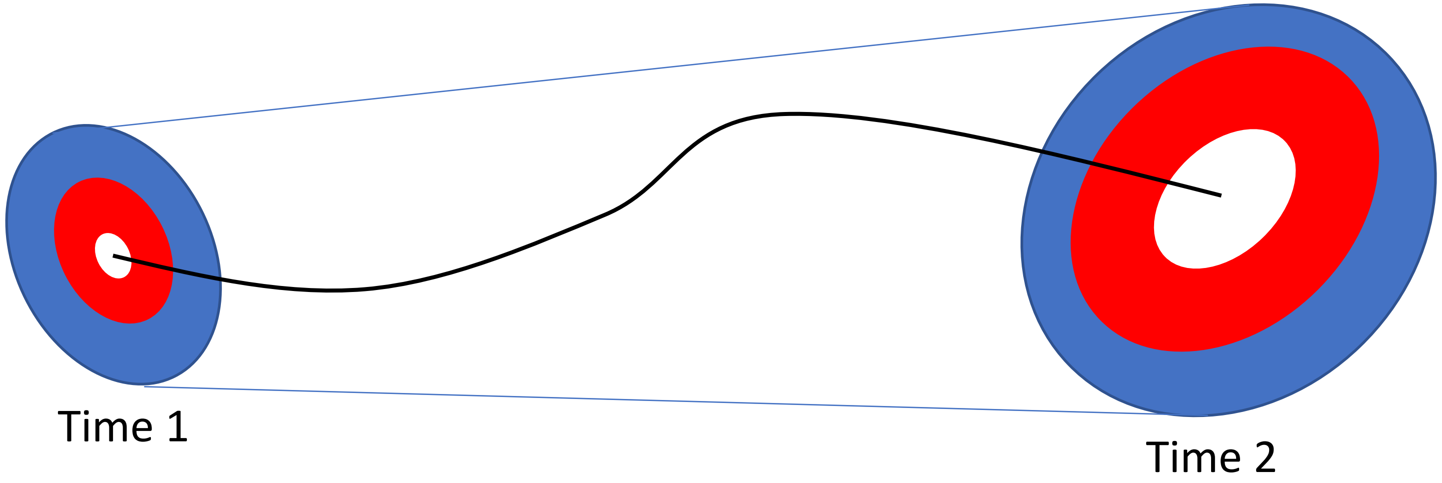
$$\mathbf{P}_k^{\mathrm{f}} = \mathbf{M}\mathbf{P}_{k-1}^{\mathrm{a}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}$$

- Due to the size of this matrix for most environmental applications, the EKF is not feasible in practice.

- An alternative approach to explicitly evolving the full covariance matrix is to instead estimate it using a sample of evolved states (known as the **ensemble**).

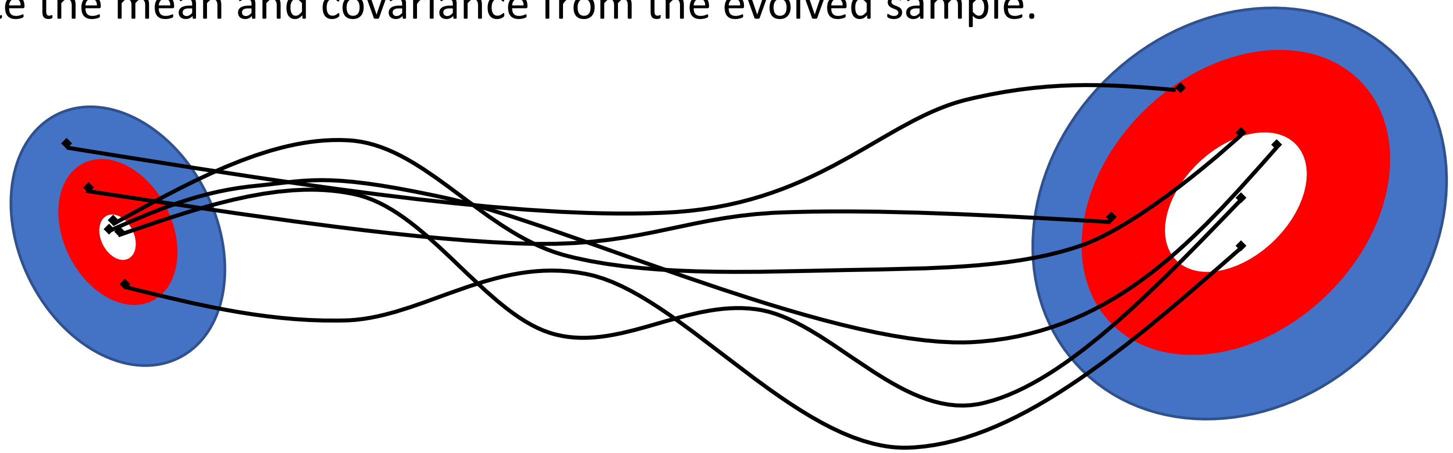
Extended Kalman filter approach

Explicitly evolve the mean and covariances forward in time using M, $\mathbf{M}$ and $\mathbf{M}^T$.



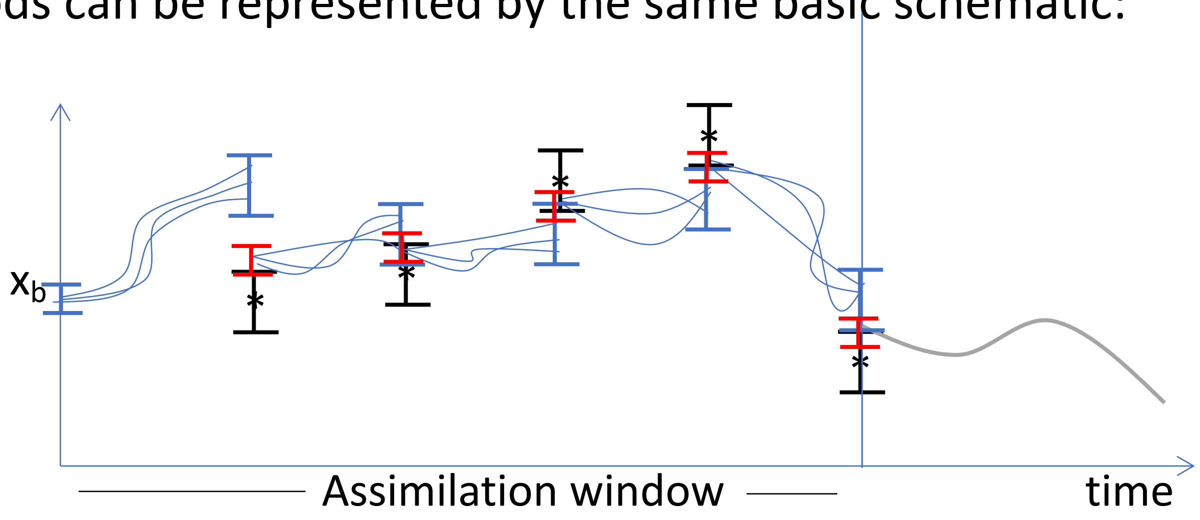
Ensemble Kalman filter approach

Sample from the initial time PDF, evolve each state forward in time using M, then estimate the mean and covariance from the evolved sample.



EnKF algorithms

- The EnKF (Envensen 1994) merges KF theory with Monte Carlo estimation methods.
- There are many many different flavours of EnKF.
- EnKF algorithms can be generalised into two main categories:
  - Stochastic algorithms (e.g. the perturbed observation Kalman filter)
  - Deterministic algorithms (e.g. the ensemble transform Kalman filter)
- All EnKF methods can be represented by the same basic schematic:



To reconstruct the PDFs from the ensemble we still assume the distributions are Gaussian!

The perturbed observation ensemble Kalman Filter

<u>Prediction step</u>

- Evolve each ensemble member i forward using the non-linear model with added noise.

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

  $i = 1, \ldots N$ where N is the ensemble size.

- Reconstruct the ensemble mean

$$\bar{\mathbf{x}}_k^{\mathrm{f}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_k^{(i),\mathrm{f}}$$

- And its covariance

$$\mathbf{P}_k^{\mathrm{f}} = \frac{1}{N-1}\mathbf{X}_k'^{\mathrm{f}}\left(\mathbf{X}_k'^{\mathrm{f}}\right)^T = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_k^{(i),\mathrm{f}} - \bar{\mathbf{x}}_k^{\mathrm{f}}\right)\left(\mathbf{x}_k^{(i),\mathrm{f}} - \bar{\mathbf{x}}_k^{\mathrm{f}}\right)^{\mathrm{T}}$$

Note there is no need to ever explicitly compute $\mathbf{P}^{\mathrm{f}} \in \mathbb{R}^{n\times n}$, just the **perturbation matrix** $\mathbf{X}'^{\mathrm{f}} \in \mathbb{R}^{n\times N}$, which is generally of a smaller dimension, e.g. typically numbers for NWP may be n=10^8, N=10^2.

The perturbed observation ensemble Kalman Filter

Update step

- Update the ensemble using perturbed observations

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

$$\text{where } \boldsymbol{\epsilon}_{\mathrm{y}} \sim N(\mathbf{0}, \mathbf{R})$$

$$\text{and } \mathbf{K}_k = \mathbf{P}_k^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}_k^{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

P^f derived from the ensemble

$$\boxed{\frac{1}{N-1}\mathbf{X}_k^{\prime\mathrm{f}}\left(\mathbf{X}_k^{\prime\mathrm{f}}\right)^{\mathrm{T}}}$$

- It is necessary to perturb the observations for the variance of the ensemble after the update step to correctly represent the uncertainty in the analysis given by

$$\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\,\mathbf{P}_k^{\mathrm{f}}$$

- This introduces additional sampling noise, which motivates the development of square-root or deterministic forms of the EnKF that do not need to perturb the observations.

Ensemble Square Root Filter

- The idea of ESRF is to create an updated ensemble with covariance consistent with $\mathbf{P}_k^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\,\mathbf{P}_k^{\mathrm{f}}$ without the need for perturbing the observations.
- Recall that the ensemble covariance matrix is given by

$$\mathbf{P}_k^{\mathrm{f}} = \frac{1}{N-1}\mathbf{X}_k'^{\mathrm{f}}\left(\mathbf{X}_k'^{\mathrm{f}}\right)^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_k^{(i),\mathrm{f}} - \bar{\mathbf{x}}_k^{\mathrm{f}}\right)\left(\mathbf{x}_k^{(i),\mathrm{f}} - \bar{\mathbf{x}}_k^{\mathrm{f}}\right)^{\mathrm{T}}$$

- We can write a similar expression for the analysis error covariance matrix in terms of the analysis perturbations

$$\mathbf{P}_k^{\mathrm{a}} = \frac{1}{N-1}\mathbf{X}_k'^{\mathrm{a}}(\mathbf{X}_k'^{\mathrm{a}})^{\mathrm{T}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_k^{(i),\mathrm{a}} - \bar{\mathbf{x}}_k^{\mathrm{a}}\right)\left(\mathbf{x}_k^{(i),\mathrm{a}} - \bar{\mathbf{x}}_k^{\mathrm{a}}\right)^{\mathrm{T}}$$

- Instead of updating each ensemble member separately, as in the perturbed observation KF, the ESRF generates the new ensemble simultaneously by updating $\bar{\mathbf{x}}_k^{\mathrm{f}}$ and $\mathbf{X}'^{\mathrm{f}}$.

Ensemble Square Root Filter

Filtering step

- Update ensemble mean

$$\bar{\mathbf{x}}_k^{\mathrm{a}} = \bar{\mathbf{x}}_k^{\mathrm{f}} + \mathbf{K}_k(\boldsymbol{y}_k - \bar{\mathbf{y}}_k^{\mathrm{f}}),$$

$$\text{where } \mathbf{K}_k = \mathbf{X}_k'^{\mathrm{f}}\left(\mathbf{Y}_k'^{\mathrm{f}}\right)^{\mathrm{T}}\left(\mathbf{Y}_k'^{\mathrm{f}}\left(\mathbf{Y}_k'^{\mathrm{f}}\right)^{\mathrm{T}} + (N-1)\mathbf{R}\right)^{-1}$$

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

$$\mathbf{Y}_k'^{\mathrm{f}} = \mathbf{H}\mathbf{X}_k'^{\mathrm{f}} \in \mathbb{R}^{p\times N}$$

- Update perturbation matrix

$$\mathbf{X}_k'^{\mathrm{a}} = \mathbf{X}_k'^{\mathrm{f}}\mathbf{T}_k$$

Need to define the matrix $\mathbf{T}$.

Ensemble Square Root Filter

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

$$\mathbf{P}_k^{\mathrm{a}} = \frac{1}{N-1}\mathbf{X}_k'^{\mathrm{a}}(\mathbf{X}_k'^{\mathrm{a}})^{\mathrm{T}}$$
$$= \frac{1}{N-1}\mathbf{X}_k'^{\mathrm{f}}\mathbf{T}_k\left(\mathbf{X}_k'^{\mathrm{f}}\mathbf{T}_k\right)^{\mathrm{T}}$$
$$\approx (\mathbf{I} - \mathbf{K}_k\mathsf{H})\,\mathbf{P}_k^{\mathrm{f}}$$

- This does not uniquely define $\mathbf{T}$ which is why there are so many different variants of the ESRF, e.g. the Ensemble Adjustment Kalman Filter (Anderson (2001)), and the Ensemble Transform Kalman Filter (Bishop et al. (2001))
- Tippet et al. (2003) review several square root filters and compare their numerical efficiency. Show that although they lead to different ensembles they all span the same subspace.

Ensemble Square Root Filter

An expression for $\mathbf{T}$ can be found by rearranging $(\mathbf{I}-\mathbf{KH})\,\mathbf{P}^{\mathrm{f}}$ using

$$\mathbf{K}=\mathbf{X}'^{\mathrm{f}}\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}\left(\mathbf{Y}'^{\mathrm{f}}\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}+(N-1)\mathbf{R}\right)^{-1} \text{ and } \mathbf{P}^{\mathrm{f}}=\frac{1}{N-1}\mathbf{X}'^{\mathrm{f}}\left(\mathbf{X}'^{\mathrm{f}}\right)^{\mathrm{T}}$$

$$\begin{aligned}
\mathbf{P}^{\mathrm{a}} &= \frac{1}{N-1}\mathbf{X}'^{\mathrm{f}}\,\mathbf{T}\left(\mathbf{X}'^{\mathrm{f}}\,\mathbf{T}\right)^{\mathrm{T}} \\
&\approx (\mathbf{I}-\mathbf{KH})\,\mathbf{P}^{\mathrm{f}} \\
&= \frac{1}{N-1}\left(\mathbf{I}-\mathbf{X}'^{\mathrm{f}}\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}\left(\mathbf{Y}'^{\mathrm{f}}\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}+(N-1)\mathbf{R}\right)^{-1}\mathbf{H}\right)\mathbf{X}'^{\mathrm{f}}\left(\mathbf{X}'^{\mathrm{f}}\right)^{\mathrm{T}} \\
&= \frac{1}{N-1}\mathbf{X}'^{\mathrm{f}}\left(\mathbf{I}-\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}\left(\mathbf{Y}'^{\mathrm{f}}\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}+(N-1)\mathbf{R}\right)^{-1}\mathbf{Y}'^{\mathrm{f}}\right)\left(\mathbf{X}'^{\mathrm{f}}\right)^{\mathrm{T}}
\end{aligned}$$

$$\Rightarrow \mathbf{T}\mathbf{T}^{\mathrm{T}}=\left(\mathbf{I}-\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}\left(\mathbf{Y}'^{\mathrm{f}}\left(\mathbf{Y}'^{\mathrm{f}}\right)^{\mathrm{T}}+(N-1)\mathbf{R}\right)^{-1}\mathbf{Y}'^{\mathrm{f}}\right)$$

The Ensemble Transform Kalman Filter

- First introduced by Bishop et al. (2001), later revised by Wang et al. (2004).
- $\mathbf{T}$ is computed using the Morrison-Woodbury identity to rewrite the previous expression for $\mathbf{T}\mathbf{T}^{\mathrm{T}}$.

$$
\begin{aligned}
\mathbf{T}\mathbf{T}^{\mathrm{T}} = \quad & (\mathbf{I} + \tfrac{1}{N-1}(\mathbf{Y}'^{\mathrm{f}})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}'^{\mathrm{f}})^{-1} \\
& = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{T}})^{-1}
\end{aligned}
$$

$$
\Longrightarrow \quad \mathbf{T} = \mathbf{U}\boldsymbol{\Sigma}^{-1/2}\mathbf{U}^{\mathrm{T}}
$$

- The revision by Wang et al. highlighted that any **T** which satisfies the estimate of the analysis error covariance does not necessarily lead to an unbiased analysis ensemble, see Livings et al. (2008) for conditions that $\mathbf{T}$ must satisfy for the analysis ensemble to be centred on the mean.

Model error

- The ensemble Kalman filter allows for an imperfect model by adding noise at each time step of the model evolution.

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

- The matrix $\mathbf{Q}$ is not explicitly needed in the algorithm, only the effect of the model error in the evolution of the state.
- There have been many different strategies to including model error in the ensemble, based on where you think the source of the error lies. A few examples are
  - **Multiphysics-** different physical models are used in each ensemble member
  - **Stochastic kinetic energy backscatter-** replaces upscale kinetic energy loss due to unresolved processes and numerical integration.
  - **Stochastically perturbed physical tendencies**
  - **Perturbed parameters**
  - Or **combinations** of the above

Summary of the Ensemble Kalman Filter

Advantages

- The a-priori uncertainty is flow-dependent.
- The code can be developed separately from the dynamical model e.g., PDAF or DART system which allows for any model to assimilate observations using ensemble techniques.
- No need to linearise the model, only linear assumption is that statistics remain close to Gaussian.
- Easy to account for model error.
- Easy to parrallelise.

Disadvantages

- Sensitive to ensemble size. Under sampling can lead to filter divergence. Ideas to mitigate this include localisation and inflation (see next EnKF lecture).
- Costly to run multiple versions of a forecast
- Assumes Gaussian statistics, for highly non-linear models this may not be a valid assumption (see Friday's lectures on particle filters)

Further reading

Kalman Filter: •Grewal and Andrews (2008) **Kalman Filtering: Theory and Practice using MATLAB**. *Wiley, New Jersey*. •Kalman (1960**) A new approach to linear filtering and prediction problems**. *J. Basic Engineering*, **82**, 32-45.

Stochastic Ensemble Kalman Filter:•Evensen (1994) **Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics**. *J. Geophys. Res.,* **99(C5)**, 10143-10162.

Determanistic Ensemble Kalman filter:•Anderson (2001) **An ensemble adjustment filter for data assimilation**. *Mon. Weather Rev*., **129**, 2884-2903. •Bishop et al. (2001) **Adaptive sampling with the ensemble transform Kalman filter**. *Mon. Wea. Rev*., **126**, 1719-1724.•Tippet et al. (2003) **Ensemble Square Root Filters**. *Mon. Wea. Rev*., **131**, 1485-1490. •Livings et al. (2008) **Unbiased ensemble square root filters.** *Physica D*. **237**, 1021-1028. •Wang et al. (2004) **Which Is Better, an Ensemble of Positive–Negative Pairs or a Centered Spherical Simplex Ensemble?** *Mon. Wea. Rev*., **132**, 1590-1605

Model error:•Berner et al. (2011) **Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations**, *Mon. Weather Rev*., **139**, 1972–1995.

Reviews:•Bannister (2017) **A review of operational methods of variational and ensemble-variational data assimilation**. Q. J. R. Meteorol. Soc., 143: 607 – 633. •Houtekamer and Zhang (2016) **Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation** . *Mon. Wea. Rev*., **144**, 4489–4532. •Vetra-Carvalho et al. (2018) **State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems.** Tellus A, https://doi.org/10.1080/16000870.2018.1445364