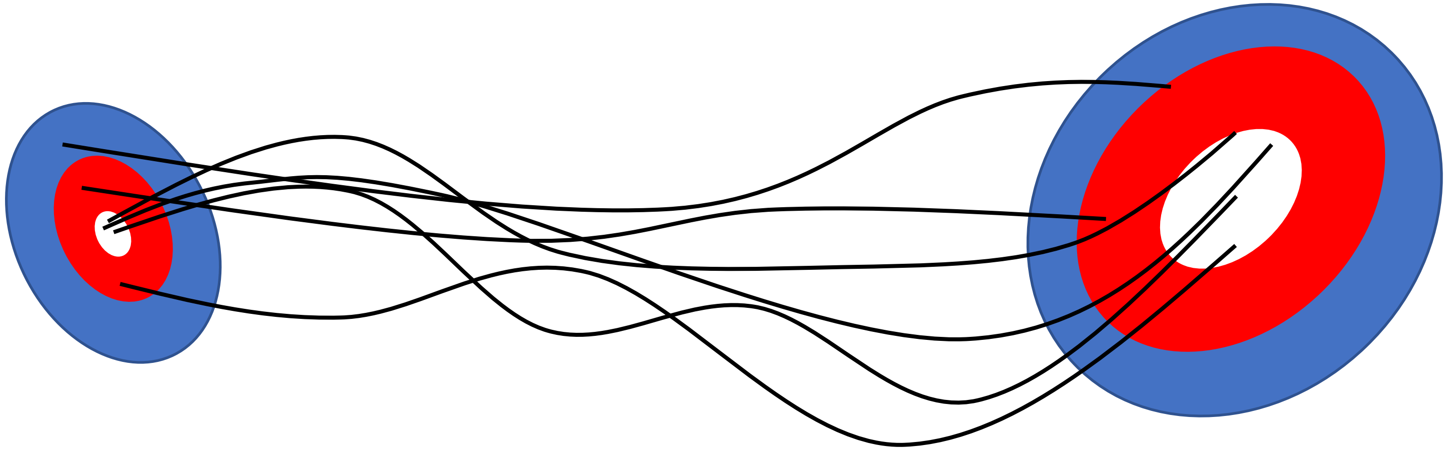
The Ensemble Kalman filter



Part I: Theory

Ivo Pasmans (based on notes by Alison Fowler)

Data-assimilation training course. 7^{th}-10^{th} May, 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:

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

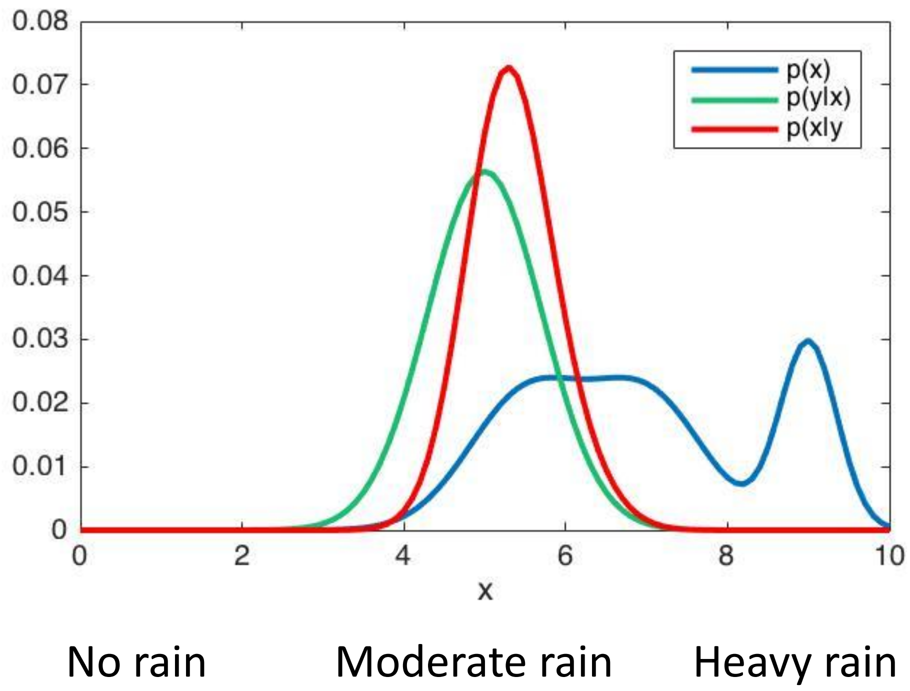
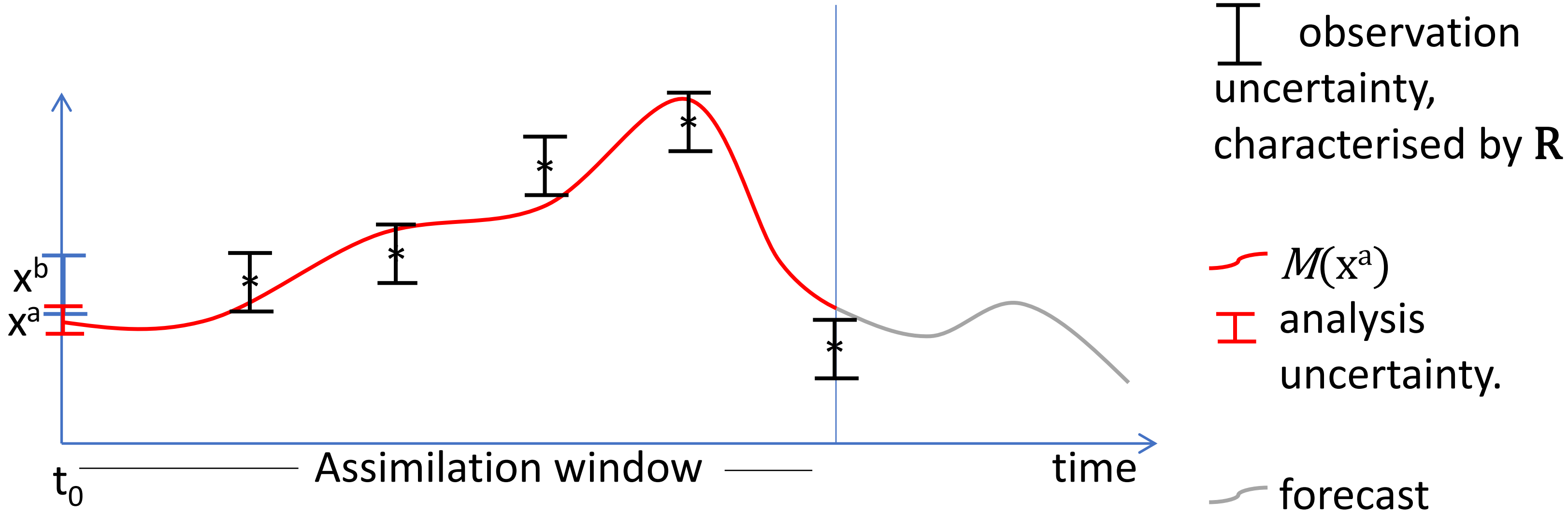


Figure: 1D example of Bayes' theorem.

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 4DVar



- 4DVar aims to find the most likely state at time t_0, given an initial estimate, $\mathbf{x}_0^{\mathrm{b}}$, 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.

$$J(\mathbf{x}_0) = \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \mathbf{B}^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)$$

$$+ \sum_{i=1}^{p} \left(\mathbf{y}_i - H(M_{t_1 \leftarrow t_0}(\mathbf{x}_0))\right)^T \mathbf{R}_i^{-1} \left(\mathbf{y}_i - H(M_{t_1 \leftarrow t_0}(\mathbf{x}_0))\right)$$

- 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).
- 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...
- Analysis is consistent with the model.
- Can use a lot of observations to generate the correction.
- Met Office and ECMWF both use methods based on 4DVar for their atmospheric assimilation.

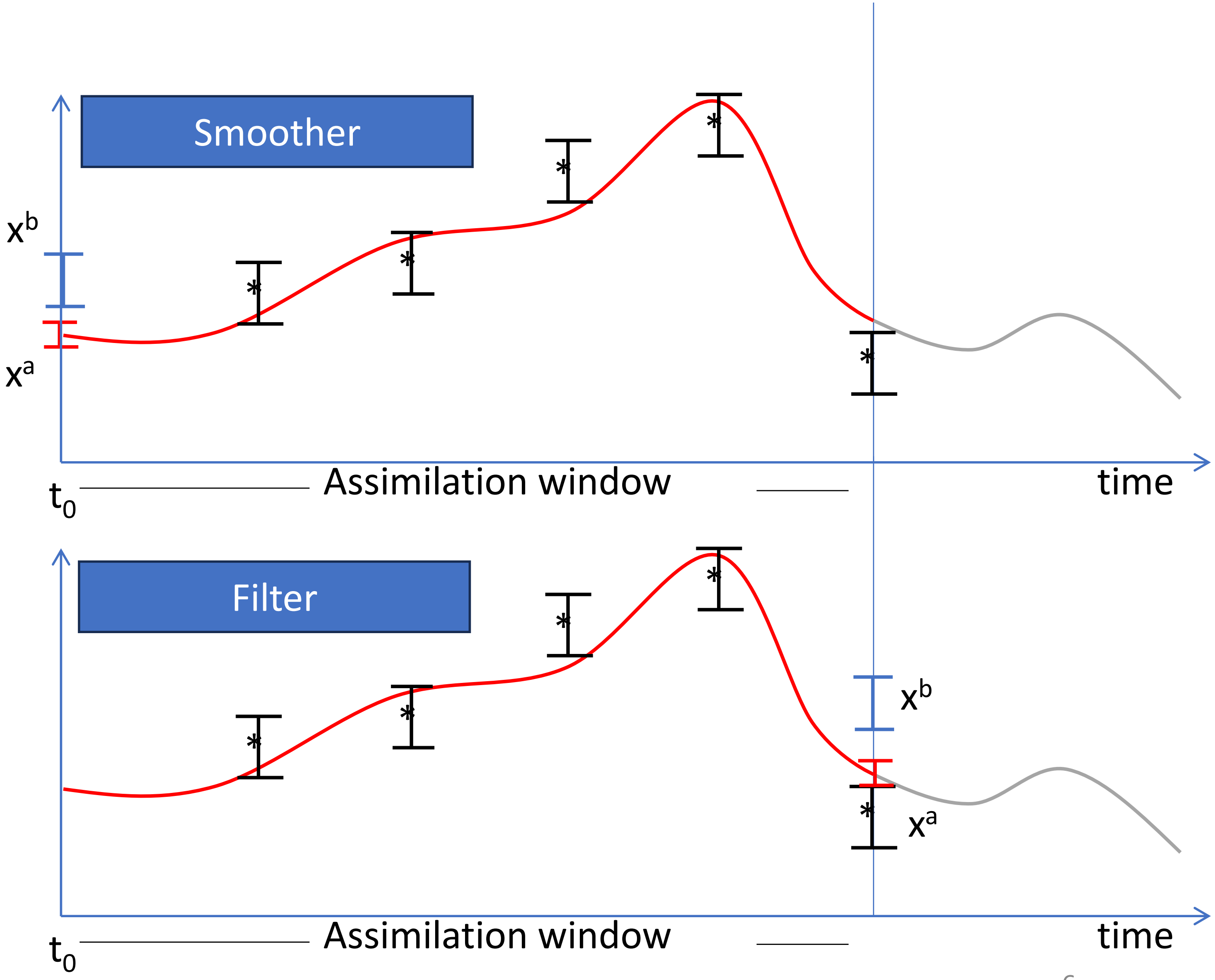
Disadvantages

- 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, **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.
- Uncertainty quantification is cumbersome.
- Requires duplication model run (background and analysis).

This motivates a different approach…

Smoother vs Filter

- Smoother:
  - Non-causal: past, present and future observations.
  - E.g. find $p(\mathbf{x}_0|\mathbf{y}_{0:T})$
- Filter:
  - Causal: past and present observations.
  - E.g. find $p(\mathbf{x}_T|\mathbf{y}_{0:T})$



Sequential DA

- Goal: find $p(\mathbf{x}_T|\mathbf{y}_{:T})$
- Bayes: $\boxed{p(\mathbf{x}_T|\mathbf{y}_{:T})}$

$$
\begin{aligned}
&\sim p(\mathbf{y}_T|\mathbf{x}_T,\mathbf{y}_{:T-1})p(\mathbf{x}_T|\mathbf{y}_{:T-1}) = p(\mathbf{y}_T|\mathbf{x}_T)p(\mathbf{x}_T|\mathbf{y}_{:T-1}) \\
&= p(\mathbf{y}_T|\mathbf{x}_T)\int p(\mathbf{x}_T|\mathbf{x}_{T-1})\boxed{p(\mathbf{x}_{T-1}|\mathbf{y}_{:T-1})}\,\mathbf{dx}_{T-1} \\
&= \ldots\int\int\ldots p(\mathbf{x}_{T-2}|\mathbf{y}_{:T-2})\,\mathbf{dx}_{T-2}\,\mathbf{dx}_{T-2}\,\mathbf{dx}_{T-1}
\end{aligned}
$$

Kalman Filter

- Analytic solution available for linear model with Gaussian errors.

$$p(\mathbf{x}_0) \sim e^{-\frac{1}{2}(\mathbf{x}_0-\mu_0)^{\mathrm{T}}(\mathbf{P}_0^f)^{-1}(\mathbf{x}_0-\mu_0)}$$ Initial error

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) \sim e^{-\frac{1}{2}(\mathbf{x}_t-\mathbf{M}_{t\leftarrow t-1}\mathbf{x}_{t-1})^{\mathrm{T}}(\mathbf{Q}_t)^{-1}(\mathbf{x}_t-\mathbf{M}_{t\leftarrow t-1}\mathbf{x}_{t-1})}$$ Model error

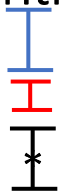
$$p(\mathbf{y}_t|\mathbf{x}_t) \sim e^{-\frac{1}{2}(\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t)^{\mathrm{T}}(\mathbf{R}_t)^{-1}(\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t)}$$ Observation error

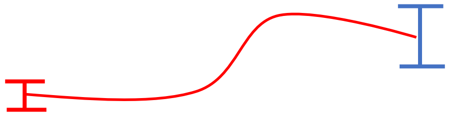
- Solution:

$$p(\mathbf{x}_t|\mathbf{y}_{:t}) \sim e^{-\frac{1}{2}(\mathbf{x}_t-\mu_t^a)^{\mathrm{T}}(\mathbf{P}_t^a)^{-1}(\mathbf{x}_t-\mu_t^a)}$$

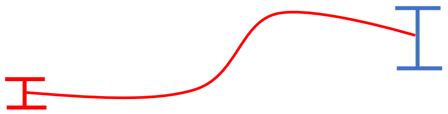
The Kalman Filter Algorithm

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.

Prediction Step



- Move probability distribution forward in time.

$$p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{:t}) \sim e^{-\frac{1}{2}(\mathbf{x}_{t+1}-\mathbf{M}_{t+1\leftarrow t}\mathbf{x}_t)^{\mathrm{T}}(\mathbf{Q}_t)^{-1}(\mathbf{x}_{t+1}-\mathbf{M}_{t+1\leftarrow t}\mathbf{x}_t)-\frac{1}{2}(\mathbf{x}_t-\mu_t)^{\mathrm{T}}(\mathbf{P}_t^a)^{-1}(\mathbf{x}_t-\mu_t)}$$

- so

$$\mathbf{x}_{t+1} = \mathbf{M}_{t+1\leftarrow t}(\mu_t + (\mathbf{P}_t^a)^{\frac{1}{2}}\epsilon) + \mathbf{Q}_t^{\frac{1}{2}}\eta_t$$

with ϵ, η drawn from a standard normal distribution

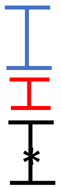
- and consequently

$$\int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{:t})\,\mathbf{dx}_t = p(\mathbf{x}_{t+1}|\mathbf{y}_{:t}) \sim e^{-\frac{1}{2}(\mathbf{x}_{t+1}-\mu_{t+1}^f)^{\mathrm{T}}(\mathbf{P}_{t+1}^f)^{-1}(\mathbf{x}_{t+1}-\mu_{t+1}^f)}$$

$$\mu_{t+1}^f = \mathbf{M}_{t+1\leftarrow t}\mu_t$$

$$\mathbf{P}_{t+1}^f = \mathbf{M}_{t+1\leftarrow t}\mathbf{P}_t^a\mathbf{M}_{t+1\leftarrow t}^T + \mathbf{Q}_t$$

Update Step



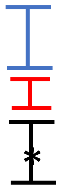
- Incorporate information from observations.

$$
\begin{aligned}
p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})p(\mathbf{x}_{t+1}|\mathbf{y}_{:t}) &\sim e^{-\frac{1}{2}(\mathbf{y}_{t+1}-\mathbf{H}_{t+1}\mathbf{x}_{t+1})^{\mathrm{T}}(\mathbf{R}_{t+1})^{-1}(\mathbf{y}_{t+1}-\mathbf{H}_{t+1}\mathbf{x}_{t+1})} e^{-\frac{1}{2}(\mathbf{x}_{t+1}-\mu^f_{t+1})^{\mathrm{T}}(\mathbf{P}^f_{t+1})^{-1}(\mathbf{x}_{t+1}-\mu^f_{t+1})} \\
&\sim \sim e^{-\frac{1}{2}(\mathbf{x}_{t+1}-\mu^a_{t+1})^{\mathrm{T}}(\mathbf{P}^a_{t+1})^{-1}(\mathbf{x}_{t+1}-\mu^a_{t+1})} e^{C} \sim e^{-\frac{1}{2}(\mathbf{x}_{t+1}-\mu^a_{t+1})^{\mathrm{T}}(\mathbf{P}^a_{t+1})^{-1}(\mathbf{x}_{t+1}-\mu^a_{t+1})}
\end{aligned}
$$

- 2nd order terms

$$
\begin{aligned}
(\mathbf{P}^a_{t+1})^{-1} &= (\mathbf{P}^f_{t+1})^{-1} + \mathbf{H}^{\mathrm{T}}_{t+1}\mathbf{R}^{-1}_{t+1}\mathbf{H}_{t+1} \\
\mathbf{P}^a_{t+1} &= \mathbf{P}^f_{t+1} - \boxed{\mathbf{P}^f_{t+1}\mathbf{H}^{\mathrm{T}}_{t+1}(\mathbf{H}_{t+1}\mathbf{P}^f_{t+1}\mathbf{H}^{\mathrm{T}}_{t+1} + \mathbf{R}_{t+1})^{-1}}\mathbf{H}_{t+1}\mathbf{P}^f_{t+1} \ (\mathrm{Sherman-Morrison-Woodbury}) \\
&= (\mathbf{I} - \boxed{\mathbf{K}_{t+1}}\mathbf{H}_{t+1})\mathbf{P}^f_{t+1}
\end{aligned}
$$

Update Step



- Incorporate information from observations.

$$\begin{aligned} p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{:t-1}) &\sim e^{-\frac{1}{2}(\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t)^{\mathrm{T}}(\mathbf{R}_t)^{-1}(\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t)} e^{-\frac{1}{2}(\mathbf{x}_t-\mu_t^f)^{\mathrm{T}}(\mathbf{P}_t^f)^{-1}(\mathbf{x}_t-\mu_t^f)} \\ &\sim e^{-\frac{1}{2}(\mathbf{x}_t-\mu_t^a)^{\mathrm{T}}(\mathbf{P}_t^a)^{-1}(\mathbf{x}_t-\mu_t^a)} \end{aligned}$$

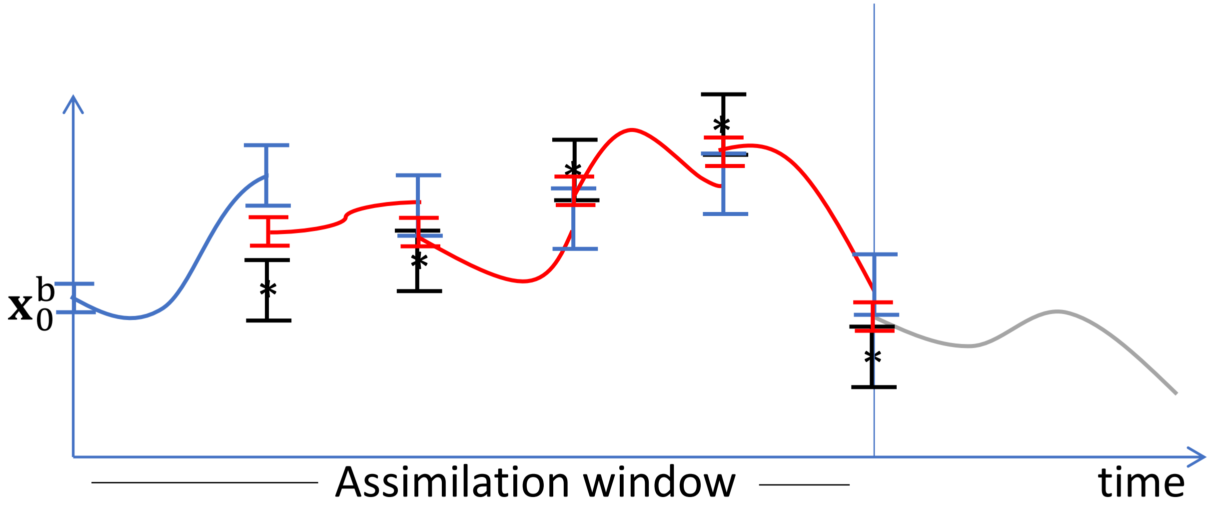
- 2nd order terms

$$\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t)^{-1}\mathbf{H}_t\mathbf{P}_t^f$$

- 1st order terms

$$\begin{aligned} (\mathbf{P}_t^a)^{-1}\mu_t^a &= \mathbf{H}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\mathbf{y}_t + (\mathbf{P}_t^f)^{-1}\mu_t^f \\ \mu_t^a &= (\mathbf{I} - \mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t)^{-1}\mathbf{H}_t)\mu_t^f \\ &- \mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t)^{-1}\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}(\mathbf{R}_t)^{-1}\mathbf{y}_t \\ &+ \mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_T)^{-1}(\mathbf{H}_T\mathbf{P}_T^f\mathbf{H}_T^{\mathrm{T}} + \mathbf{R}_t)(\mathbf{R}_t)^{-1}\mathbf{y}_t \\ &= \mu_t^f + \mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t)^{-1}\boxed{(\mathbf{y}_t - \mathbf{H}_t\mu_t^f)} \quad \text{Innovation vector} \\ &= \mu_t^f + \mathbf{K}_t\boxed{\mathbf{d}_t} \end{aligned}$$

Sequential DA (or filter)



*Observations

$M(\mathbf{x}^b)$

$M(\mathbf{x}^a)$

forecast

a-priori uncertainty, characterised by $\mathbf{P}_f$

observation uncertainty, characterised by $\mathbf{R}$

analysis uncertainty.

- Instead of assimilating all observations at one time, assimilate them sequentially in time.
- It can be shown that solution for $\mathbf{x}_T$ 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.

Extended Kalman Filter (EKF)

- Kalman filter assumptions are not satisfied for nonlinear models.
- Possible solution: Taylor model around a background $M_{t+1\leftarrow t}(\mathbf{x}_t) \approx \mathbf{x}_t^b + \mathbf{M}_{t+1\leftarrow t}(\mathbf{x}_t - \mathbf{x}_t^b)$
- The extended Kalman Filter still needs the TL and adjoint model to propagate the covariance matrix.

$$\mathbf{P}_{t+1}^{\mathrm{f}} = \mathbf{M}\mathbf{P}_t^{\mathrm{a}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}$$

- 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**).

The Ensemble Kalman Filter

- Assume that distribution is Gaussian with μ and $\mathbf{P}$ estimated from an ensemble $\{\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \ldots, \mathbf{x}_t^{(N)}\}$

$$\mu_t = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_t^{(n)}$$

$$\mathbf{P}_t = \frac{1}{N-1}\sum_{n=1}^{N} (\mathbf{x}_t^{(n)} - \mu_t)(\mathbf{x}_t^{(n)} - \mu_t)^{\mathrm{T}}$$

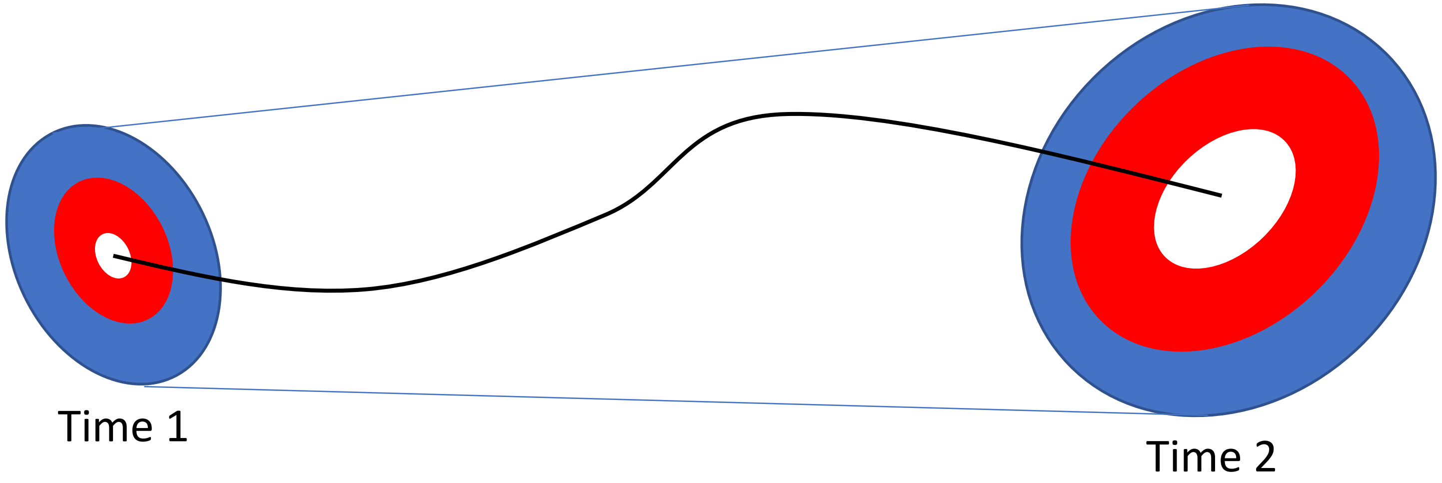
- The ensemble Kalman filter aims to overcome the following disadvantages of 4DVar:

	4DVAR	Kalman	ETK	EnKF
Tangent/adjoint model	Yes	**No**	Yes	**No**
Model error	No (strong)	**Yes**	**Yes**	**Yes**
Static **B/P**	Yes	**No**	**No**	**No**
Output	Solution	**Probability**	**Probability**	**Probability**
Non-linear runs	>=2	**1**	2	**1**

- 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

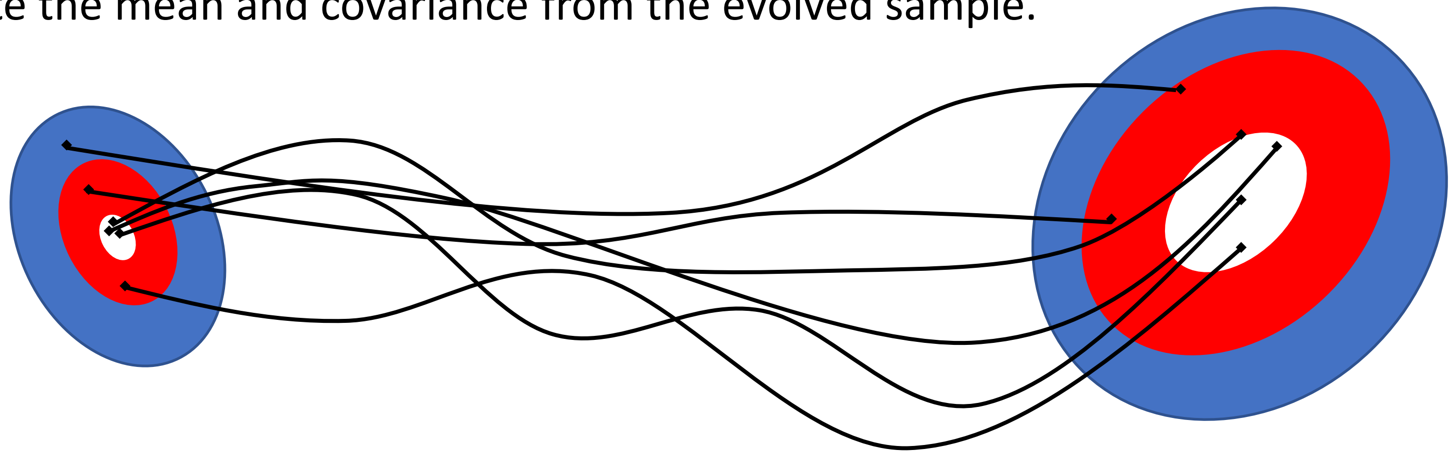
Extended Kalman filter approach

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



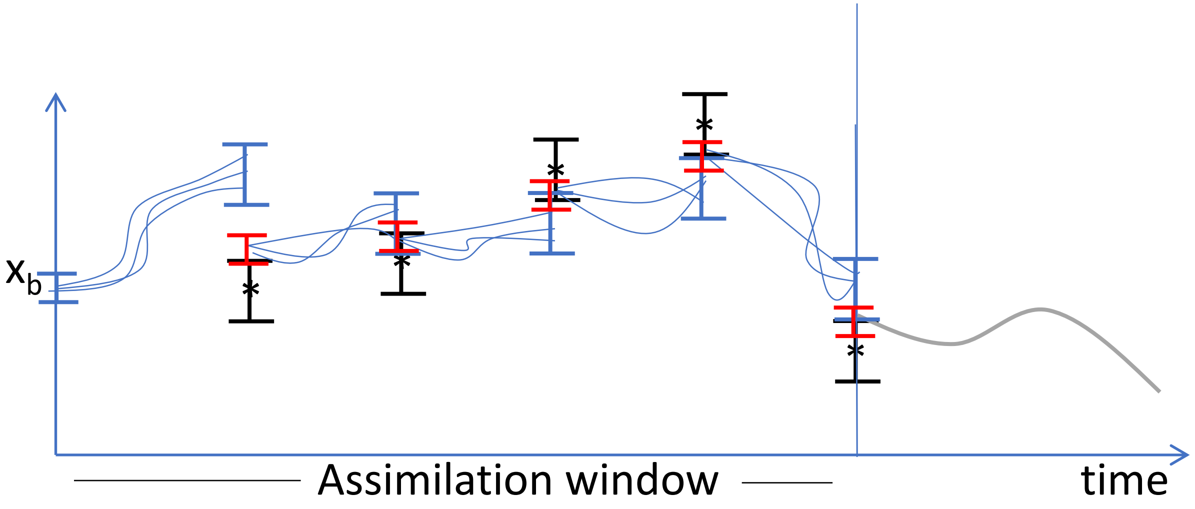
Ensemble Kalman filter approach

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!

EnKF: classical prediction step

- Prediction step is the same for both EnKF families.
- Evolve each ensemble member n forward using the non-linear model with added noise.

$$\mathbf{x}_t^{f,(n)} = M_{t \leftarrow t-1}(\mathbf{x}_{t-1}^{a,(n)}) + \eta_{t-1}^{(n)}$$

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

- Reconstruct the ensemble mean

$$\mu_t^f = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_t^{f,(n)}$$

- And its covariance

$$\mathbf{P}_t^f = \frac{1}{N-1} \sum_{n=1}^{N} (\mathbf{x}_t^{f,(n)} - \mu_t^f)(\mathbf{x}_t^{f,(n)} - \mu_t^f)^{\mathrm{T}} = \sum_{n=1}^{N} \mathbf{x}'^{f,(n)} (\mathbf{x}'^{f,(n)})^{\mathrm{T}} = \frac{1}{N-1} \mathbf{X}'^f (\mathbf{X}'^f)^{\mathrm{T}}$$

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

$$\mathbf{P}_t^f \mathbf{q} = \frac{1}{N-1} \mathbf{X}'^f \left(\mathbf{q}^{\mathrm{T}} \mathbf{X}'^f\right)^{\mathrm{T}}$$

Model error

- Different ways to generate the model errors are in use.
- To match the classical Kalman filter the model error should be drawn from a Gaussian:

$$\eta_t^{(n)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

- 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**
  - **Perturb boundary conditions**
  - Or **combinations** of the above

Stochastic EnKF: update step

- Naïve approach:

$$
\begin{aligned}
\mathbf{x}^{a,(n)} &= \mathbf{x}^{f,(n)} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^{f,(n)}) \\
\mathbf{K} &= \mathbf{P}_N^f \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{P}_N^f \mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \approx \mathbf{X}'^f \mathbf{Y}^{\mathrm{T}} (\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1} \\
\mathbf{Y} &= \mathbf{H}\mathbf{X}'^f
\end{aligned}
$$

$$
\mathbf{X}' = [\mathbf{x}^{(1)} \dots \mathbf{x}^{(N)}] - \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}^{(n)} \mathbf{1}_N^{\mathrm{T}}
$$

- Resulting analysis covariance

$$
\begin{aligned}
\mathbf{X}'^a &= \mathbf{X}'^f - \mathbf{K}\mathbf{H}\mathbf{X}'^f \\
\mathbf{P}_N^a &= \frac{1}{N-1} \mathbf{X}'^a (\mathbf{X}'^a)^{\mathrm{T}} = \frac{1}{N-1} (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{X}'^f (\mathbf{X}'^f)^{\mathrm{T}} (\mathbf{I} - \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}}) \\
&= \boxed{(\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{P}_N^f} - (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{P}_N^f \mathbf{H}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}}
\end{aligned}
$$

Stochastic EnKF: update step

- Correct approach:

$$\mathbf{x}^{a,(n)} = \mathbf{x}^{f,(n)} + \mathbf{K}(\mathbf{y} + \boxed{\xi^{(n)}} - \mathbf{H}\mathbf{x}^{f,(n)}) \text{ with } \xi^{(n)} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

$$\mathbf{K} = \mathbf{P}^f_N \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{P}^f_N \mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \approx \mathbf{X}'^f \mathbf{Y}^{\mathrm{T}} (\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1}$$

$$\mathbf{Y} = \mathbf{H}\mathbf{X}'^f$$

- Resulting analysis covariance

$$\mathbf{X}'^a = \mathbf{X}'^f - \mathbf{K}\mathbf{H}\mathbf{X}'^f + \mathbf{K}\Xi'$$

$$\begin{aligned}
\mathbf{P}^a_N &= \frac{1}{N-1}\mathbf{X}'^a(\mathbf{X}'^a)^{\mathrm{T}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N - (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N \mathbf{H}^{\mathrm{T}}\mathbf{K}^T + \frac{1}{N-1}\mathbf{K}\Xi'\Xi'^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} \\
&- \frac{1}{N-1}(\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{X}'\Xi'^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} - \frac{1}{N-1}\mathbf{K}\Xi'(\mathbf{X}'^f)^{\mathrm{T}}(\mathbf{I} - \mathbf{K}^T\mathbf{H}^T) \\
&\overset{N\to\infty}{=} (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N - (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N \mathbf{H}^{\mathrm{T}}\mathbf{K}^T + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}} \\
&= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N - \mathbf{P}^f_N \mathbf{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} + \mathbf{K}(\mathbf{H}\mathbf{P}^f_N\mathbf{H}^{\mathrm{T}} + \mathbf{R})\mathbf{K}^{\mathrm{T}} \\
&= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N - \mathbf{P}^f_N \mathbf{H}^{\mathrm{T}}(\mathbf{I} - (\mathbf{H}\mathbf{P}^f_N\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{H}\mathbf{P}^f_N\mathbf{H}^{\mathrm{T}} + \mathbf{R}))\mathbf{K}^T = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f_N
\end{aligned}$$

Ensemble Square Root Filter: update step

- The idea of ESRF is to create an updated ensemble with covariance consistent with

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}^f$$

- Using ensemble approximations for the covariances this gives

$$\begin{aligned}
(N-1)\mathbf{P}_N^a &= \mathbf{X}'^a(\mathbf{X}'^a)^{\mathrm{T}} = \mathbf{X}'^f(\mathbf{X}'^f)^{\mathrm{T}} - \mathbf{X}'^f\mathbf{Y}^{\mathrm{T}}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1}\mathbf{Y}(\mathbf{X}'^f)^{\mathrm{T}} \\
&= \mathbf{X}^f\left[\mathbf{I} - \mathbf{Y}^{\mathrm{T}}(\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1}\mathbf{Y}\right](\mathbf{X}'^f)^{\mathrm{T}} \\
&= \mathbf{X}'^f[\frac{1}{N-1}\mathbf{Y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y} + \mathbf{I}]^{-1}(\mathbf{X}'^f)^{\mathrm{T}} \quad \text{(Shermann-Morrison-Woodley)} \\
\mathbf{Y} &= \mathbf{H}\mathbf{X}'^f \\
\mathbf{X}' &= \left[\mathbf{x}^{(1)} \ldots \mathbf{x}^{(N)}\right] - \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}^{(n)}\mathbf{1}_N^{\mathrm{T}}
\end{aligned}$$

Ensemble Transform Kalman Filter: update step

- Idea: pick analysis anomalies as linear combinations forecast anomalies, i.e. $\mathbf{X}'^a = \mathbf{X}'^f \mathbf{T}$
- From $\mathbf{X}'^a (\mathbf{X}'^a)^{\mathrm{T}} = \mathbf{X}^f [\frac{1}{N-1} \mathbf{Y}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y} + \mathbf{I}]^{-1} (\mathbf{X}'^f)^{\mathrm{T}}$
  it follows this works if $\mathbf{T}\mathbf{T}^{\mathrm{T}} = [\frac{1}{N-1} \mathbf{Y}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y} + \mathbf{I}]^{-1}$
- 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.
- Possible solutions if $[\frac{1}{N-1} \mathbf{Y}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y} + \mathbf{I}] = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\mathrm{T}}$
  are
  $$\mathbf{T} = \mathbf{U} \mathbf{\Lambda}^{-\frac{1}{2}}$$
  $$\mathbf{T} = \mathbf{U} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{U}^{\mathrm{T}}$$
- Be aware $\mathbf{X}'^a \mathbf{1}_N = \mathbf{X}'^f \mathbf{T} \mathbf{1}_N = 0$ so $\mathbf{1}_N$ must be an eigenvector of $\mathbf{T}$.

Ensemble Transform Kalman Filter: update step

- Calculated analysis ensemble anomalies around the mean.
- Update analysis mean is straightforward

$$\begin{aligned}\overline{\mathbf{x}}^a &= \overline{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f) = \overline{\mathbf{x}}^f + \mathbf{X}'^f \mathbf{Y}^{\mathrm{T}} (\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1} (\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^f) \\ \overline{\mathbf{x}}^f &= \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}^{f,(n)}\end{aligned}$$

Summary of the Ensemble Kalman Filter

Advantages

- The apriori 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 parallelize.

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.

Further reading

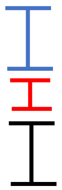
Kalman Filter: •Grewal and Andrews (2008) **Kalman Filtering: Theory and Practice using MATLAB**. *Wiley, New Jersey*. •Kalman (1960**) A new approach to linear filtering and prediction problems**. *J. Basic Engineering*, **82**, 32-45.

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

Bayesian conjugates 

- *p* is Bayesian conjugate prior for $p(\mathbf{y}_t|\mathbf{x}_t)$ if $p(\mathbf{x}_{t+1}|\mathbf{y}_{:t})$ and $p(\mathbf{x}_t|\mathbf{y}_{:t})$ are in the same family.
- Gaussian is a Bayesian conjugate prior for Gaussians, i.e. if $p(\mathbf{y}_t|\mathbf{x}_t)$ and $p(\mathbf{x}_{t+1}|\mathbf{y}_{:t})$ are Gaussians so is $p(\mathbf{x}_{t+1}|\mathbf{y}_{:t+1})$

Ensemble Transform Kalman Filter: update step

- Idea: pick analysis anomalies as linear combinations forecast anomalies, i.e. $\mathbf{X}'^a = \mathbf{X}'^f \mathbf{T}$
- From $\mathbf{X}'^a (\mathbf{X}'^a)^{\mathrm{T}} = \mathbf{X}^f \left[\mathbf{I} - \mathbf{Y}^{\mathrm{T}} (\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1} \mathbf{Y}\right] (\mathbf{X}'^f)^{\mathrm{T}}$
  it follows this works if $\mathbf{T}\mathbf{T}^{\mathrm{T}} = \mathbf{I} - \mathbf{Y}^{\mathrm{T}} (\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + (N-1)\mathbf{R})^{-1} \mathbf{Y}$
- 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 $\mathbf{T}\mathbf{T}^{\mathrm{T}}$.

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

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

- 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.