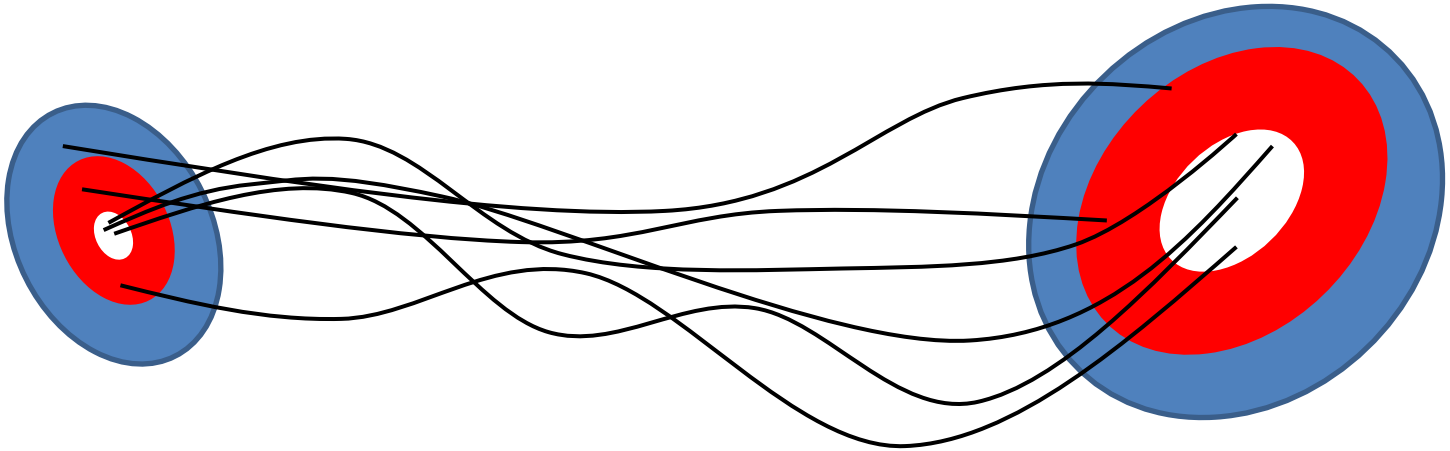
The Ensemble Kalman filter



Part II: Practicalities

Dr Sanita Vetra-Carvalho
(Based on notes by Dr Alison Flower and Dr Ross Bannister)

Data-assimilation training course. 5-8th March 2019, University of Reading

Outlook

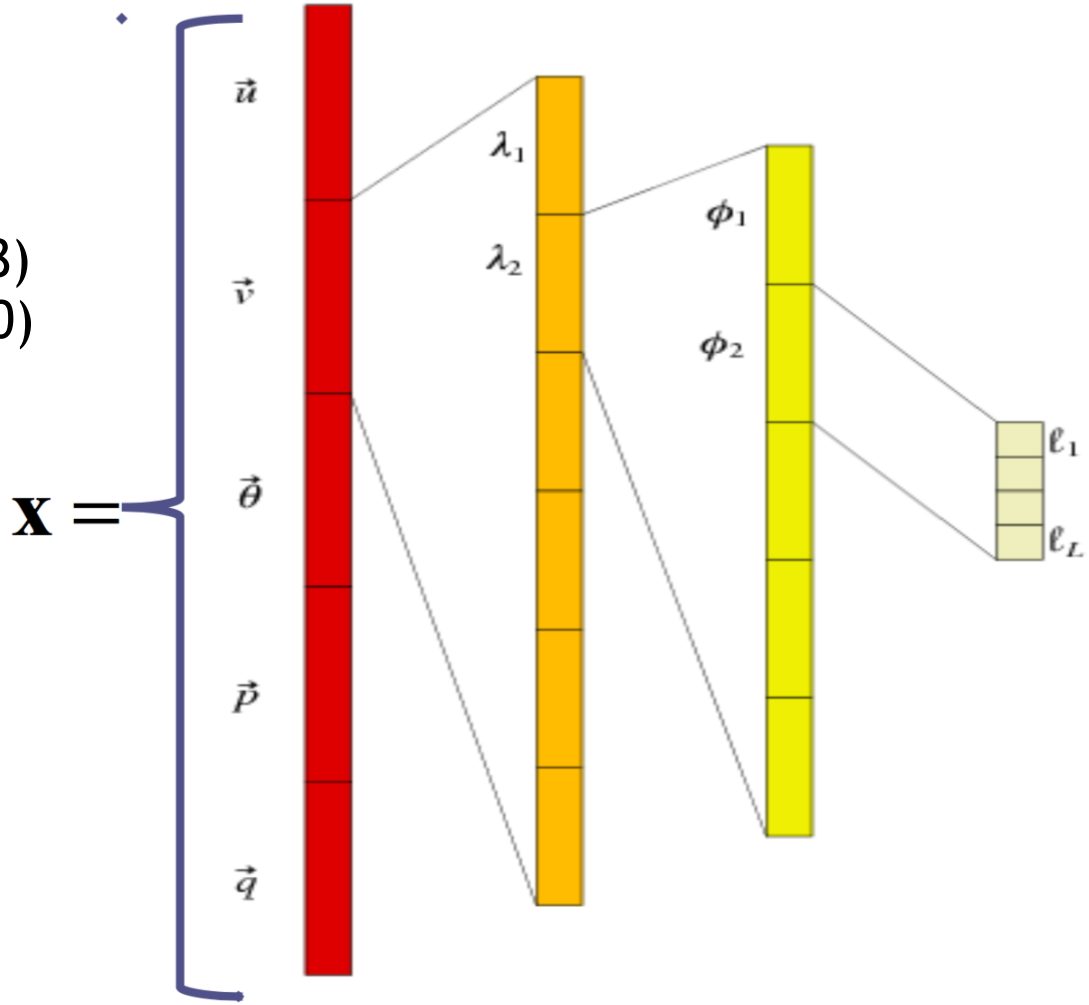
- Issued due to ensemble size
- Possible solutions
  - *Localisation*
  - *Inflation*
  - *Hybrid methods*
- Other issues
  - *Parallelisation*
  - *Nonlinearity*

Issues due to limited ensemble size

Ensemble vs. State size

In NWP applications:
- state dimension O(10^8)
- ensemble size O(10-100)

Can ensemble with e.g. 100 members represent sufficiently well ensemble covariances in an NWP model?



EnKF Recap

- Recall at each analysis time we update the ensemble using ensemble mean and perturbations

$$\overline{\mathbf{x}}^{\mathbf{f}} = \frac{1}{N_e}\sum_{i=1}^{N_e}\mathbf{x}^{\mathbf{f}}_i$$

$$\mathbf{P}^{\mathbf{f}}_e = \frac{1}{N_e - 1}\mathbf{X}'^{\mathbf{f}}\left(\mathbf{X}'^{\mathbf{f}}\right)^{\mathbf{T}} = \frac{1}{N_e - 1}\left(\mathbf{x}^{\mathbf{f}}_i - \overline{\mathbf{x}}^{\mathbf{f}}\right)\left(\mathbf{x}^{\mathbf{f}}_i - \overline{\mathbf{x}}^{\mathbf{f}}\right)^{\mathbf{T}}$$

- N_e is the size of the ensemble (typically 20-100)
- N_x is the size of the state vector (typically 10^9 in NWP)

- Therefore $\mathbf{X}^{\mathbf{f}} \in \mathbb{R}^{N_x \times N_e}$ and $rank(\mathbf{P}^{\mathbf{f}}) \ll N_x$.

Sampling error

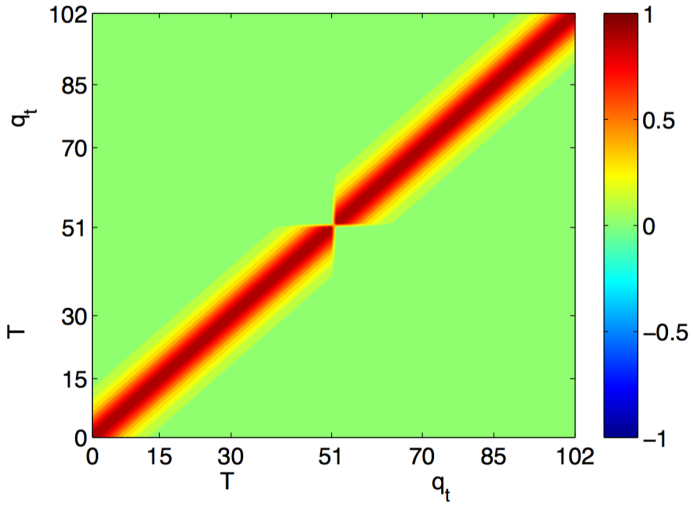
The ensemble Kalman Filter theory assumes that the ensemble is large enough to give an accurate estimate of the sample mean and covariance.

The success of the EnKF methods is highly dependent on the size of the ensemble being adequate for the system we apply these methods to.

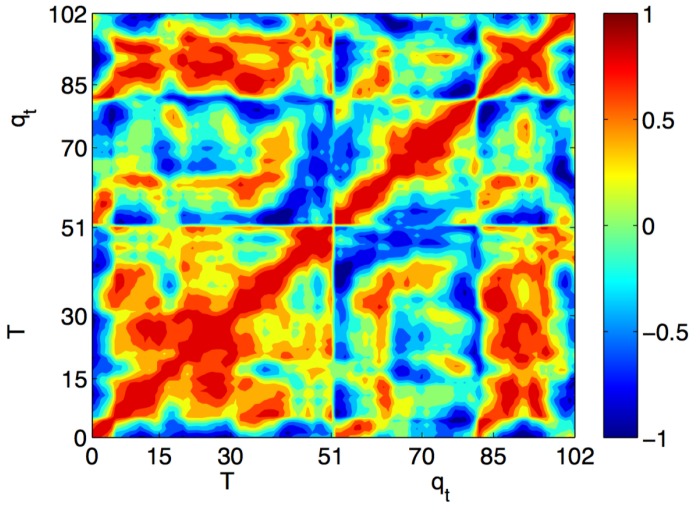
For large scale problems, where $N_e \ll N_x$, ensemble undersampling can cause major problems in EnDA methods:

- underestimated ensemble variance,
- filter divergence,
- errors in estimated correlations, in particular spurious long-range correlations.

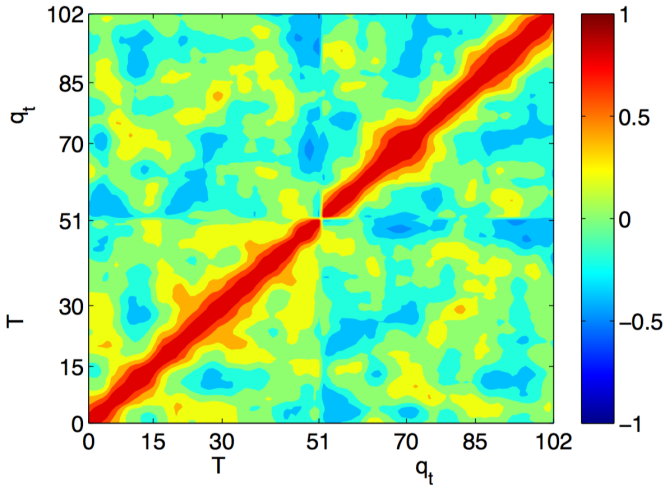
Sampling error



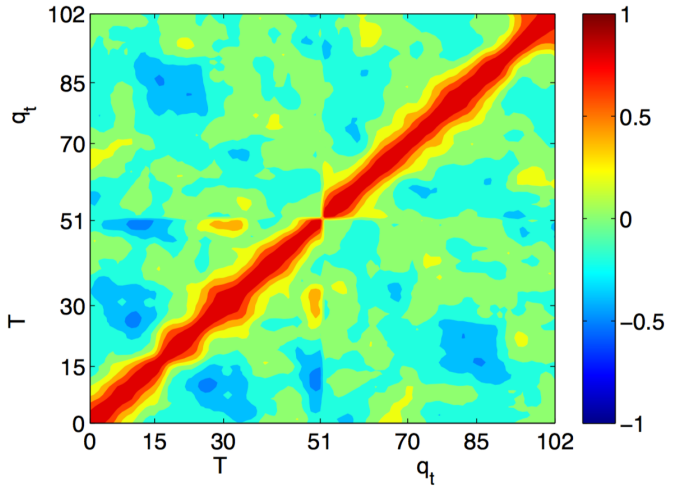
(a) Original



(b) $N = 6$



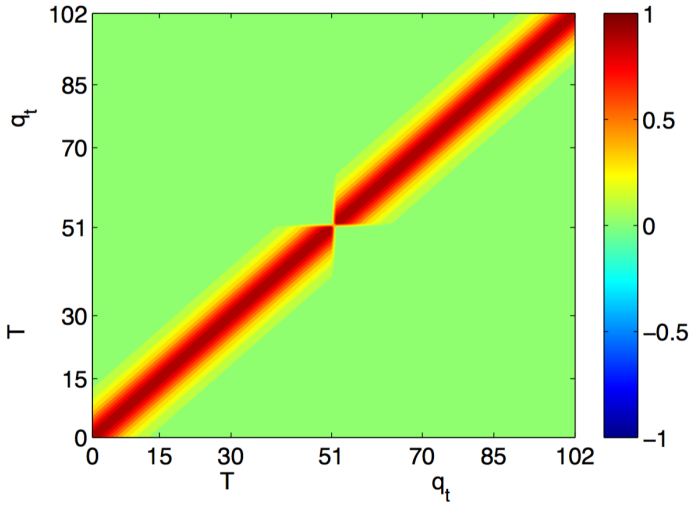
(c) $N = 30$



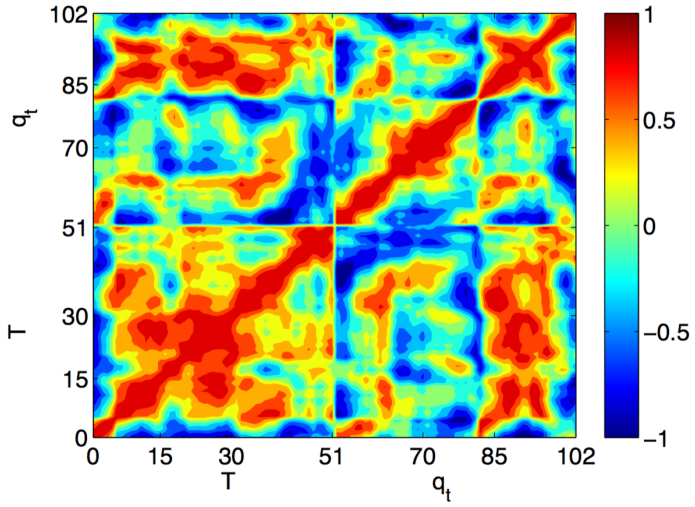
(d) $N = 51$

Images from Vetra-Carvalho (2012) Properties of the Ensemble Kalman Filter for Convective-Scale Numerical Weather Forecasting, PhD Thesis, University of Reading

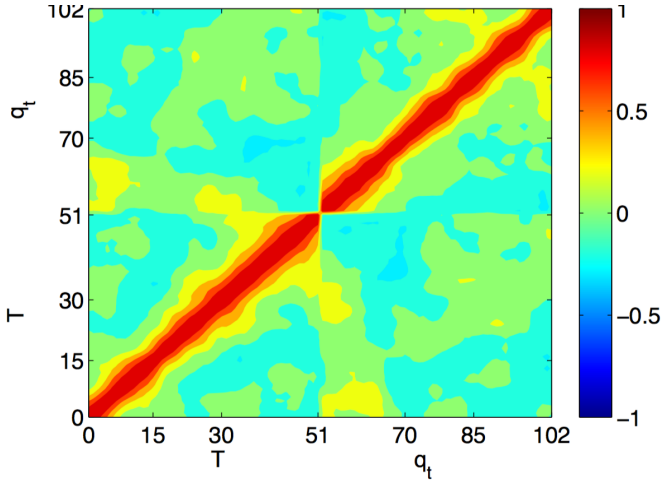
Sampling error



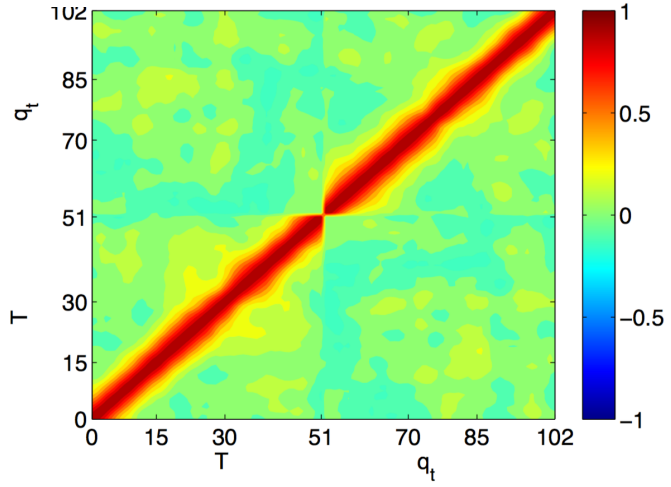
(a) Original



(b) $N = 6$



(e) $N = 102$

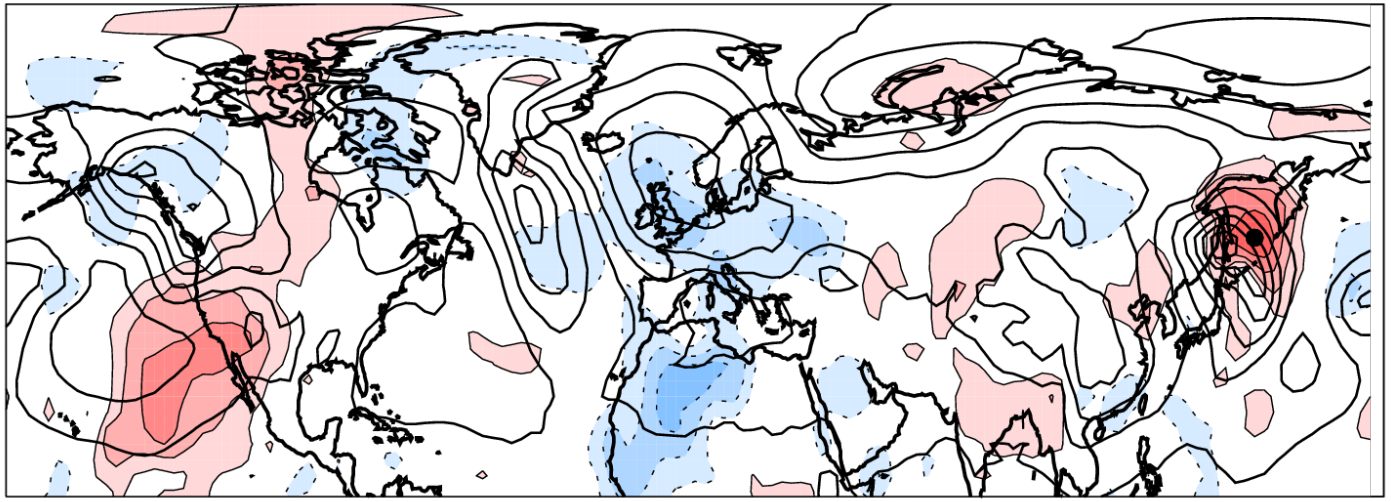


(f) $N = 204$

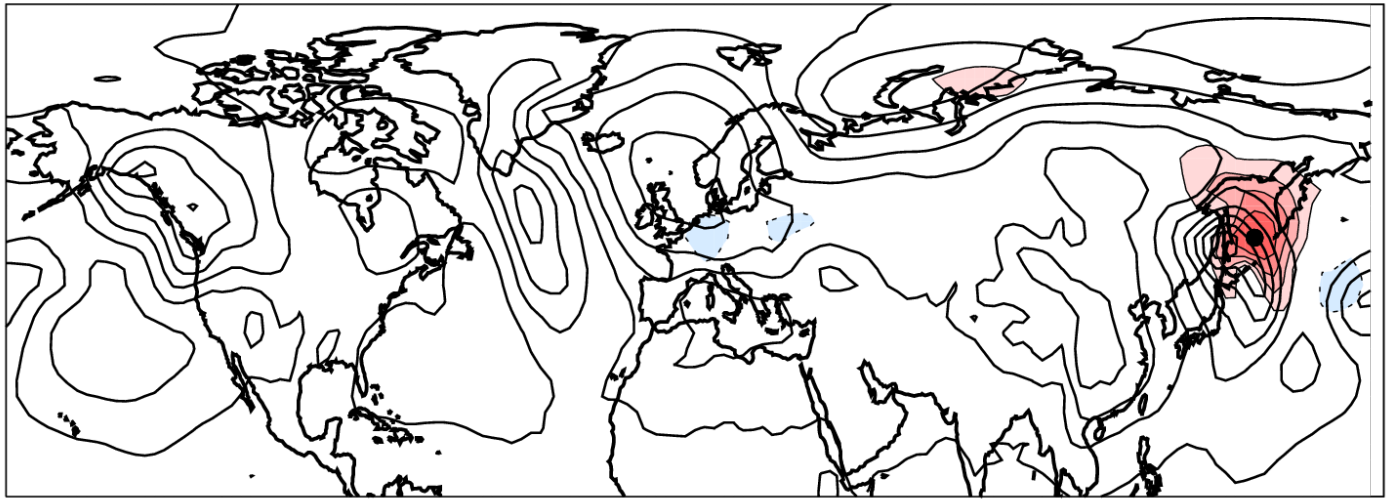
Images from Vetra-Carvalho (2012) Properties of the Ensemble Kalman Filter for Convective-Scale Numerical Weather Forecasting, PhD Thesis, University of Reading

Spurious correlations

(a) Correlations in P^b, 25−member ensemble



(b) Correlations in P^b, 200−member ensemble



from Hamill, Chapter 6 of "*Predictability of Weather and Climate*"

Underestimated ensemble variance

The analysis increments are in the sub-space spanned by the forecast ensemble

- The analysis increments are given by

$$\mathbf{x}_i^{\mathbf{a}} - \mathbf{x}_i^{\mathbf{f}} = \mathbf{P}_e^{\mathbf{f}}\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{P}_e^{\mathbf{f}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right)^{-1}(\mathbf{y} - H(\mathbf{x}_i^{\mathbf{f}}))$$

- The analysis increments are therefore a linear combination of the forecast error ensemble.

- Therefore, even if the observations indicate otherwise, the analysis is restricted to the space spanned by the ensemble which has at the most a dimension of $N_e - 1$.

Underestimated ensemble variance

The forecast ensemble spread will be subject to sampling error.

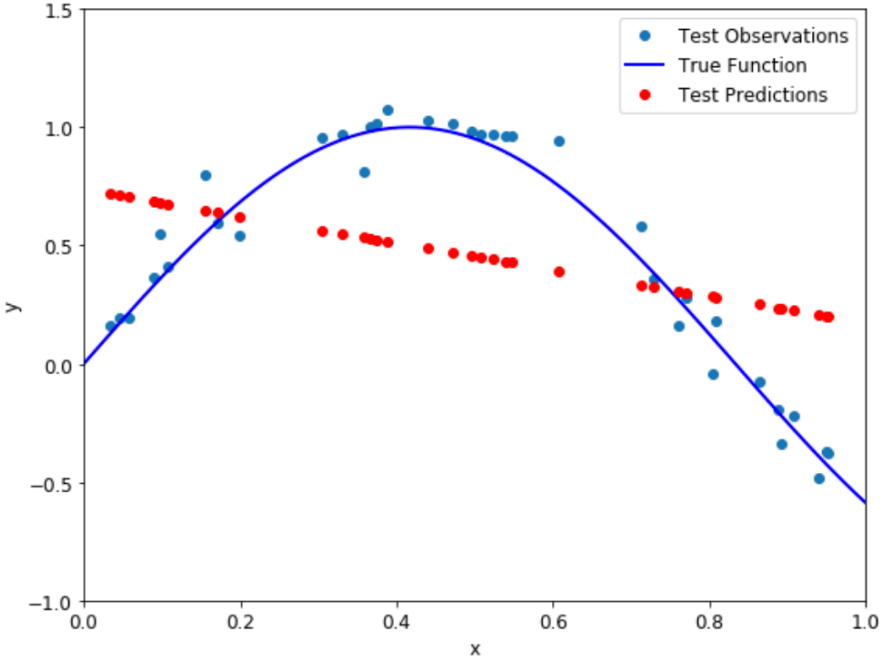
If the spread is too small, the ensemble will underfit to the observations.

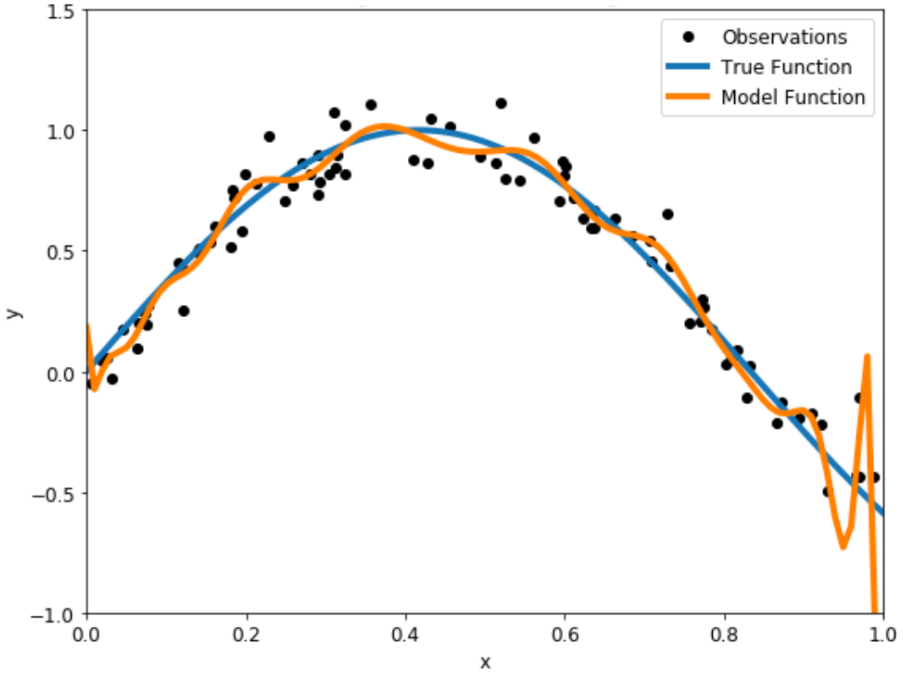
If the ensemble repeatedly underestimates the forecast error and the information in the observations is ignored then it is difficult to regain spread in the ensemble. This is called ‘filter divergence’.

If the spread is too large, the analysis ensemble will overfit to the observations.

Underestimated ensemble variance

Underfitting **vs.** **Overfitting**





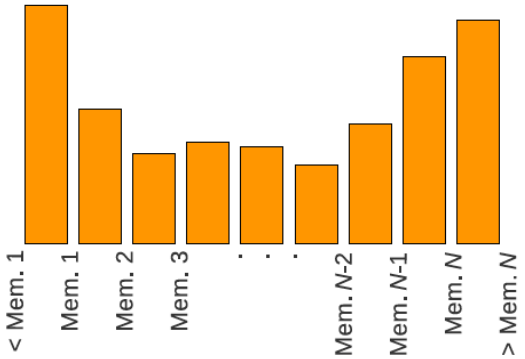
Validation of ensemble spread

Method 1: rank histograms

- For the ensemble to be reliable it is assumed that it is sampling the same distribution as the truth.
- A rank histrogram is constructed by considering a point in space that is well observed.
  - The values of the ensemble members at that point are ranked from highest to lowest creating N-1 bins.
  - Then each observation is binned to give a frequency diagram.

Interpretation:

- Concave shape- the ensemble is underspread
- Convex shaped- the ensemble is overspread
- Flat- the ensemble is correctly spread
- Asymmetric- the ensemble is biased



Need for caution (Hamill, T., 2001):

- A flat histogram does not necessarily indicate reliability of the ensemble
- A flat histogram does not ensure the covariances are correctly specified
- A concave shaped histogram could also be the result of conditional biases- need to look at sub-populations
- Observation errors should be accounted for by adding random noise to each ensemble member consistent with the observational error statistics

Possible solutions

1. Use more ensemble members (see Miyoshi et al. 2014)

2. Localisation
   - Addresses problem of spurious correlations
   - Splits problem into quasi-independent problems
   - Increases the effective ensemble size

3. Ensemble inflation
   - Addresses problem of filter divergence

4. Combine ensemble with variational approaches
   - These are known as hybrid methods

Localisation

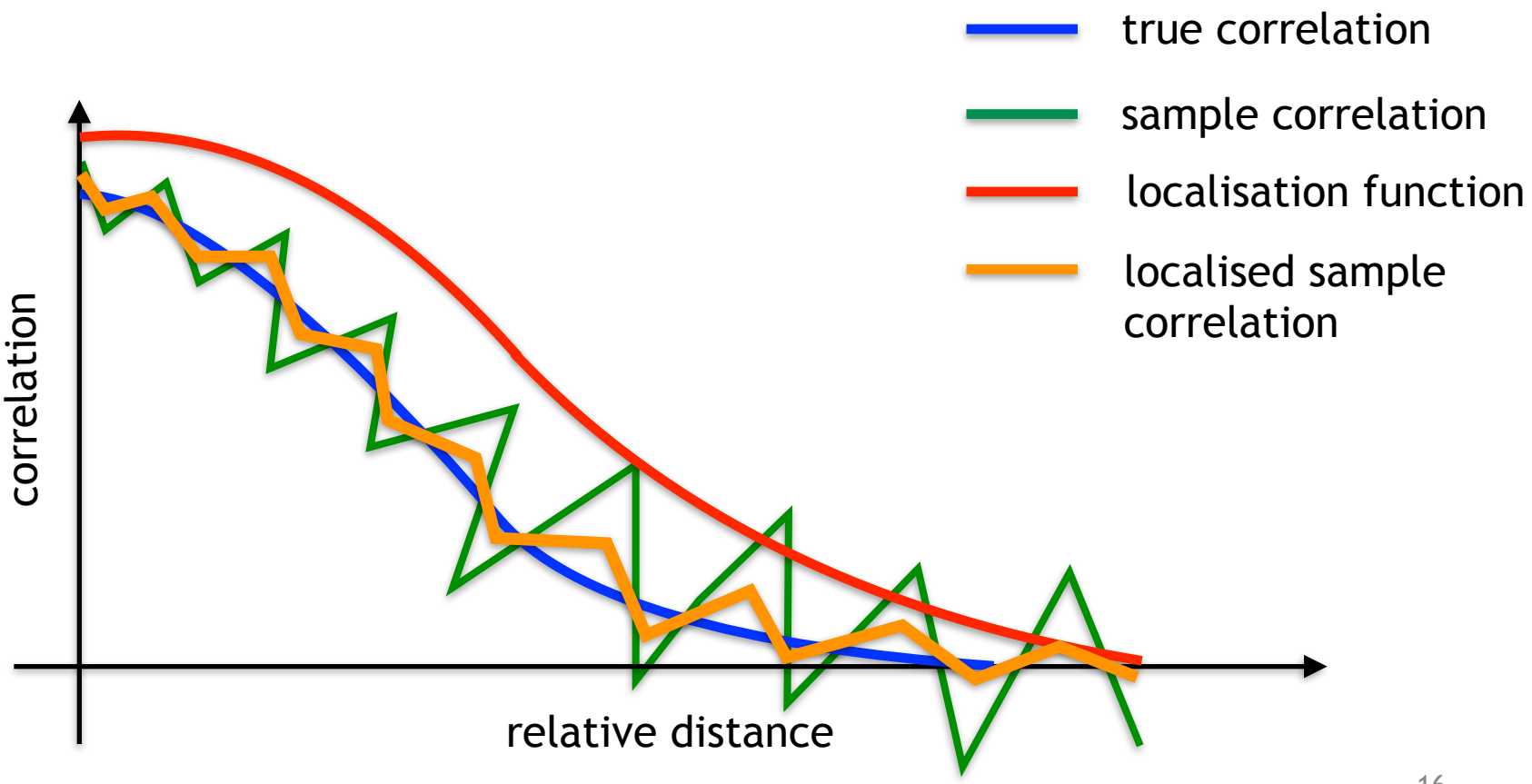
The underlying assumption of localisation is that correlation length scales are much shorter than the extent of the model grid so that only correlations over short distances are relevant while for long distances the sampling error in the ensemble-estimated covariance matrices dominates (see, e.g., Morzfeld et al., 2017).

The aim of localisation is to minimise the issues caused by the rank problem. Two most popular localisation methods are:

- **Covariance localisation**: this modifies the forecast error covariance matrix to reduce long-range correlations.
- **Observation localisation:** This restricts observations which are allowed to influence each grid point.

Covariance localisation

Localisation can be performed by multiplying the ensemble covariances with a smooth correlation function.



Covariance localisation

Since in practice we don't form the ensemble covariance matrix direct application of a localisation function ρ is not possible,

$$\mathbf{P}_L^f = \rho \circ \mathbf{P}_e^f .$$

Petrie (2008) showed that we cannot apply localisation to ensemble perturbation matrix directly, i.e.

$$\left(\rho\rho^{\mathbf{T}}\right) \circ \left(\mathbf{X}'^f \left(\mathbf{X}'^f\right)^T\right) \neq \left(\rho \circ \mathbf{X}'^f\right)\left(\rho \circ \mathbf{X}'^f\right)^{\mathbf{T}} .$$

However, if structure of H is close to diagonal we can justify an approximation such that

$$\left(\rho \circ \mathbf{P}_e^{\mathbf{f}}\right)\mathbf{H}^{\mathbf{T}} = \rho \circ \left(\mathbf{P}_e^{\mathbf{f}}\mathbf{H}^{\mathbf{T}}\right)$$

Hence,

$$\mathbf{K}_e = \rho \circ \left(\mathbf{P}_e^{\mathbf{f}}\mathbf{H}^{\mathbf{T}}\right)\left[\rho \circ \left(\mathbf{H}\mathbf{P}_e^{\mathbf{f}}\mathbf{H}^{\mathbf{T}}\right) + \mathbf{R}\right]^{-1} .$$

Covariance localisation

In the case of the *stochastic ensemble Kalman filter* where analysis update is computed separately for each ensemble member and model predicted observations are perturbed we can solve for ϕ_i instead of inverting the matrix

$$\rho \circ \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^{\mathbf{T}} + \mathbf{R} \right) \phi_i = y_i - y_{m,i}^f, \qquad i = 1, \ldots, N_e$$

Then our analysis update is

$$\mathbf{x}_i^{\mathbf{a}} = \mathbf{x}_i^{\mathbf{f}} + \left[\rho \circ \left(\mathbf{P}_e^{\mathbf{f}} \mathbf{H}^{\mathbf{T}} \right) \right] \phi_{i.}$$

However, for ensemble square root filters we cannot do this since the updated ensemble is constructed from analysis mean and analysis perturbations, i.e.

$$\mathbf{X}^{\mathbf{a}} = \bar{\mathbf{X}}^{\mathbf{f}} + \mathbf{X}'^{\mathbf{f}} \left(\bar{\mathbf{W}} + \mathbf{W}' \right).$$

Covariance localisation

For *square root filters* such as Ensemble Adjustment Kalman Filter (EAKF, Anderson 2001) and Ensemble Square Root Filter (EnSRF, Whitaker and Hamill 2002) which have been developed with **serial observations** it is possible to apply covariance localisation.

Let *j* indicate the *j*th observation loop, then analysis mean is

$$\overline{\mathbf{x}}^{\mathbf{a}}_{(j)} = \overline{\mathbf{x}}^{\mathbf{f}}_{(j)} + K_{(j)}\left(\mathbf{y}_{(j)} - H(\overline{\mathbf{x}}^{\mathbf{f}}_{(j)})\right)$$

where

$$\overline{\mathbf{x}}^{\mathbf{a,f}}_{(j)} \in \mathbb{R}^{N_x \times 1},\ K_{(j)} \in \mathbb{R}^{N_x \times 1},\ \mathbf{y}_{(j)}, H(\mathbf{x}^{\mathbf{f}}_{(j)}) \in \mathbb{R}.$$

With the Kalman gain given by

$$\mathbf{K}_{(j)} = \mathbf{P}^{\mathbf{f}}_{(j)}\mathbf{H}^{\mathbf{T}}_{(j)}\left(\mathbf{H}_{(j)}\mathbf{P}^{\mathbf{f}}_{(j)}\mathbf{H}^{\mathbf{T}}_{(j)} + \mathbf{R}_{(j)}\right)^{-1}.$$

vector, $\mathbb{R}^{N_x \times 1}$ *scalar*

Note, that serial filter formulation requires that R is diagonal.

Covariance localisation

And analysis ensemble perturbations are given by

$$\mathbf{X}'^{\mathbf{a}}_{(j)} = \mathbf{X}'^{\mathbf{f}}_{(j)} + \hat{\mathbf{K}}_{(j)}\mathbf{H}_{(j)}\mathbf{X}'^{\mathbf{f}}_{(j)}$$

with

$$\hat{\mathbf{K}}_{(j)} = \left[1 + \sqrt{\frac{\mathbf{R}_{(j)}}{\mathbf{H}_{(j)}\mathbf{P}^{\mathbf{f}}_{(j)}\mathbf{H}^{\mathbf{T}}_{(j)} + \mathbf{R}_{(j)}}} \right]^{-1} \mathbf{K}_{(j)}.$$

The scalar in front of the gain $\mathbf{K}_{(j)}$ reduces the Kalman gain.

This reduction is required for statistical consistency as without it the analysis error variances would be underestimated unless an ensemble of perturbed observations would be used (Burgers et al. 1998).

For more detail see Whitaker and Hamill (2002) and Nerger (2015).

Covariance localisation

- Test runs with Lorenz-96 model
- Vary ensemble size and localization radius

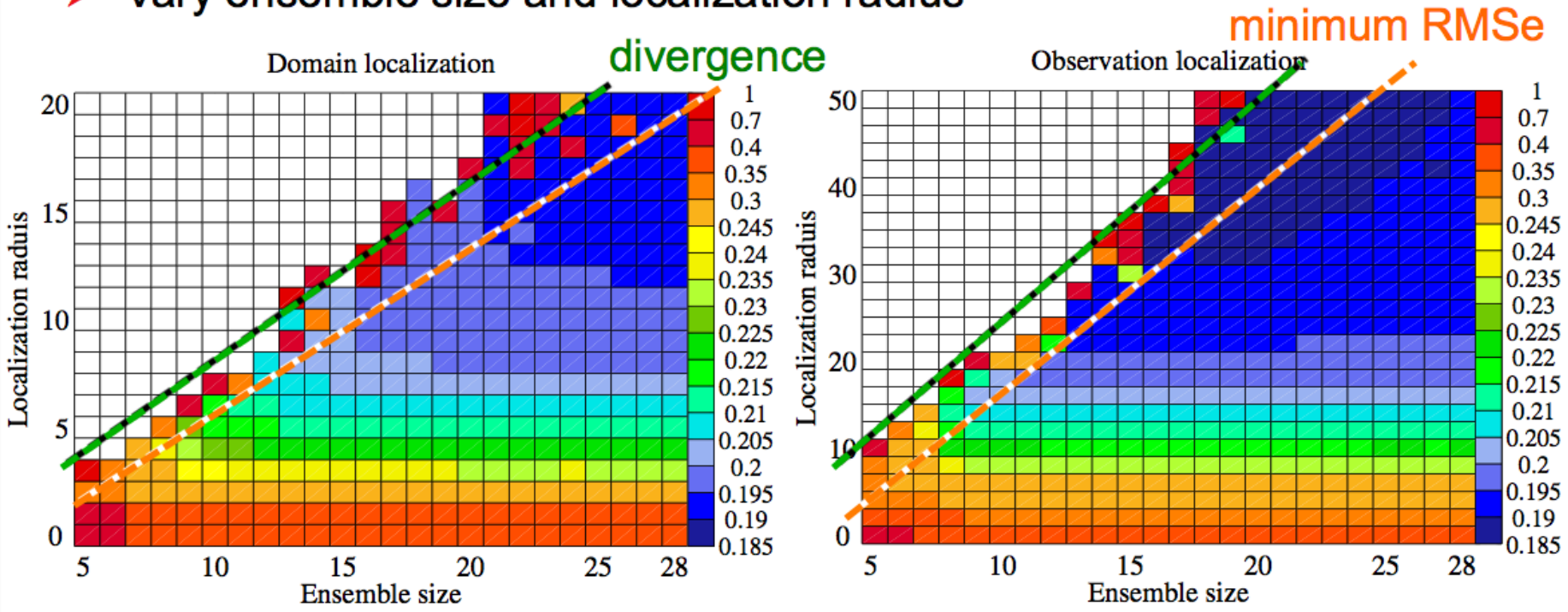
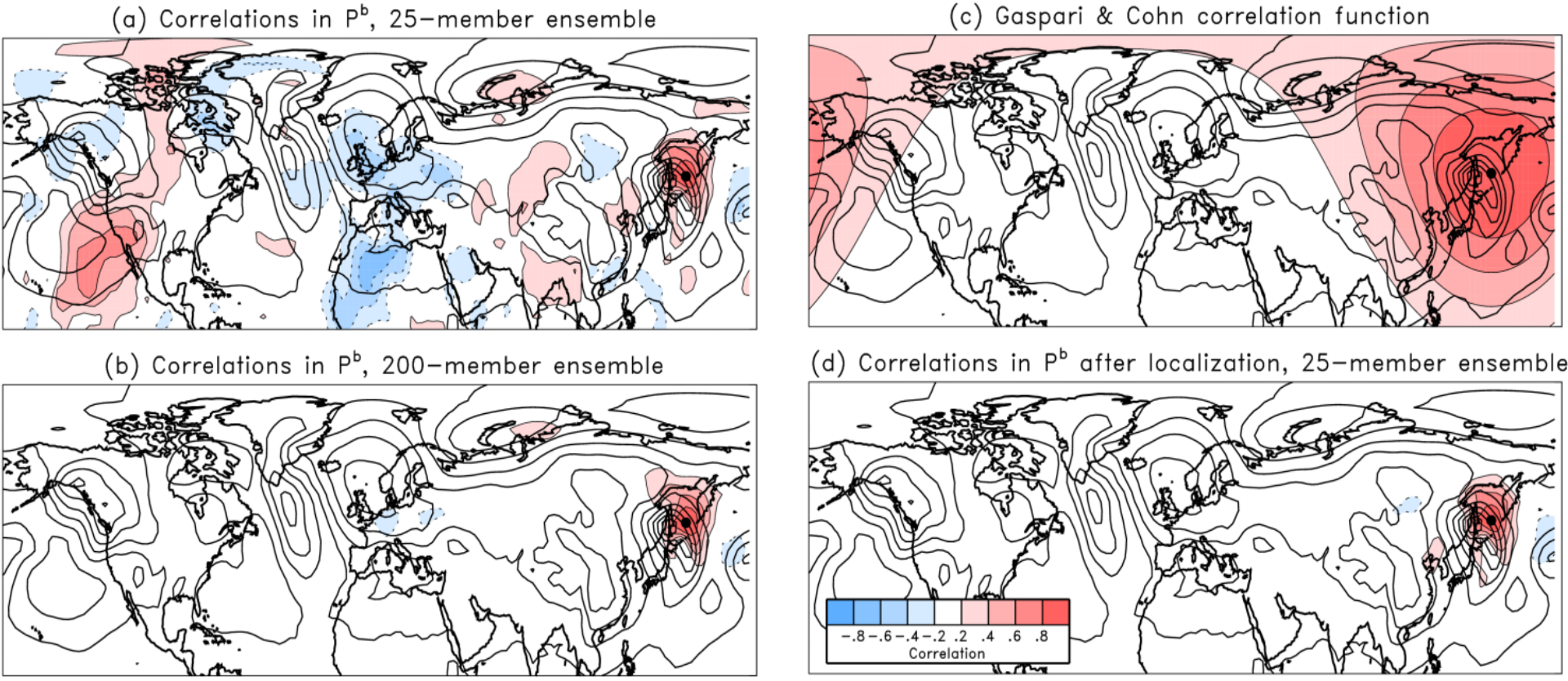


Image credit: Lars Nerger, AWI

Covariance localisation



From Fig. 6.4 of Hamill, 2006

Summary of covariance localisation

+ Can be applied to stochastic EnKF and serial versions of EAKF and EnSRF.
+ Reduces the spurious ensemble correlations.
+ Increases the effective degrees of freedom - more ways the ensemble can adjust to the data.
- Can affect balance e.g. geostrophic balance which has lenghtscale O(1000)km
- Need to choose the localisation function and its length scale, this may be state-dependent.
- Not clear how to define distance between observations which have no clear defined location in space, e.g. satellite observations.
- In case if the observations have a strong influence, Nerger (2015) showed that the interaction of serial observation processing and localisation can destabilise the analysis process.

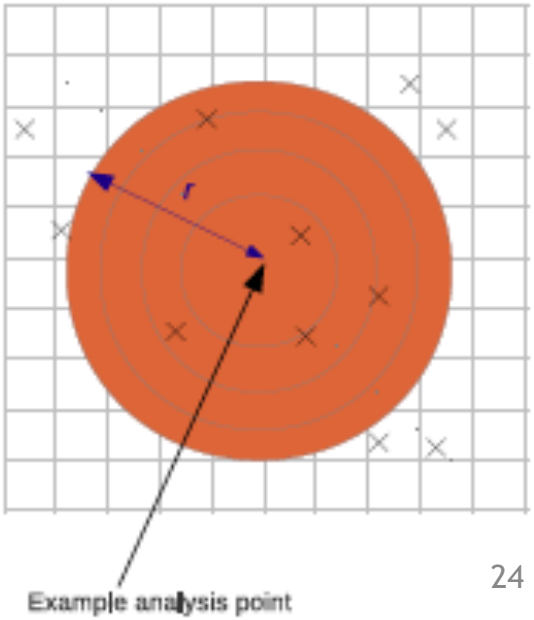
Observation localisation

Observation localisation reduces the influence of observations far away from analysis point by inverse scaling observation-error covariances.

For many square root ensemble Kalman filters, e.g. LETKF (Hunt et al., 2007), ESTKF (error subspace transform Kalman filter, Nerger et al. 2012), observation localisation is a much more natural choice since covariances are never explicitly calculated in these filters.

For a smooth analysis Hunt et al. (2007) proposed to use a gradual observation localisation in the LETKF acting on R^{-1}.

This smoothly reduces the observation influence and excludes observations outside a defined radius by prescribing their error to be infinitely large.



Localisation and LETKF

LETKF analysis updates, similarly to ETKF, are given by

$$\overline{\mathbf{x}}_{\sigma}^{\mathbf{a}} = \overline{\mathbf{x}}_{\sigma}^{\ \mathbf{f}} + \mathbf{X}'^{\mathbf{f}}_{\sigma}\overline{\mathbf{w}}_{\delta}, \quad \mathbf{X}'^{\mathbf{a}}_{\sigma} = \mathbf{X}'^{\mathbf{f}}_{\sigma}\mathbf{W}'_{\delta},$$

through finding

$$\mathbf{T}_{\delta}\mathbf{T}_{\delta}^{\mathbf{T}} = \left(\mathbf{I} + \frac{1}{N_e - 1}\mathbf{S}_{\delta}^{\mathbf{T}}\left(\mathbf{D}_{\delta} \circ \mathbf{R}_{\delta}^{-1}\right)\mathbf{S}_{\delta}\right)^{-1}, \quad \left(\mathbf{T}_{\delta}\mathbf{T}_{\delta}^{\mathbf{T}}\right)^{-1} = \mathbf{U}\Sigma\mathbf{U}^{\mathbf{T}}.$$

This gave the LETKF weight vector and matrix as follows,

$$\overline{\mathbf{w}}_{\delta} = \frac{1}{\sqrt{N_e - 1}}\mathbf{U}\Sigma^{-1}\mathbf{U}^{\mathbf{T}}\mathbf{S}_{\delta}^{\mathbf{T}}\left(\mathbf{D}_{\delta} \circ \mathbf{R}_{\delta}^{-1}\right)\mathbf{d}_{\delta}, \quad \mathbf{W}'_{\delta} = \mathbf{U}\Sigma^{-1/2}\mathbf{U}^{\mathbf{T}}.$$

- σ denotes the local analysis domain
- δ denotes the domain of the corresponding observations of nonzero weight

Localisation and LETKF

To obtain the LETKF as a localised form of the ETKF, the analysis and the ensemble transformation are performed in a loop through disjoint local analysis domains.

For each local analysis domain, the observations are weighted by their distance from this domain using

$$\tilde{\mathbf{R}} = \tilde{\mathbf{D}} \circ \mathbf{R}$$

where localisation matrix is usually constructed from a correlation function with compact support, e.g. see Gaspari and Cohn, (1999).

Note that $\tilde{\mathbf{D}}$ is not a correlation matrix, because the diagonal elements vary with the distance.

Summary of observation localisation

+ Observation localisation can be implemented in more filters.
+ The optimal lengthscale for R-localisation is found to be shorter than for Pf-localisation (Greybush, 2011).
+ Using the optimal lengthscales, R-localisation and Pf-localisation have comparable performance in terms of analysis RMSE and balance.
+ Reduces spurious ensemble correlations.
+ Increases effective ensemble size.
- Optimal lengthscales need to be found.
- Need to take care to not produce a patchy analysis.
- In case if the observations have a strong influence, Nerger (2015) showed that the interaction of serial observation processing and localisation can destabilise the analysis process.

Localisation length scale

The ***optimal localisation*** radius is a priori unknown and needs to be tuned in numerical experiments:

- perform several DA experiments with different localisation radii.
- select the localisation length scale which results in smaller estimation errors.

"Localisation scale is a tuning parameter that trades off computational effort, sampling error, and imbalance error."

*(Ref: Zhou et al., 2008, MWR, **136**, 678-698.)*

Kirchgessner et al. (2014) showed that the optimal localisation radius should be reached when the sum over the observation weights (using Gaspari and Cohn, 1999 correlation function) equals the ensemble size.

- allows for a simple form of adaptivity or a starting point for further tuning.

Other localisation ideas

- **Multiscale filter**: at each update time replaces the sample covariance with a multiscale tree composed of nodes distributed over a relatively small number of discrete scales, Zhou et al. (2008).
- **Spectral localisation**: filter covariances after transformation to spectral space, or combine spectral and grid-point localisation, Buehner and Charron, (2007).
- **Adaptive localisation:** adaptively specify the localisation function or radius according to the dynamically generated covariance structure.
- **Localisation in different variables** (i.e. stream function, velocity potential, Kepert (2006)).

Active research field!

Many interesting ideas but none yet accepted as standard. Alternatives tend to have their own set of strengths and weaknesses.

Ensemble inflation

Localisation is one method to reduce the undersampling.

However, for high-dimensional systems, localisation alone is not sufficient to ensure a stable assimilation process and **covariance inflation** is applied to further increase the sampled variance and thus stabilise the filter.

In addition, the inflation can partly account for model error in case of an imperfect model (Pham et al., 1998b; Hamill, 2001; Anderson, 2001; Whitaker and Hamill, 2002; Hunt et al., 2007).

Ensemble inflation

Most common is a ***fixed multiplicative covariance inflation*** (Anderson and Anderson, 1999).

The method uses the inflation factor *r* to perform a multiplicative inflation for each ensemble member $\mathbf{x}^{a,f}$ with i = 1, ..., Ne

$$\mathbf{x}_i^{\mathbf{a},\mathbf{f}} = r\left(\mathbf{x}_i^{\mathbf{a},\mathbf{f}} - \overline{\mathbf{x}}^{\mathbf{a},\mathbf{f}}\right) + \overline{\mathbf{x}}^{\mathbf{a},\mathbf{f}}$$

where *r* is slightly greater than 1.

The specification of an optimal inflation factor may vary according to the size of the ensemble (Hamill, 2001; Whitaker and Hamill, 2002) and the choice of *r* will depend on various factors, such as dynamics of the model, type of the ensemble filter used as well as the length scale of covariance localisation.

Ensemble inflation

Related to covariance inflation is the so-called ***forgetting factor*** ρ introduced by Pham et al. (1998b).

The forgetting factor is usually chosen to be slightly lower than one and is typically applied in the square-root filters like the ETKF, SEIK, and ESTKF. For example, in the ETKF it is applied to

$$\mathbf{T}\mathbf{T}^{\mathrm{T}} = \left(r\mathbf{I} + \frac{1}{N_e - 1}\mathbf{S}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{S} \right)^{-1},$$

In this way, the inflation and forgetting factors are related as $\rho = r^{-2}$.

Forgetting factor allows one to apply inflation in a computationally very efficient way because $\mathbf{T}\mathbf{T}^{\mathrm{T}}$ is much smaller than the ensemble states to which the inflation is applied to.

Ensemble inflation

The multiplicative inflation leads to an inflation that is relative to the variance level. Thus, large variances will be inflated much more than small variances.

An ***additive inflation*** (Ott et al., 2004) inflates all variances by the same amount, rather than a relative factor, ,

$$\hat{\mathbf{x}}_i^{\mathbf{f}} = \mathbf{x}_i^{\mathbf{f}} + \eta_i, \qquad \eta_i \sim N(0, \mathbf{Q})$$

This difference can be useful if the variances vary strongly as in this case the additive inflation acts stronger on the very small variances.

Apply additive and multiplicative inflations separately or together depending on your system.

Hybrid methods

Hybrid methods combine the best parts of the EnKF (flow-dependent P^f) with the best parts of variational methods (full rank P^s).

The earliest hybrid method was proposed by Hamill and Snyder (2004), in which the representation of the error covariance of the prior information is a weighted combination of the flow-dependent estimate from the EnKF and the full rank estimate used in variational methods

$$\mathbf{P}_{hybrid} = \beta \mathbf{P}^s + (1 - \beta)\mathbf{P}^f$$

where β is a tuneable parameter.

Note localisation and inflation are still necessary.

Summary 1.

The success of the EnDA methods is highly dependent on the size of the ensemble being adequate for the system we apply these methods to.

However, in reality the ensemble size is much much smaller than the state space resulting in sampling errors in ensemble covariances causing:

- Rank deficiency,
- Filter divergence,
- Spurious correlations,
- Analysis degradation.

Hence, to make ensemble DA practical

- Ensemble inflation
- Localisation
- Hybrid methods

Other practical issues

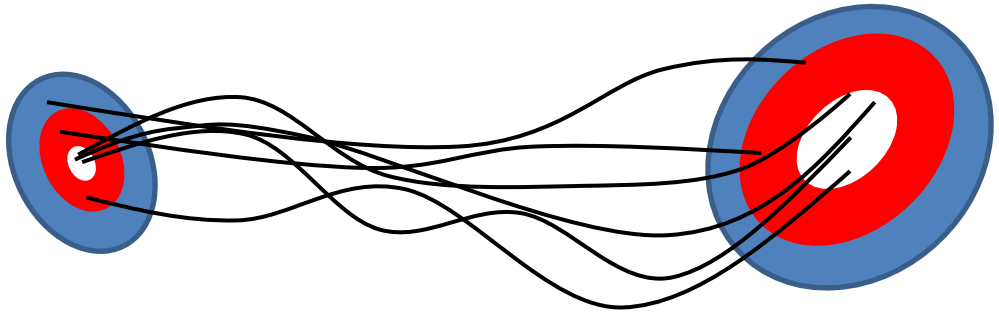
Parallelisation of EnKF

The most costly part of the ensemble data assimilation is the model runs between observations.

However, the integration of each ensemble state is independent from the other states -> run the numerical model N_e times.

We can parallelise:

- offline where the dynamical model is kept independent of the data assimilation system.
- by coupling the dynamical model with the data assimilation system, e.g. EMPIRE, PDAF.

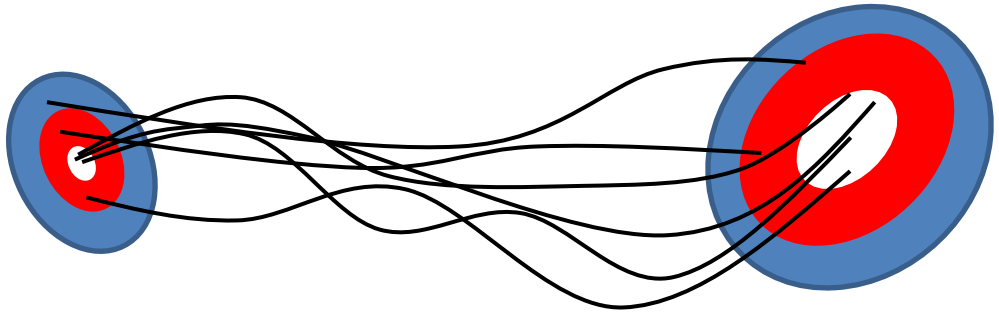


Parallelisation of EnKF

Using for example Message Passing Interfaces (MPI) to couple the dynamical model to the DA system parallelising not only the model runs but also the analysis step.

Parallelisation differs depending:

- which filter is used;
- if localisation is used;
- if observations are assimilated serially or in bulk; using the domain-decomposition of a model was found to be more efficient than using serial observations since the amount of data that has to be exchanged using MPI is smaller for domain-decomposition (Nerger et al., 2005a)



Non-linearity and EnKF

Due to technological and scientific advances, three significant developments have occurred in the last decade that force us to look beyond standard Ensemble Kalman Filtering, which is based on linear and/or Gaussian assumptions.

- Continuous **increase in computational capability** has recently allowed to run operational models at high resolutions so that the dynamical models have become increasingly nonlinear due to the direct resolution of small-scale nonlinear processes in these models, e.g. small-scale turbulence.
- In several fields such as atmosphere, ocean, land surface, hydrology, and sea-ice, it is of interest to estimate **bounded variables or parameters** requiring DA methods that can deal with non-Gaussian distributions.
- **observational network** around the world has increased manyfold for weather, ocean and land surface areas providing more information about the real system with greater accuracy and higher spatial and temporal resolution.

Non-linearity and EnKF

All EnKFs update ensemble using Gaussian assumption. **Is this a valid assumption when the dynamical model is non-linear?**

The optimality of the Kalman filters is no longer preserved when applied to non-linear problems resulting in an suboptimal analysis state and the error estimates.

This is a common issue for all ensemble filters whose analysis step is based on the equations of the Kalman filter.

Nonetheless, the many existing data assimilation studies with nonlinear models, e.g. of the ocean or atmosphere, with different formulations of the ensemble Kalman Filters show that these filters are rather stable with regard to nonlinearity.

Non-linearity and EnKF

This is a very hot topic right now!

Current methods are assessed for the application to non-linear problems and numerous new data assimilation methods are being developed within ensemble data assimilation framework and other fields, e.g. particle filters.

As well as hybrids between the various data assimilation methods to bring the best of the methods in one.

Summary 2.

The idea of the data assimilation is easy!

To make it work in the real world is the hard part.

Further reading

- Bannister et al. 2011: **Ensemble prediction for nowcasting with a convection-permitting model - II: forecast error statistics**, Tellus A, 497-512 .
- Greybush, 2011: **Balance and ensemble Kalman filter localisation techniques**, Mon. Wea. Rev., 139, 511-522.
- Hamill, T., 2001: **Interpretation of Rank Histograms for Verifying Ensemble Forecasts**, Mon. Wea. Rev., 129,550-560.
- Houtekamer and Zhang (2016) **Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation,** *Mon. Weather Rev.*, **144**, 4489–4532
- Houtekamer and Mitchell 1998: **Data assimilation using an ensemble Kalman Filter technique**, Mon. Wea. Rev. 126, 796-811.
- Miyoshi et al. 2014: **The 10,240-member ensemble Kalman filtering with an intermediate AGCM**, Geophys. Res. Lett., 41, 5264-5271.
- P. Kirchgessner, L. Nerger, and A. Bunse-Gerstner. **On the choice of an optimal localization radius in ensemble Kalman filter methods.** *Mon. Wea. Rev.*, 142:2165–2175, 2014.
- L. Nerger. ***Parallel Filter Algorithms for Data Assimilation in Oceanography***, Number 487 in Reports on Polar and Marine Research. Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany, 2004. PhD Thesis, University of Bremen, Germany.
- L. Nerger. **On serial observation processing on localized ensemble Kalman filters**, *Mon. Wea. Rev.*, 143:1554–1567, 2015.
- 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

Open source DA code/systems

- **PDAF** - MPI based DA system developed by L. Nerger (2004) http://pdaf.awi.de/trac/wiki
- **SANGOMA DA library** in Matlab and Fortran - consistently written and well tested, http://www.data-assimilation.net/Tools/