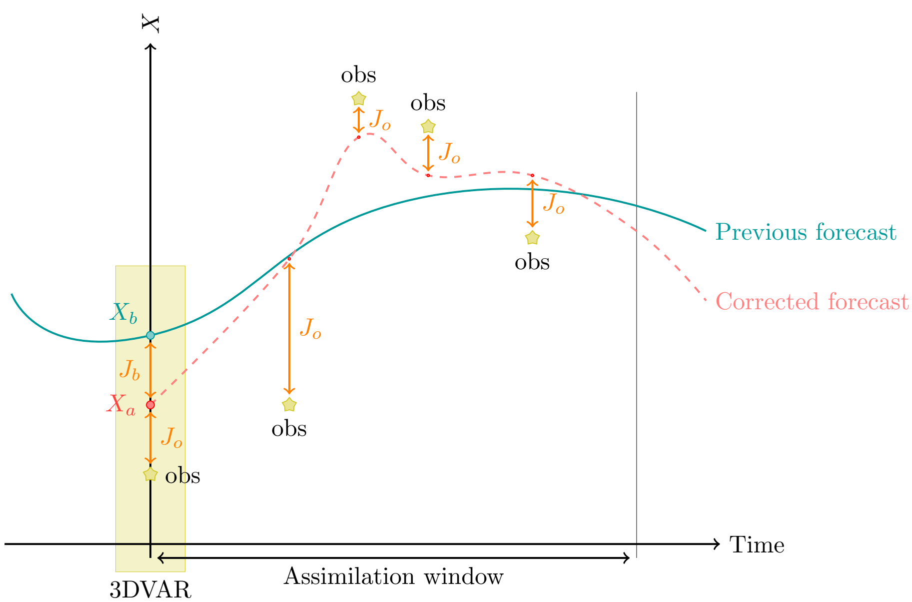
Hybrid Data Assimilation

Mark Asch - CSU/IMU/2023



Outline of the course (I)

Adjoint methods and variational data assimilation (4h)

1. Introduction to data assimilation: setting, history, overview, definitions.

2. Adjoint method.

3. Variational data assimilation methods:

   (a) 3D-Var,
   (b) 4D-Var.

4. EnsVar - Hybrid Ensemble Variational DA

Outline of the course (II)

Statistical estimation, Kalman filters and sequential data assimilation (4h)

1. Introduction to statistical DA.
2. Statistical estimation.
3. The Kalman filter.
4. Nonlinear extensions and ensemble filters.

Reacll: Variational DA - formulation

- In variational data assimilation we describe the state of the system by
  - ⇒ a state variable, $\mathbf{x}(t) \in \mathcal{X}$, a function of space and time that
  - ⇒ represents the physical variables of interest, such as current velocity (in oceanography), temperature, sea-surface height, salinity, biological species concentration, chemical concentration, etc.

- Evolution of the state is described by a system of (in general nonlinear) differential equations in a region Ω,

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathcal{M}(\mathbf{x}) \quad \text{in } \Omega \times [0, T], \\ \mathbf{x}(t=0) = \mathbf{x}_0, \end{cases} \tag{1}$$

where the initial condition is unknown (or poorly known).

- Suppose that we are in possession of observations $\mathbf{y}(t) \in \mathcal{O}$ and an observation operator $\mathcal{H}$ that describes the available observations.

- Then, to characterize the difference between the observations and the state, we define the objective (or cost) function,

$$J(\mathbf{x}_0) = \frac{1}{2}\int_0^T \|\mathbf{y}(t) - \mathcal{H}\left(\mathbf{x}(\mathbf{x}_0, t)\right)\|_{\mathcal{O}}^2 \,\mathrm{d}t + \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}\right\|_{\mathcal{X}}^2 \tag{2}$$

where

⇒ $\mathbf{x}^{\mathrm{b}}$ is the background (or first guess)

⇒ and the second term plays the role of a regularization (in the sense of Tikhonov—see previous Lecture.

⇒ The two norms under the integral, in the finite-dimensional case, will be represented by the error covariance matrices $\mathbf{R}$ and $\mathbf{B}$ respectively, and will be described below.

$\Rightarrow$ Note that for mathematical rigor we have indicated, as subscripts, the relevant functional spaces on which the norms are defined.

- In the continuous context, the data assimilation problem is formulated as follows:

Find the analyzed state $\mathbf{x}_0^{\mathrm{a}}$ that minimizes J and satisfies

$$\mathbf{x}_0^{\mathrm{a}} = \operatorname{argmin} J(\mathbf{x}_0).$$

- The necessary condition for the existence of a (local) minimum is (as usual...)

$$\nabla J(\mathbf{x}_0^{\mathrm{a}}) = 0.$$

Variational DA - 3D Var

- Finite-dimensional version of the cost function (2),

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right) \tag{3}$$

$$+\frac{1}{2}\left(\mathbf{H}\mathbf{x}-\mathbf{y}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{x}-\mathbf{y}\right), \tag{4}$$

  where

  ⇒ $\mathbf{x}$, $\mathbf{x}^{\mathrm{b}}$, and $\mathbf{y}$ are the state, the background state, and the measured state respectively;

  ⇒ $\mathbf{H}$ is the observation matrix (a linearization of the observation operator $\mathcal{H}$);

  ⇒ $\mathbf{R}$ and $\mathbf{B}$ are the observation and background error covariance matrices respectively.

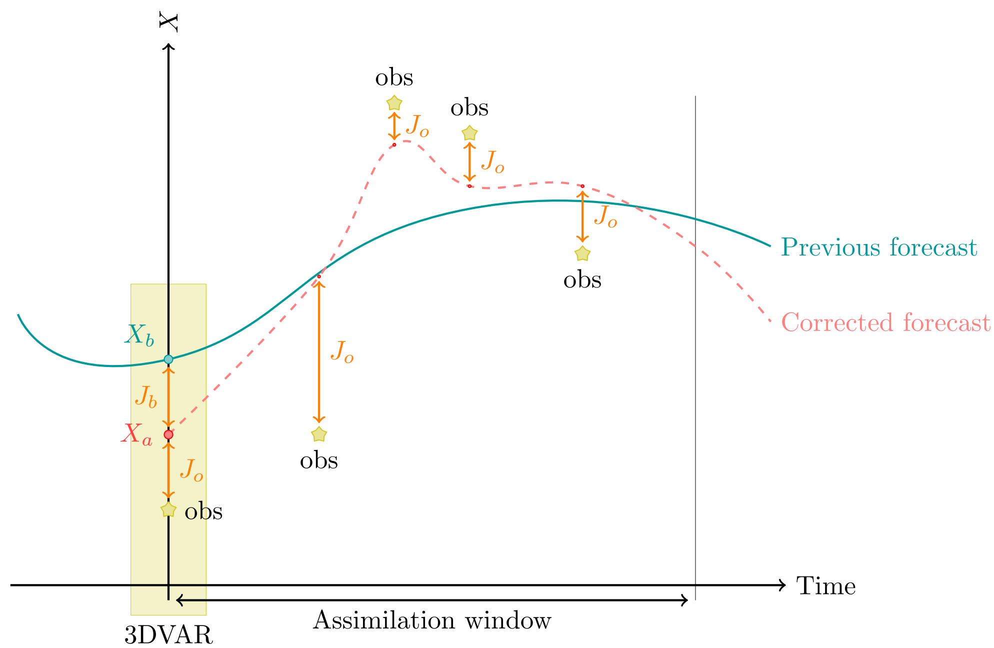
- This quadratic function attempts to strike a balance between some *a priori* knowledge about a background (or historical) state and the actual measured, or observed, state.

- It also assumes that we know and that we can invert the matrices $\mathbf{R}$ and $\mathbf{B}$. This, as we will be pointed out below, is not always obvious.

- Furthermore, it represents the sum of the (weighted) background deviations and the (weighted) observation deviations. The basic methodology is presented in the Algorithm below, which is nothing more than a classical gradient descent algorithm.

Variational DA - 3D Var Algorithm

$j = 0, \;\; x = x_0$
while $\|\nabla J\| > \epsilon$ or $j \le j_{\max}$
    compute J
    compute ∇J
    gradient descent and update of x_{j+1}
    $j = j + 1$
end



Variational DA - 4D Var

- A more realistic, but complicated situation arises when one wants to assimilate observations that are acquired over a time interval, during which the system dynamics (flow, for example) cannot be neglected.

- Suppose that the measurements are available at a succession of instants, $t_k,\ k = 0, 1, \ldots, K$ and are of the form

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\epsilon}_k^{\mathrm{o}}, \tag{5}$$

  where

  ⇒ $\mathbf{H}_k$ is a linear observation operator and

  ⇒ $\boldsymbol{\epsilon}_k^{\mathrm{o}}$ is the observation error with covariance matrix $\mathbf{R}_k$,

  ⇒ and suppose that these observation errors are uncorrelated in time.

- Now we add the dynamics described by the state equation,

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}_k, \tag{6}$$

where we have neglected any model error.[1]

- We suppose also that at time index $k = 0$ we know

$\Rightarrow$ the background state $\mathbf{x}_0^{\mathrm{b}}$ and
$\Rightarrow$ its error covariance matrix $\mathbf{P}_0^{\mathrm{b}}$
$\Rightarrow$ and we suppose that the errors are uncorrelated with the observations in (5).

- Then a given initial condition, $\mathbf{x}_0$, defines a unique model solution $\mathbf{x}_{k+1}$ according to (6).

- We can now generalize the objective function (3),

[1] This will be taken into account below.

which becomes

$$J(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}} \left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \tag{7}$$

$$+ \frac{1}{2}\sum_{k=0}^{K} \left(\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1} \left(\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k\right). \tag{8}$$

- The minimization of $J(\mathbf{x}_0)$ will provide the initial condition of the model that fits the data most closely.

- This analysis is called "strong constraint four-dimensional variational assimilation," abbreviated as *strong constraint 4D-Var*. The term "strong constraint" implies that the model found by the state equation (6) must be exactly satisfied by the sequence of estimated state vectors.

- In the presence of model uncertainty, the state equation becomes

$$\mathbf{x}^{\mathrm{t}}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}^{\mathrm{t}}_{k} + \boldsymbol{\eta}_{k+1}, \tag{9}$$

where

$\Rightarrow$ the model noise $\boldsymbol{\eta}_k$ has covariance matrix $\mathbf{Q}_k$,

$\Rightarrow$ which we suppose to be uncorrelated in time and uncorrelated with the background and observation errors.

- The objective function for the best, linear unbiased estimator (BLUE) for the sequence of states

$$\{\mathbf{x}_k,\ k = 0, 1, \ldots, K\}$$

is of the form

$$
\begin{aligned}
J(\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_K) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right)^{\mathrm{T}}\left(\mathbf{P}_0^{\mathrm{b}}\right)^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) \\
+\frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right) \quad (10) \\
+\frac{1}{2}\sum_{k=0}^{K-1}\left(\mathbf{x}_{k+1} - \mathbf{M}_{k+1}\mathbf{x}_k\right)^{\mathrm{T}}\mathbf{Q}_{k+1}^{-1}\left(\mathbf{x}_{k+1} - \mathbf{M}_{k+1}\mathbf{x}_k\right).
\end{aligned}
$$

- This objective function has become a function of the complete sequence of states

$$
\{\mathbf{x}_k,\ k = 0, 1, \ldots, K\},
$$

and its minimization is known as "weak constraint four-dimensional variational assimilation," abbreviated as *weak constraint 4D-Var*.

- Equations (7) and (10), with an appropriate reformulation of the state and observation spaces, are special cases of the BLUE objective function.

Ensemble KF - the Three Steps

1. **Initialization:** generate an ensemble of m random states $\left\{\mathbf{x}_{i,0}^{\mathrm{f}}\right\}_{i=1,\ldots,m}$ at time $t=0.$

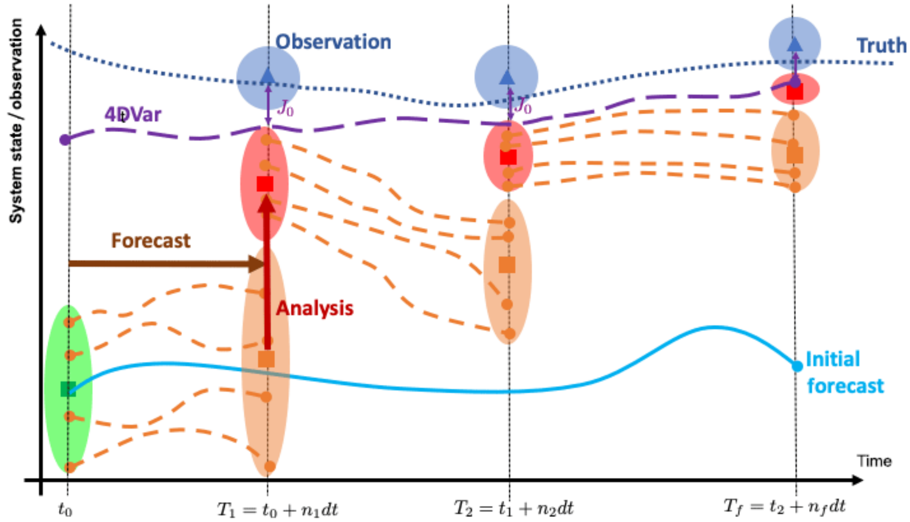
2. **Forecast:** compute the prediction for each member of the ensemble.

3. **Analysis:** correct the prediction in light of the observations.

- **Notes**:
  1. Propagation can equivalently be performed either at the end of the analysis step or at the beginning of the forecast step.
  2. The Kalman gain is not computed directly, but estimated from the ensemble statistics.
  3. With the important exception of the Kalman gain computation, all operations on the ensemble members are independent. As a result, parallelization is straightforward.

4. This is one of the main reasons for the success/popularity of the EnKF.

Comparison: EnKf and 4D-Var



- Principle of data assimilation: Having a physical model able to forecast the evolution of a system from time $t = t_0$ to time$t = T_f$ (cyan curve), the aim of DA is to use available observations (blue triangles) to correct the model projections and get closer to the (unknown) truth (dotted line).

- In EnKFs, the initial system state and its uncertainty (green square and ellipsoid) are represented by N_e members.

- $\Rightarrow$ The members are propagated forward in time during n_1 model time steps dt to $t = T_1$ where observations are available (forecast phase, orange dashed lines).
- $\Rightarrow$ At $t = T_1$ the analysis uses the observations and their uncertainty (blue triangle and ellipsoid) to produce a new system state that is closer to the observations and with a lower uncertainty (red square and ellipsoid).
- $\Rightarrow$ A new forecast is issued from the analysed state and this procedure is repeated until the end of the assimilation window at $t = T_f$.
- $\Rightarrow$ The model state should get closer to the truth and with lower uncertainty as more observations are assimilated.

- Time-dependent variational methods (4D-Var) iterate over the assimilation window to find the trajectory that minimises the misfit (J_0) between the model and all observations available from t_0 toT_f (violet curve).

- For linear dynamics, Gaussian errors and infinite ensemble sizes, the states produced at the end of the assimilation window by the two methods should be equivalent (Li and Navon, 2001).

Hybrid - EnsVar - DA

- The term hybrid DA refers to a system where two DA methods run concurrently, exchanging information about errors and estimated model states, to obtain improved estimations of these.

- For challenging DA problems, neither pure EnKF nor pure 4D-Var can do the job.

- These methods have cross-fertilized to combine the benefits of variational and ensemble Kalman methods, and at the same time, overcome the limitations of the individual methods.

- Fullest details of the different hybridizations are beyond the scope of this lecture, and [Asch2016] should be consulted for full explanations of the various options as well as references.

EnsVar - Motivation

- Here, we look into the many ways to combine the benefits of variational methods and of the ensemble Kalman approaches.

- The focus has mainly been in the domain of numerical weather prediction (NWP).

- Combining the advantages of these methods, one also wishes to avoid some of the drawbacks of both classes of methods.

  ⇒ From a theoretical standpoint, the EnKF propagates the errors and has a dynamical, flow-dependent, representation of those errors. It also does not require the tangent linear and adjoint model of the observation operator, as seen in the Basic Course.

  ⇒ On the other hand, 4D-Var by definition operates on a time data assimilation window over which asynchronous observations can consistently, i.e. model-wise, be assimilated. Moreover, 4D-Var can perform a full nonlinear analysis within its data assimilation window thanks to numerical optimization techniques.

- On the downsides,

  ⇒ the EnKF requires the use of regularization techniques, inflation and localization specifically, to filter out sampling errors and address the rank-deficiency issue of the ensemble (collapse...).

  ⇒ The 4D-Var requires the use of the tangent linear and adjoint models (evolution and observation) which are very time-consuming to derive and maintain.

Hybrid Methods

- Hybrid EnKF with 3D-Var (EnsVAR)
- Ensemble of variational DAs (EDA)
- Hybrid EnKF with 4D-Var, known as 4DEnVar
- Iterative ensemble Kalman smoother (IEnKS), derived from Bayes' Law.

Hybrid EnKF with 3D-Var

- The 3D-Var system relies on a full rank static (and/or predetermined) error covariance matrix $\mathbf{C}$.

- The EnKF estimates the flow-dependent errors through the perturbations $\mathbf{P}^{\mathrm{f}} = \frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}} = \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}$, where $\mathbf{X}_{\mathrm{f}}$ is the normalized perturbation matrix (see Ensemble Kalman Filter Lecture).

- Whatever the type of EnKF, the simplest idea is to perform a state analysis using a linear combination of these error covariances,

$$\mathbf{B} = \gamma\mathbf{C} + (1-\gamma)\mathbf{P}^{\mathrm{f}},$$

  where $\gamma \in [0, 1]$ is a scalar parameter that controls the blending of the covariances:

$\Rightarrow$ using $\mathbf{B}$ with $\gamma = 1$ corresponds to a 3D-Var state analysis while

$\Rightarrow$ using $\mathbf{B}$ with $\gamma = 0$ corresponds to the pure EnKF state analysis.

- If the EnKF is stochastic, then one can update each member $i = 1, \ldots, m$ of the ensemble using a state analysis with $\mathbf{B}$, which essentially solves the following variational problem,

$$\mathcal{L}_i(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} + \boldsymbol{\epsilon}_i - \mathcal{H}(\mathbf{x})\|^2_{\mathbf{R}} + \frac{1}{2} \|\mathbf{x} - \mathbf{x}_i\|^2_{\mathbf{B}} ,$$

where $\mathbf{x}_i$ is the first-guess of member i.

- We recall that $\|\mathbf{z}\|^2_{\mathbf{B}} = \mathbf{z}^{\mathbf{T}}\mathbf{B}^{-1}\mathbf{z}$.

- The elegance of the scheme lies in the fact that it yields a statistically consistent update of the ensemble thanks to the stochastic representation of errors.

- This EnKF-3D-Var scheme was shown to improve the performance of data assimilation

$\Rightarrow$ over the EnKF when the ensemble is small, and

$\Rightarrow$ over 3D-Var when the observation network is not dense enough. Of course, this requires a proper tuning of γ.

- If the EnKF is deterministic, the construction of the hybrid is not as straightforward, especially concerning the update of the perturbation ensemble—please consult [Asch2016] and references therein.

Ensemble of variational DAs

- Idea: Process an *ensemble of data assimilations* or *EDA*.
  - ⇒ More generally an EDA system denotes a data assimilation system that processes several variational analyses in parallel.
  - ⇒ The goal is to introduce some flow-dependence in the 4D-Var operational schemes that were initially only based on the static background error covariance matrix.
  - ⇒ This scheme actually mimics closely the stochastic EnKF and, even more to the point, the hybrid EnKF-3D-Var scheme described above.

- The main idea is to maintain an ensemble of Var, which will be assumed to be a 4D-Var in the following.
  - ⇒ This is usually numerically costly for 4D-Var's

and may require the degrading of the model resolution.

- ⇒ Each analysis uses the same background error covariance matrix $\mathbf{B}$, which may have been obtained from the sampled covariances whose variances have been properly filtered and whose correlations have been regularized, as well as the static background covariances.
- ⇒ Just as in the stochastic EnKF, it is necessary for each analysis to have perturbed observations so as to maintain statistical consistency.
- ⇒ The strong-constraint 4D-Var cost function for each analysis $i = 1, \ldots, m$ has the form

$$\begin{aligned}\mathcal{L}_i(\mathbf{x}_0) = \frac{1}{2}\sum_{k=1}^{K} \left\| \mathbf{y}_k + \boldsymbol{\epsilon}_k^i - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0) \right\|_{\mathbf{R}_k}^2 \\ + \frac{1}{2}\left\| \mathbf{x}_0 - \mathbf{x}_0^i \right\|_{\mathbf{B}}^2,\end{aligned}$$

where $\mathcal{M}_{k:0}$ is the resolvent of the forecast model from t_0 to t_k, $\mathcal{H}_k$ is the observation operator at t_k, $\boldsymbol{\epsilon}_k^i$ is the random noise added to observation $\mathbf{y}_k$ and is related to the i-th member analysis. The symbol $\circ$ stands for the composition operator.

- $\Rightarrow$ This generates an ensemble of updates for $\mathbf{x}_0^i$ similarly to the stochastic EnKF.
- $\Rightarrow$ It is also possible to perturb each member of the ensemble in the forecast step so as to account for (parametric or not) model error.
- $\Rightarrow$ Hence, the ensemble will be instrumental in accounting for flow-dependence and model error.

Hybrid EnKF with 4D-Var (4DEnVar)

- The future of 4DVar as an operational tool is uncertain because of its

  ⇒ poor scalability and of the
  ⇒ cost of the adjoint model maintenance.

- The 4DEnVar method has emerged as a way to circumvent the development of the adjoint of the dynamical model.

  ⇒ The key idea is based on the ability of the EnKF to estimate the sensitivities of the observation to the state variables using the full observation model in place of the tangent linear one and of its adjoint in the computation of the Kalman gain.

- Many variants of the 4DEnVar are possible depending on the way the perturbations are generated, or if the adjoint model is available or not.

- Full 4DEnVar operational systems are now implemented or are in the course of being so.

- For implementation details, please consult [Asch2016].

Iterative ensemble Kalman smoother (IEnKS)

- Most of these EnVar methods, with the noticeable exception of the more firmly grounded EDA ones (see above), have been designed heuristically blending theoretical insights and operational constraints. This led to many variants of the schemes, even when this is not mathematically justified. Most of these ideas stemmed from the variational DA community.

- By contrast, the iterative ensemble Kalman smoother (IEnKS, Bocquet and Sakov, 2014), is a four-dimensional EnVar method that is derived from Bayes' rule and where all approximations are understood at a theoretical level.

- It comes from ensemble-based DA and, specifically, extends the iterative ensemble Kalman filter (Sakov et al., 2012b) to a full data assimilation window (DAW) as in 4DVar. The name smoother reminds us that the method smooths trajectories like 4DVar. However, it can equally be used for smoothing and filtering.

- Basically, the IEnKS can be seen as an EnKF, for which each analysis corresponds to a nonlinear 4DVar analysis but within the reduced subspace defined by the ensemble.

- Hence, the associated cost function is of the form

$$\mathcal{J}(\mathbf{w}) = \sum_{k=L-S+1}^{L} \frac{1}{2} \left\| \mathbf{y}_k - \mathcal{F}_{k:0}\overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w} \right\|_{\mathbf{R}_k}^2 + \frac{1}{2} \left\| \mathbf{w} \right\|^2.$$

where

⇒ $\mathcal{F}_{k:0}$ stands for the composition of $\mathcal{H}_k$ and the resolvent $\mathcal{M}_{k:0}$.
⇒ L is the length of the DAW
⇒ S is he length of the forecast between cycles, in units of $t_{k+1} - t_k$

- Because the IEnKS catches the best of 4DVar (nonlinear analysis) and EnKF (flow-dependence of the error statistics), both these parameters could be critical.

- The minimization of $\mathcal{J}$ can be performed in the ensemble subspace using any nonlinear optimization method, such as Gauss-Newton , Levenberg-Marquardt or trust-region methods

- In chaotic systems, the IEnKS outperforms any reasonably scalable DA method in terms of accuracy.

  $\Rightarrow$ By construction, it outperforms the EnKF, the EnKS and 4DVar for smoothing but also filtering.

EnVar Techniques - table

Credit: Carrassi, Bertino, Bocquet, Evensen. 2018

Table 1: Comparison of EnVar data assimilation techniques. This table answers the following questions: (i) Is the analysis based on a linear or nonlinear scheme? (ii) Is the adjoint of the evolution model required? (iii) Is the adjoint of the observation operator required? (iv) Is the background flow-dependent? (v) Are the updated perturbations stochastic or deterministic? (vi) Are the updated perturbations fully consistent with the analysis, i.e., are they a faithful representation of the analysis uncertainty? (vii) Is localization of the ensemble analysis required? (viii) Is a static background used? To some extent, all algorithms can accommodate a static background; the answer tells whether the published algorithm has a static background. Blank answers correspond to irrelevant questions.

algorithm	analysis type	evol. model adjoint required?	obs. operator adjoint required?	background flow-dependence?	sto. or det. perturbations?	consistent perturbations?	localization required?	static background
EnKF	linear		no	yes	both	yes	yes	no[4]
3DVar	nonlinear		yes	no				yes
4DVar	nonlinear	yes	yes	no				yes
EDA with 4DVar	nonlinear	yes[1]	yes[1]	yes	sto.	yes	part.	yes[3]
4DEnVar	linear	no	no	yes	sto.	no[2]	yes	yes[3]
IEnKS	nonlinear	no	no	yes	det.	yes	yes	no[4]
MLEF	nonlinear		no	yes	det.	yes	yes	no[4]
4D-ETKF	linear	no	no	yes	det.	yes	yes	no[4]

[1] The adjoint models could be avoided considering an EDA of 4DEnVar.

[2] It depends on the implementation of 4DEnVar; the perturbation are often generated by a concomitant EnKF.

[3] With an hybridization of the covariances.

[4] But possible with an hybridization of the covariances.

EXAMPLES

Codes

Various open-source repositories and codes are available for both academic and operational data assimilation.

1. DARC: `https://research.reading.ac.uk/met-darc/` from Reading, UK.

2. DAPPER: `https://github.com/nansencenter/DAPPER` from Nansen, Norway.

3. DART: `https://dart.ucar.edu/` from NCAR, US, specialized in ensemble DA.

4. OpenDA: `https://www.openda.org/`.

5. Verdandi: `http://verdandi.sourceforge.net/` from INRIA, France.

6. PyDA: `https://github.com/Shady-Ahmed/PyDA`, a Python implementation for academic use.

7. Filterpy: `https://github.com/rlabbe/filterpy`, dedicated to KF variants.

8. EnKF; `https://enkf.nersc.no/`, the original Ensemble KF from Geir Evensen.

References

1. K. Law, A. Stuart, K. Zygalakis. *Data Assimilation. A Mathematical Introduction*. Springer, 2015.

2. G. Evensen. *Data assimilation, The Ensemble Kalman Filter*, 2nd ed., Springer, 2009.

3. A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* SIAM. 2005.

4. O. Talagrand. Assimilation of observations, an introduction. *J. Meteorological Soc. Japan*, **75**, 191–209, 1997.

5. F.X. Le Dimet, O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus,* **38**(2), 97–110, 1986.

6. J.-L. Lions. Exact controllability, stabilization and perturbations for distributed systems. *SIAM Rev.*, **30**(1):1–68, 1988.

7. J. Nocedal, S.J. Wright. *Numerical Optimization*. Springer, 2006.

8. F. Tröltzsch. *Optimal Control of Partial Differential Equations*. AMS, 2010.