Deutscher Wetterdienst

Variants of Ensemble Kalman Filter algorithms – pros and cons

Andreas Rhodin

Deutscher Wetterdienst

COSMO General Meeting, Offenbach
September 7, 2009

Motivation

- An ensemble data assimilation system is under development for COSMO.
- As a first candidate the LETKF variant following Hunt et al. (2007) has been chosen.
- Ensemble Kalman filter in general and the LETKF in particular has been criticised to have a number of shortcomings.
  (cf. subsequent talk of Mikhail Tsyrulnikov)
- **The goal of this talk is to revisit the rational for this choice and for possible alternatives.**

Why LETKF (for COSMO)

- An Ensemble Forecasting system is currently build up in order to quantify the uncertainties of the forecast.
- A more theoretically founded assimilation system than the current (nudging-) scheme is desired.
- The prerequisites to build up a 4D-Var are not given. (adjoint model, 'smooth' physics)
- **An Ensemble Assimilation Systems is a natural choice for a data assimilation system that at the same time provides initial values for a forecast ensemble.**
- Ensemble data assimilation systems came up (relative) recently. There is a chance to participate in the early development phase and not to lag behind the developments of others.
- **The LETKF is a particularly fast EnKF implementation.**
- Alternatives to the EnKF (currently SIR-Filter) are pursued. (with lower priority)

Foundations of Ensemble Data Assimilation Systems

- Ensemble Data Assimilation Systems use an ensemble of forecasts (with members $\mathbf{x}_k^b$) to infer (flow dependent) information on the background error covariance matrix $\mathbf{B}$ for the data assimilation system.

  $\mathbf{B} \approx \mathbf{X}\mathbf{X}^T \quad \text{with} \quad \mathbf{X}_k = \sqrt{\frac{1}{n_k - 1}} \left(\mathbf{x}_k^b - \overline{\mathbf{x}_k^b} \right)$
- In the analysis step an ensemble of analyses is provided. The $\mathbf{x}_k^a$ represent the uncertainty of the analysis, based on :
  - the forecast error derived from the $\mathbf{x}_k^f$,
  - and the prescribed observational error.
  - Model error has to be introduced explicitly into the $\mathbf{x}_k^f$.
- The $\mathbf{x}_k^a$ are the initial values for cycling the ensemble to the next analysis time.
- The $\mathbf{x}_k^a$ can be used as the initial values for forecast ensembles in general.

Limitations of Ensemble Data Assimilation Systems

- $\mathbf{B} = \mathbf{X}\mathbf{X}^T$ is of low rank.
  (compare ensemble size to number of degrees of freedom of the atmosphere).
- Small off-diagonal elements (correlations) of $\mathbf{X}\mathbf{X}^T$ are noisy.
- nonlinearities are accounted for in the ensemble forecast, but : assumptions on linearity and normal distributions are made in the analysis step.

Precautions to deal with the low rank/noise problem are required.

This is accomplished by localisation.

Localisation: Schur product

Multiply $\mathbf{B} = \mathbf{X}\mathbf{X}^T$ element by element with a matrix $\mathbf{C}$:
$\mathbf{B} \rightarrow \mathbf{C} \circ \mathbf{B}$.

Requirements on $\mathbf{C}$:

- $\mathbf{C}_{ij} \approx 1$ for large correlations in $\mathbf{B}$
- $\mathbf{C}_{ij} \approx 0$ for small correlations in $\mathbf{B}$
- $\mathbf{C} \circ \mathbf{B}$ must be positive definite.
  ($\mathbf{C}$ positive definite is a sufficient condition).

Common choice: piecewise rational function (Gaspari & Cohn)

- $\mathbf{C}_{ij}$ is defined as a smooth function of $\triangle x = x_i - x_j$.
- $\mathbf{C}(\triangle x)$: Gaussian like function characterised by a localisation length scale λ_l.
- $\mathbf{C} = 0$ for $\triangle x$ large. (facilitates computations)
- $\lambda_l >> $ correlation length scale λ_c.
  - Do not impair covariances provided by the ensemble
  - Maintain balances

Remarks on localisation in physical space

Remarks:

- Definition of $\mathbf{C}$ just as a function of $\triangle x$ may be sub-optimal. (correlation length scales may vary for different model variables)
- Statistical considerations could be used to chose $\mathbf{C}_{ij}$.

More remarks:

- To apply the Schur product on $\mathbf{B}$ in physical space is an arbitrary choice.
- Application in spectral or in wavelet representation has totally different effects. (Buehner and Charron 2007)

Ensemble Transform Kalman Filter (ETKF)

Perform the analysis using the gain matrix $\mathbf{K}$:

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{K}\ (\mathbf{o} - H(\mathbf{x}^b))$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T\ (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

set

$$\mathbf{B} = \mathbf{X}\mathbf{X}^T \quad \text{with} \quad \mathbf{X}_k = \sqrt{\tfrac{1}{n_k - 1}} \left(\mathbf{x}_k^b - \overline{\mathbf{x}_k^b}\right)$$

and

$$\mathbf{H}{=}\mathbf{Y}\mathbf{X} \quad \text{with} \quad \mathbf{Y}_k = \sqrt{\tfrac{1}{n_k - 1}} \left(H(\mathbf{x}_k^b) - \overline{H(\mathbf{x}_k^b)}\right)$$

Then the gain matrix becomes :

$$\mathbf{K} = \mathbf{X}\mathbf{Y}^T\ (\mathbf{Y}\mathbf{Y}^T + \mathbf{R})^{-1}$$

Finally derive the analysis ensemble deviations $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{Y}^T\ (\mathbf{Y}\mathbf{Y}^T + \mathbf{R})^{-1} (\mathbf{o} - H(\mathbf{x}^b)) \quad \text{with} \quad \mathbf{Z} = \sqrt{\tfrac{1}{n_k - 1}} \left(\mathbf{x}_k^a - \mathbf{x}_k^b\right)$$

Variants of localisation

For computational efficiency :
apply $\mathbf{C}\circ$ on $\mathbf{BH}^T$, $\mathbf{HBH}^T$ or $\mathbf{R}^{-1}$ instead of $\mathbf{B}$

- Kalman Gain matrix

  $$\mathbf{K} = \mathbf{BH}^T\ (\mathbf{HBH}^T + \mathbf{R})^{-1}$$

  or

  $$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\ \mathbf{R}^{-1}$$

- pure EnKF (small set of linear equations)

  $$\mathbf{K} = \mathbf{XY}^T\ (\mathbf{YY}^T + \mathbf{R})^{-1}$$

  or

  $$\mathbf{K} = \mathbf{X}(\mathbf{1} + \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y})^{-1}\ \mathbf{Y}^T\mathbf{R}^{-1}$$

- localisation on $\mathbf{B}$ (requires $\mathbf{H}$)

  $$\mathbf{K} = \mathbf{C}\circ\mathbf{XX}^T\mathbf{H}^T\ (\mathbf{H}\,\mathbf{C}\circ\mathbf{XX}^T\mathbf{H}^T + \mathbf{R})^{-1}$$

- localisation on $\mathbf{BH}^T$, $\mathbf{HBH}^T$

  $$\mathbf{K} = \mathbf{C}\circ\mathbf{XY}^T\ (\mathbf{C}\circ\mathbf{YY}^T + \mathbf{R})^{-1}$$

- localisation on $\mathbf{R}^{-1}$ (LETKF, Hunt et al. 2007)

  $$\mathbf{K} = \mathbf{X}\,(\mathbf{1} + \mathbf{Y}^T\mathbf{C}\circ\mathbf{R}^{-1}\,\mathbf{Y})^{-1}\ \mathbf{Y}^T\mathbf{C}\circ\mathbf{R}^{-1}$$

Localisation of non-local observations

- For in-situ observations localisation on $\mathbf{B}$ and on $\mathbf{HBH}^T$ is fully equivalent.
- For non-local (remote sensing) observations the two approaches are different.
- For non-local observations and localisation on $\mathbf{B}$ and on $\mathbf{HBH}^T$ the following parameters must be prescribed:
  - The spatial location of the observation.
  - The λ_l used for in-situ observations may be not suitable as the scale of the footprint λ_o may be much larger than the correlation length scale λ_c.

It has been proposed to (Fertig et al. 2007)

- Increase λ_l to be at least as large as λ_o for non-local observations.

Note:

- There is the mathematical constrained that $\mathbf{C} \circ \mathbf{HBH}^T$ is positive definite.
- There is no such constraint on $\mathbf{C} \circ \mathbf{R}^{-1}$ for the LETKF.
- Using different localisation scales for different observation types will lead to inconsistencies.

Practical Implementations considered

Variants of localised ETKF :

- process observations in **patches**
- **EnSRF** Ensemble Square Root Filter
- **LETKF** following Hunt et al. 2007
- **VarETKF** following Buehner 2005

Variants of 3D-Var :

- **VarETKF** following Buehner 2005 (again)
- **3D-Var**: use parameterised **B**

Process observations in patches

- For linear H the analysis step may be split into multiple steps without changing the final result:
  1. Derive the Ensemble deviations $\mathbf{Y}$ in observation space from the background ensemble deviations $\mathbf{X}$.
  2. Use only a subset of the observations to derive the analysis ensemble deviations $\mathbf{Z}$ from $\mathbf{Y}$ and $\mathbf{X}$.
  3. replace the background ensemble $\mathbf{X}$ by the $\mathbf{Z}$, repeat step 1,2,3 with another subset.
- Processing patches with a limited number of observations at a time turns the large problem into a number of smaller problems this may be utilised for parallelisation.
- The 4D-EnKF (to be used for COSMO) makes use of the opposite relationship: Observations made at different times max be processed together at a later time.

EnSRF: Ensemble Square Root Filter

- The EnSRF (Anderson 2003, Anderson and Collins 2007) also processes the observations in patches.
  In the extreme case: only one observation at a time.
- In contrast to the previous algorithm the EnSRF does not recalculate the $\mathbf{Y}$ from the $\mathbf{X}$ in each iteration, but directly updates the $\mathbf{Y}$:
  1. Derive the Ensemble deviations $\mathbf{Y}$ in observation space from the background ensemble deviations $\mathbf{X}$.
  2. Use only a subset of the observations to derive the analysis ensemble deviations $\mathbf{Z}$ from $\mathbf{Y}$ and $\mathbf{X}$.
     At the same time update the $\mathbf{Y}$ by a similar procedure.
  3. replace the background ensemble $\mathbf{X}$ by the $\mathbf{Z}$, repeat steps 2,3 with another subset.
- Consequences:
  - $H(\mathbf{x})$ is applied only once.
  - Localisation is performed on $\mathbf{HBH}^T$.

LETKF (Hunt et al. 2007)

- The LETKF makes independent analyses for every grid-point using only observations within a certain localisation radius.
- Consequences:
  - Localisation is performed on $\mathbf{R}^{-1}$.
  - Technically different localisation length scales may be used for different observation types.
    However, this may lead to inconsistencies.
- The algorithm can be accelerated considerably by calculating the weight matrices $(\mathbf{1} + \mathbf{Y}^T \mathbf{C} \circ \mathbf{R}^{-1} \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{C} \circ \mathbf{R}^{-1}$ on a coarser grid and then interpolate to the model grid.

VarETKF: Variational ETKF following Buehner 2005

- The algorithm utilises that
  $\mathbf{B} = \mathbf{C} \circ \mathbf{X}\mathbf{X}^T$
  may be reformulated as
  $\mathbf{B} = \sum_k \mathbf{X}'_k \mathbf{C}^{1/2} \mathbf{C}^{1/2\,T} \mathbf{X}'^{\,T}_k$.
  Here $\mathbf{X}'_k$ is a diagonal matrix consisting of the forecast ensemble deviations.
- If an operator representation of $\mathbf{C}^{1/2}$ (square root of the localisation matrix) is available, this $\mathbf{B}$ can be used in the usual 3D-Var framework.
- Consequences:
  - Localisation is performed on $\mathbf{B} = \mathbf{C} \circ \mathbf{X}\mathbf{X}^T$
  - The $\mathbf{Y}$ are not derived by a linear regression.
    Instead the full nonlinear $H(\mathbf{x})$ are used.
  - Technically localisation may be performed in any representation (spectral, wavelet) if the $\mathbf{X}'_k$ are represented respectively.

3D-Var: parameterise **B**

- Ensemble Assimilation Systems based on 3D/4D-Var
  - Run ensemble of assimilation cycle with disturbed observations and model.
  - Use parameterisation or model: $\mathbf{B} = \mathbf{B}(p)$.
    Fit the parameters p (variances, length-scales,...) to $\mathbf{X}\mathbf{X}^T$.

Pros and Cons – Criteria

Computational demands
Computational complexity ($\mathbf{H}$ required, $H(\mathbf{x})$ iterated)
Consistent application of localisation
Localisation in $\triangle x$ only
Usage of nonlinear H in analysis

Computational demands

CPU-time requirements:

- The analysis step must be fast (few number of minutes) as is has to be cycled in real time. (cycle period down to 15 min ?)
- The requirements of the LETKF and the EnSRF have been compared by Whitaker (2008) :
  The CPU-time requirements are comparable, but depend on number of grid-points, number of observations, etc.
  The enhancement of the LETKF by performing analyses on a coarser grid has not been taken into account.
- CPU-time requirements of other algorithms are not considered in detail, but are probably larger.

Computational complexity

Computational complexity:

- The current 4D-EnKF design for COSMO separates :
  - Application of the observation operators $H(\mathbf{x})$ at the appropriate time in the model.
  - Performing the analysis at a later time in the LETKF.

  Consequently algorithms which repeatedly apply $H(\mathbf{x})$ cannot be used without changing the design.

Application of localisation

- Consistent application of localisation
  If localisation is not applied on $\mathbf{B}$ but on $\mathbf{HBH}^T$ or on $\mathbf{R}^{-1}$ there is the risk of inconsistencies in the algorithm, especially if different localisation length scales are applied for different observation types.
- Optimal application of localisation
  If the localisation function is applied merely in physical space it may be sub-optimal, especially for observations which measure integrated quantities. Algorithms which do not rely on localisation in space may have advantages.

Pros and Cons

algorithm to be used for localisation	patches $\mathbf{B}$	EnSRF $\mathbf{HBH}^T$	LETKF COSMO $\mathbf{R}^{-1}$	Var-EnKF GME/ICON $\mathbf{B}$	3D-Var $\mathbf{B}(p)$
requirements					
requires $\mathbf{H}$	no	no	no	yes	yes
iterates $H(\mathbf{x})$	yes	no	no	yes	yes
requires $\mathbf{B}(p)$	no	no	no	no	yes
is fast	no?	?	yes	no?	no?
functionality					
consistent $\mathbf{C} \circ \mathbf{HBH}^T$	yes	?	no	yes	yes
localisation in $\triangle x$	yes	yes	yes	no	no
nonlinear H	no	no	no	yes	yes

Conclusions ?

- We cannot get everything at the same time :
  A fast algorithm without drawbacks
  that does not conflict with the 4D-EnKF-Design.
- LETKF may be replaced by the EnSRF without changing the design.
  Not clear if EnSRF cures the localisation problems.
- Wait for first experiences with the COSMO-LETKF.
  We currently do not know relevant parameters of the setup.
  Localisation strength has to compromise noise and balance.
- Wait for first experiences with the GME-VarETKF.

.

References

Jeffrey L. Anderson and Nancy Collins.
Scalable implementations of ensemble filter algorithms for data assimilation.
Journal of Atmospheric and Oceanic Technology, 24:1452–1463, 2007.

Jeffrey L. Anderson.
A local least squares framework for ensemble filtering.
Mon. Wea. Rev., 131:634–642, 2003.

Mark Buehner and Martin Charron.
Spectral and spatial localization of background-error correlations for data assimilation.
Q.J,R.Meteorol.Soc, 133:615–630, 2007.

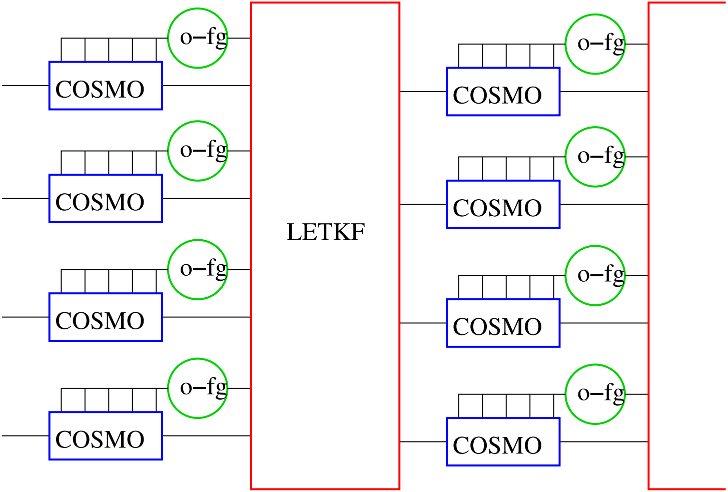
Mark Buehner.
Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting.
Q.J,R.Meteorol.Soc, 131:1013–1043, 2005.

Elana J. Fertig, Brian R. Hunt, Edward Ott, and Istvan Szunyogh.
Assimilating non-local observations with a local ensemble Kalman filter.
Tellus, 59A:719–730, 2007.

B. R. Hunt, E. J. Kostelich, and I. Szunyogh.
Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter.
Physica D, 230:112–126, 2007.

Jeff Whitaker.
Computational issues: An EnKF perspective.
Talk at the WWRP/THORPEX workshop on 4D-Var and Ensemble Kalman Filter inter-comparisons, Buenos Aires - Argentina, 2008.

EnKF for COSMO: 4D-EnKF



EnKF for GME/ICON: hybrid 3D-Var/EnKF (VarETKF)