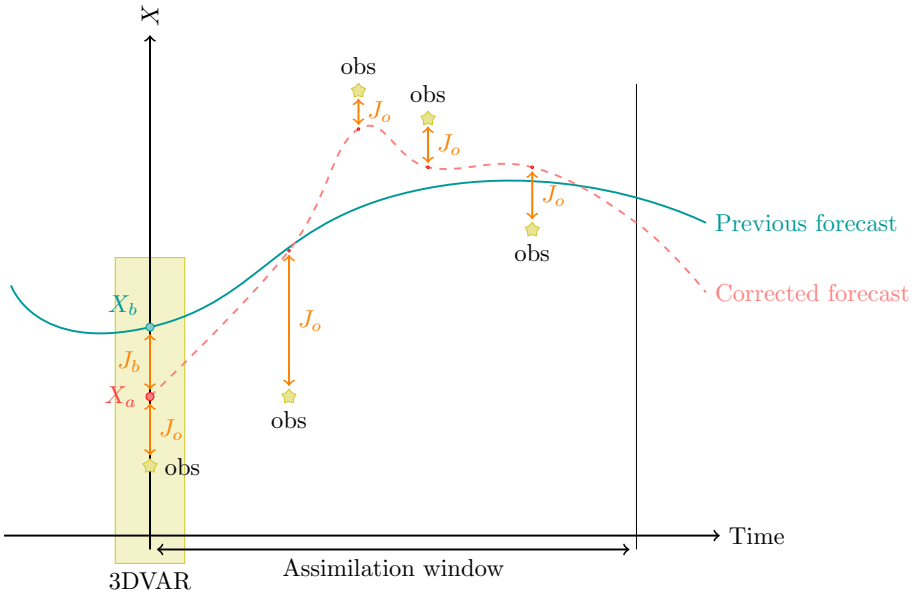
Statistical Data Assimilation

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. Optimization methods.

3. Adjoint method.

4. Variational data assimilation methods:

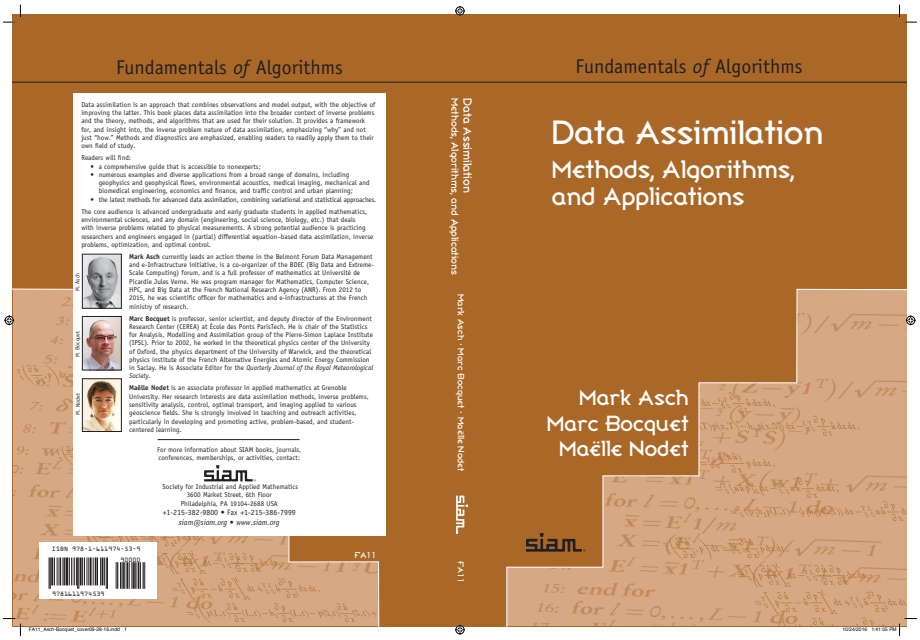
   (a) 3D-Var,
   (b) 4D-Var.

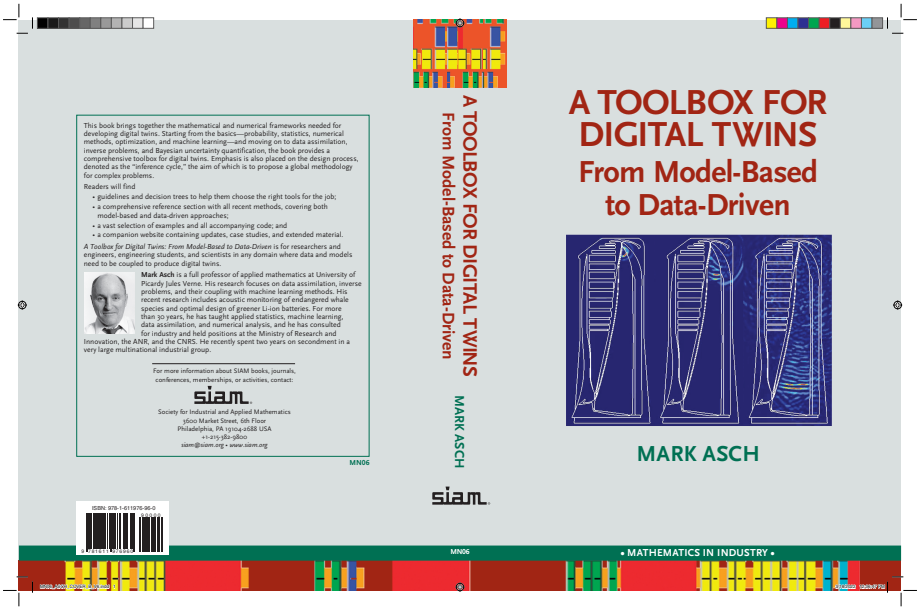
Outline of the course (II)

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.

Reference Textbooks





Statistical DA: introduction

- Now we will generalize the variational approach to deal with errors and noise in

  $\Rightarrow$ the models,
  $\Rightarrow$ the observations and
  $\Rightarrow$ the initial conditions.

- The variational results could of course be derived as a special case of statistical DA, in the limit where the noise disappears.

- Even the statistical results can be derived in a very general way, using SDEs and/or Bayesian analysis, and then specialized to the various Kalman-type filters that we will study here.

- Practical inverse problems and data assimilation problems involve measured data.

$\Rightarrow$ These data are inexact and are mixed with random noise.

$\Rightarrow$ Only statistical models can provide rigorous, effective means for dealing with this measurement error.

Statistical DA: a "simple" example

We want to estimate a scalar quantity, say the temperature or the ozone level, at a fixed point in space.

Suppose we have:

- a model forecast, x^{b} (background, or *a priori* value)
- and a measured value, x^{obs} (observation).

The simplest possible approach is to try a linear combination of the two,

$$x^{\mathrm{a}} = x^{\mathrm{b}} + w(x^{\mathrm{obs}} - x^{\mathrm{b}}),$$

where x^{a} denotes the analysis that we seek and $0 \leq w \leq 1$ is a weight factor. We subtract the (always unknown) true state x^{t} from both sides,

$$x^{\mathrm{a}} - x^{\mathrm{t}} = x^{\mathrm{b}} - x^{\mathrm{t}} + w(x^{\mathrm{obs}} - x^{\mathrm{t}} - x^{\mathrm{b}} + x^{\mathrm{t}})$$

and defining the three errors (analysis, background, observation) as

$$e^{\mathrm{a}} = x^{\mathrm{a}} - x^{\mathrm{t}}, \quad \mathrm{e}^{\mathrm{b}} = x^{\mathrm{b}} - x^{\mathrm{t}}, \quad e^{\mathrm{obs}} = x^{\mathrm{obs}} - x^{\mathrm{t}},$$

we obtain

$$e^{\mathrm{a}} = e^{\mathrm{b}} + w(e^{\mathrm{obs}} - e^{\mathrm{b}}) = we^{\mathrm{obs}} + (1-w)e^{\mathrm{b}}.$$

If we have many realizations, we can take an ensemble average, or expectation, denoted by $\langle \cdot \rangle$,

$$\langle e^{\mathrm{a}} \rangle = \langle e^{\mathrm{b}} \rangle + w(\langle e^{\mathrm{obs}} \rangle - \langle e^{\mathrm{b}} \rangle).$$

Now if these errors are centred (have zero mean, or the estimates of the true state are unbiased), then

$$\langle e^{\mathrm{a}} \rangle = 0$$

also. So we must look at the variance and demand that it be as small as possible. The variance is defined,

using the above notation, as

$$\sigma^2 = \left\langle \left(e - \langle e \rangle\right)^2 \right\rangle .$$

Now, taking variances of the error equation, and using the zero-mean property, we obtain

$$\sigma_{\rm a}^2 = \sigma_{\rm b}^2 + w^2 \left\langle \left(e^{\rm obs} - e^{\rm b}\right)^2 \right\rangle + 2w \left\langle e^{\rm b} \left(e^{\rm obs} - e^{\rm b}\right) \right\rangle .$$

This reduces to

$$\sigma_{\rm a}^2 = \sigma_{\rm b}^2 + w^2 \left(\sigma_{\rm o}^2 + \sigma_{\rm b}^2\right) - 2w\sigma_{\rm b}^2$$

if $e^{\rm o}$ and $e^{\rm b}$ are uncorrelated.

Now, to compute a minimum, take the derivative with respect to w and equate to zero, to obtain

$$0 = 2w \left(\sigma_{\rm obs}^2 + \sigma_{\rm b}^2\right) - 2\sigma_{\rm b}^2,$$

where we have ignored all cross terms (errors are

assumed independent). Finally, solving this last equation, we can write the optimal weight,

$$w_* = \frac{\sigma_{\rm b}^2}{\sigma_{\rm obs}^2 + \sigma_{\rm b}^2} = \frac{1}{1 + \sigma_{\rm o}^2/\sigma_{\rm b}^2}$$

which depends on the ratio of the background and the observation errors. Clearly $0 \leq w_* \leq 1$ and

- if the observation is perfect, $\sigma_{\rm obs}^2 = 0$ and thus $w_* = 1$, the maximum weight;

- if the background is perfect, $\sigma_{\rm b}^2 = 0$ and $w_* = 0$, so the observation will not be taken into account.

We can now rewrite the analysis error variance as,

$$\begin{aligned}\sigma_{\mathrm{a}}^2 &= w_*^2\sigma_{\mathrm{obs}}^2 + (1-w_*)^2\sigma_{\mathrm{b}}^2 \\ &= \frac{\sigma_{\mathrm{b}}^2\sigma_{\mathrm{obs}}^2}{\sigma_{\mathrm{obs}}^2+\sigma_{\mathrm{b}}^2} \\ &= (1-w_*)\sigma_{\mathrm{b}}^2 \\ &= \frac{1}{\sigma_{\mathrm{obs}}^{-2}+\sigma_{\mathrm{b}}^{-2}},\end{aligned}$$

where we suppose that $\sigma_{\mathrm{b}}^2,\ \sigma_{\mathrm{o}}^2 > 0$. In other words,

$$\frac{1}{\sigma_{\mathrm{a}}^2} = \frac{1}{\sigma_{\mathrm{o}}^2} + \frac{1}{\sigma_{\mathrm{b}}^2}.$$

This is a very fundamental result, implying that the overall *precision*, $\tau = 1/\sigma^2$, (reciprocal of the variance) is the sum of the background and measurement precisions. Finally, the analysis equation becomes

$$x^{\mathrm{a}} = x^{\mathrm{b}} + \frac{1}{1+\alpha}(x^{\mathrm{obs}} - x^{\mathrm{b}}),$$

where $\alpha = \sigma^2_{\rm obs}/\sigma^2_{\rm b}$. This is called the BLUE - Best Linear Unbiased Estimator - because it gives an unbiased, optimal weighting for a linear combination of two independent measurements.

Statistical DA: 3 special cases and conclusions

We can isolate three special cases:

- if the observation is very accurate, $\sigma^2_{\rm obs} \ll \sigma^2_{\rm b}$, $\alpha \ll 1$ and thus $x^{\rm a} \approx x^{\rm obs}$
- if the background is accurate, $\alpha \gg 1$ and $x^{\rm a} \approx x^{\rm b}$
- and finally, if observation and background varaiances are approximately equal, $\alpha \approx 1$ and $x^{\rm a}$ is the arithmetic average of $x^{\rm b}$ and $x^{\rm obs}$.

Conclusion: this simple, linear model does indeed capture the full range of possible solutions in a statistically rigorous manner, thus providing us with an "enriched" solution when compared with a non-probabilistic, scalar response such as the arithmetic average of observation and background, which would correspond to only the last of the above three special cases.

KALMAN FILTERS

Kalman Filters - background and history

- DA is concerned with dynamic systems, where (noisy) observations are acquired over time.

- Question: Is there some statistically optimal way to combine the dynamic model and the observations?

- One answer is provided by Kalman filters

  $\Rightarrow$ They are linear models for state estimation of noisy dynamic systems.

  $\Rightarrow$ They have been the *de facto* standard in many robotics and tracking/prediction applications because they are well-suited for systems where there is uncertainty about an observable dynamic process.

  $\Rightarrow$ They are also the basis of many data assimilation systems.

⇒ They use a paradigm of "observe, predict, correct" to extract information from a noisy signal.

- The Kalman filter was invented[1] in 1960 by R. E. Kálmán to solve this sort of problem in a mathematically optimal way.

- Its first use was on the Apollo missions to the moon, and since then it has been used in an enormous variety of domains.

  ⇒ There are Kalman filters in aircraft and autonomous vehicles, on submarines, and, in cruise missiles.
  ⇒ Wall Street uses them to track the market.
  ⇒ They are used in robots, in IoT (Internet of Things) sensors, and in laboratory instruments.
  ⇒ Chemical plants use them to control and monitor reactions.
  ⇒ They are used to perform medical imaging and to remove noise from cardiac signals.
  ⇒ Weather forecasting is based on Kalman filters.

[1] Apparently, following a prior invention by Stratonovich, one year earlier.

$\Rightarrow$ They can effectively be used for modeling in epidemiology.

- In summary, if it involves a sensor and/or time-series data, a Kalman filter or a close relative of the Kalman filter is usually involved.

Kalman Filters - formulation

- Consider a dynamical system that evolves in time and we would like to estimate a series of *true* states, $\mathbf{x}_k^{\mathrm{t}}$ (a sequence of random vectors) where discrete time is indexed by the letter k.

- These times are those when the observations or measurements are taken, as shown in the Figure.

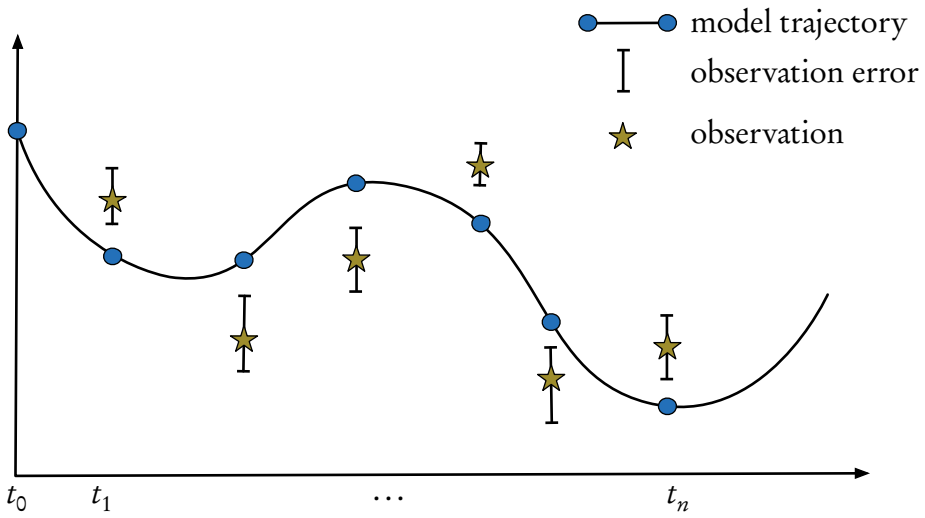


Figure 1: Sequential assimilation: a computed model trajectory, observations, and their error bars.

- The assimilation starts with an unconstrained model trajectory from $t_0, t_1, \ldots, t_{k-1}, t_k, \ldots, t_n$ and aims to provide an optimal fit to the available observations/measurements given their uncertainties (error bars).

  $\Rightarrow$ For example, in current, synoptic scale weather forecasts, $t_k - t_{k-1} = 6$ hours and is less for the convective scale.

  $\Rightarrow$ In robotics, or autonomous vehicles, the time intervals are of the order of the instrumental frequency, which can be a few milliseconds.

Kalman Filters - stochastic model

- We seek to estimate the state $\mathbf{x} \in \mathbb{R}^n$ of a discrete-time dynamic process that is governed by the linear stochastic difference equation

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

- with a measurement/observation $\mathbf{y} \in \mathbb{R}^m$,

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k. \tag{2}$$

- Note:

  $\Rightarrow$ $\mathbf{M}_{k+1}$ and $\mathbf{H}_k$ are considered linear, here.

  $\Rightarrow$ The random vectors, $\mathbf{w}_k$ and $\mathbf{v}_k$, represent the process/modeling and measurement/observation errors respectively.

$\Rightarrow$ They are assumed to be independent, white noise processes with Gaussian/normal probability distributions,

$$\begin{aligned} \mathbf{w}_k &\sim \mathcal{N}(0, \mathbf{Q}_k), \\ \mathbf{v}_k &\sim \mathcal{N}(0, \mathbf{R}_k), \end{aligned}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are the covariance matrices (supposed known) of the modeling and observation errors respectively.

- All these assumptions about unbiased and uncorrelated errors (in time and between each other) are not limiting, since extensions of the standard Kalman filter can be developed should any of these not be valid—see Advanced Course.

- We note that, for a broader mathematical view on the above system, we could formulate all of statistical DA in terms of stochastic differential equations (SDEs).

$\Rightarrow$ Then the theory of Itô provides a detailed solution of the problem of optimal filtering as well as rigorous existence and uniqueness results... see [Law, Sarkka].

Kalman Filters - sequential assimilation scheme

The typical assimilation scheme is made up of two major steps:

1. a prediction/forecast step, and

2. a correction/analysis step.

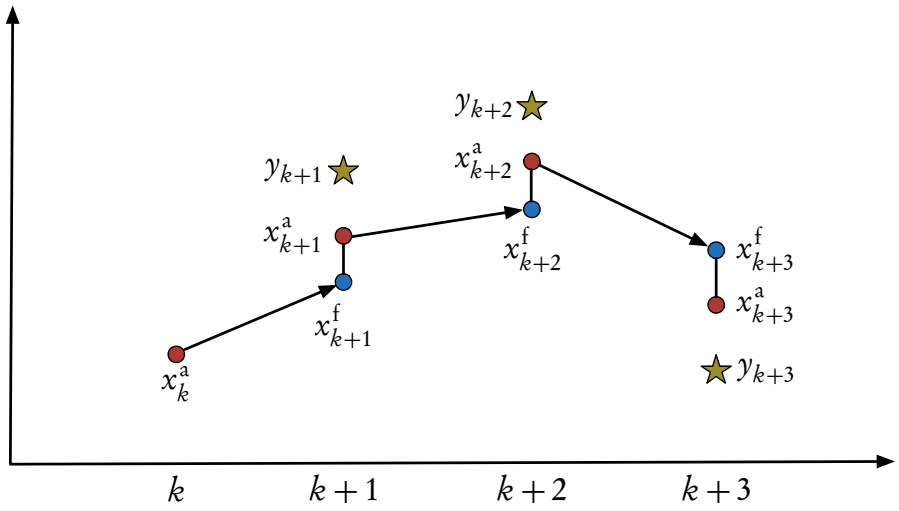


Figure 2: Sequential assimilation scheme for the Kalman filter. The x-axis denotes time, the y-axis denotes the values of the state and observations vectors.

- At time t_k we have the result of a previous forecast, $\mathbf{x}_k^{\mathrm{f}}$, (the analogue of the background state $\mathbf{x}_k^{\mathrm{b}}$) and the result of an ensemble of observations in $\mathbf{y}_k$.

- Based on these two vectors, we perform an analysis that produces $\mathbf{x}_k^{\mathrm{a}}$.

- We then use the evolution model to obtain a prediction of the state at time t_{k+1}.

- The result of the forecast is denoted $\mathbf{x}_{k+1}^{\mathrm{f}}$, and becomes the background, or initial guess, for the next time-step—see Figure 2.

- The Kalman filter problem can be resumed as follows:

  $\Rightarrow$ given a prior/background estimate $\mathbf{x}^{\mathrm{f}}$ of the system state at time t_k,
  $\Rightarrow$ what is the best update/analysis $\mathbf{x}_k^{\mathrm{a}}$ based on the currently available measurements $\mathbf{y}_k$?

Kalman Filters - the filter

- The goal of the Kalman filter is:

  $\Rightarrow$ to compute an optimal *a posteriori* estimate $\mathbf{x}_k^{\mathrm{a}}$

  $\Rightarrow$ that is a linear combination of an *a priori* estimate $\mathbf{x}_k^{\mathrm{f}}$ and a weighted difference between the actual measurement $\mathbf{y}_k$ and the measurement prediction $\mathbf{H}_k\mathbf{x}_k^{\mathrm{f}}$.

- This is none other than the BLUE that we have seen above.

- The filter is thus of the linear, recursive form

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^{\mathrm{f}} \right). \qquad (3)$$

  $\Rightarrow$ The difference $\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k^{\mathrm{f}}$ is called the *innovation* and reflects the discrepancy between the actual and the predicted measurements at time t_k.

- Note that, for generality, the matrices are shown with a time-dependence. When this is not the case, the subscripts k can be dropped. The *Kalman gain* matrix, $\mathbf{K}$, is chosen to minimize the *a posteriori* error covariance equation (4).

  $\Rightarrow$ We define forecast (*a priori*) and analysis (*a posteriori*) estimate errors as

$$\begin{aligned} \mathbf{e}_k^{\mathrm{f}} &= \mathbf{x}_k^{\mathrm{f}} - \mathbf{x}_k^{\mathrm{t}}, \\ \mathbf{e}_k^{\mathrm{a}} &= \mathbf{x}_k^{\mathrm{a}} - \mathbf{x}_k^{\mathrm{t}}, \end{aligned}$$

  where $\mathbf{x}_k^{\mathrm{t}}$ is the (unknown) true state.

  $\Rightarrow$ Their respective error covariance matrices are

$$\begin{aligned} \mathbf{P}_k^{\mathrm{f}} &= \mathrm{Cov}(\mathbf{e}_k^{\mathrm{f}}) = \mathrm{E}\left[\mathbf{e}_k^{\mathrm{f}}(\mathbf{e}_k^{\mathrm{f}})^{\mathrm{T}}\right], \\ \mathbf{P}_k^{\mathrm{a}} &= \mathrm{Cov}(\mathbf{e}_k^{\mathrm{a}}) = \mathrm{E}\left[\mathbf{e}_k^{\mathrm{a}}(\mathbf{e}_k^{\mathrm{a}})^{\mathrm{T}}\right]. \quad (4) \end{aligned}$$

- To compute this *optimal gain* requires a careful derivation, that is beyond our scope here (see [Asch2016, 2022]).

Kalman Filters - optimal gain

- The *Kalman gain* matrix, $\mathbf{K}$, is chosen to minimize the *a posteriori* error covariance equation (4).

- The resulting $\mathbf{K}$ that minimizes equation (4) is given by

$$\mathbf{K}_k = \mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^{\mathrm{T}} \left(\mathbf{H}_k \mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k \right)^{-1}, \quad (5)$$

  where we remark that $\mathbf{H}\mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k = \mathrm{E}\left[\mathbf{d}_k \mathbf{d}_k^{\mathrm{T}} \right]$ is the covariance of the innovation.

- Looking at this expression for $\mathbf{K}_k$, we see:

  $\Rightarrow$ when the measurement error covariance $\mathbf{R}_k$ approaches zero, the gain $\mathbf{K}_k$ weights the innovation more heavily, since

$$\lim_{\mathbf{R} \to 0} \mathbf{K}_k = \mathbf{H}_k^{-1}.$$

$\Rightarrow$ On the other hand, as the *a priori* error estimate covariance $\mathbf{P}_k^{\mathrm{f}}$ approaches zero, the gain $\mathbf{K}_k$ weights the innovation less heavily, and

$$\lim_{\mathbf{P}_k^{\mathrm{f}} \to 0} \mathbf{K}_k = 0.$$

$\Rightarrow$ Another way of thinking about the weighting of $\mathbf{K}$ is that as the measurement error covariance $\mathbf{R}$ approaches zero, the actual measurement $\mathbf{y}_k$ is "trusted" more and more, while the predicted measurement $\mathbf{H}_k \mathbf{x}_k^{\mathrm{f}}$ is trusted less and less.

$\Rightarrow$ On the other hand, as the *a priori* error estimate covariance $\mathbf{P}_k^{\mathrm{f}}$ approaches zero, the actual measurement $\mathbf{y}_k$ is trusted less and less, while the predicted measurement $\mathbf{H}_k \mathbf{x}_k^{\mathrm{f}}$ is trusted more and more—this will be illustrated in the computational example below.

Kalman Filters - 2-step procedure

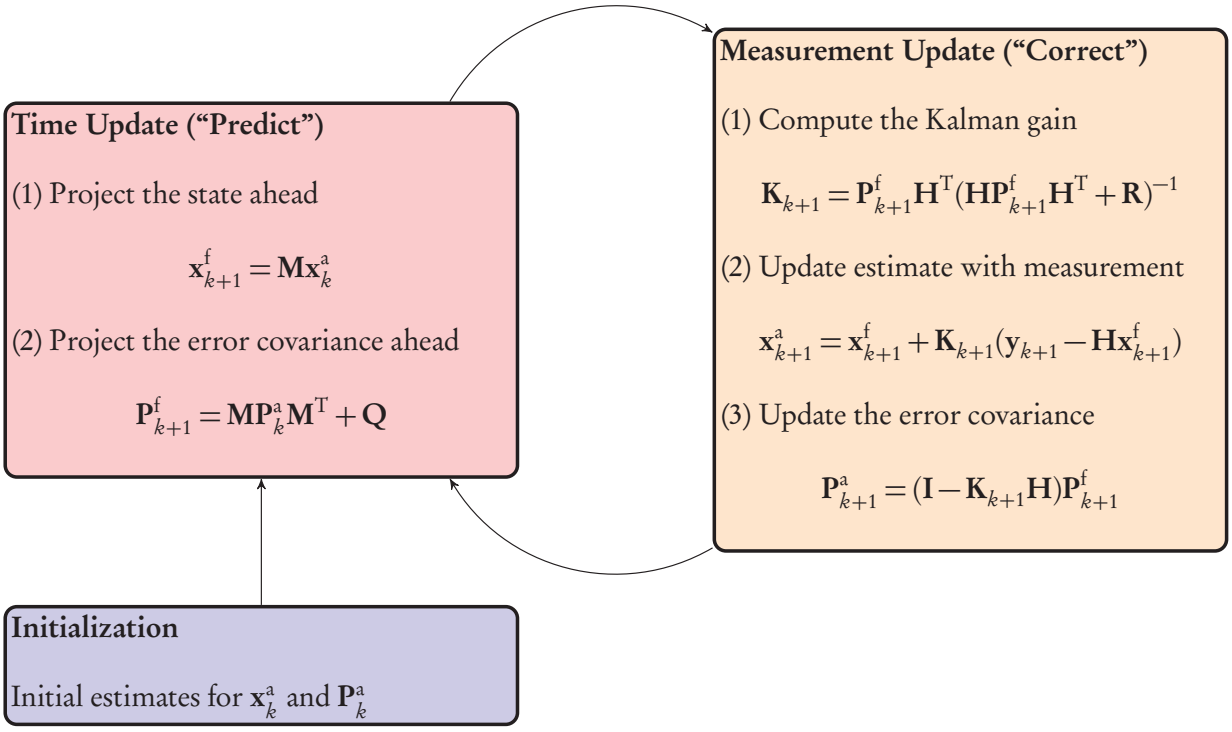
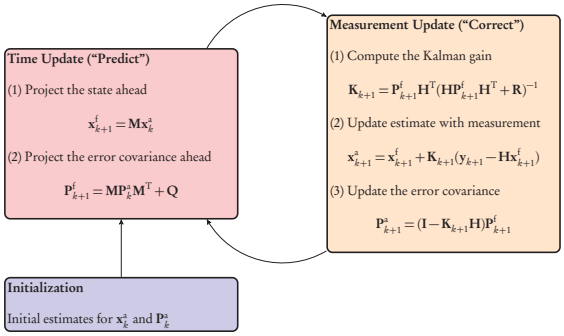


Figure 3: Kalman filter loop, showing the two phases, predict and correct, preceded by an initialization step.

The predictor-corrector loop is illustrated in the Figure and can be transposed, as is, into an operational algorithm.

KF - predictor/forecast step



- Start from a previous analyzed state, $\mathbf{x}^{\mathrm{a}}_k$, or from the initial state if $k = 0$, characterized by the Gaussian pdf $p(\mathbf{x}^{\mathrm{a}}_k \mid \mathbf{y}^{\mathrm{o}}_{1:k})$ of mean $\mathbf{x}^{\mathrm{a}}_k$ and covariance matrix $\mathbf{P}^a_k$.[2]

- An estimate of $\mathbf{x}^{\mathrm{t}}_{k+1}$ is given by the dynamical model which defines the forecast as

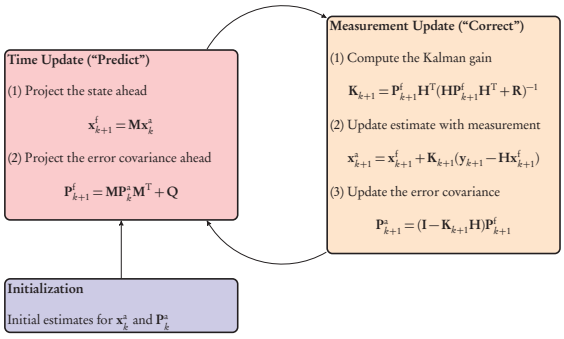
$$\mathbf{x}^{\mathrm{f}}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}^{\mathrm{a}}_k, \quad (6)$$

$$\mathbf{P}^{\mathrm{f}}_{k+1} = \mathbf{M}_{k+1}\mathbf{P}^{\mathrm{a}}_k\mathbf{M}^{\mathrm{T}}_{k+1} + \mathbf{Q}_{k+1}, \quad (7)$$

[2] We use here the classical notation $\mathbf{y}_{i:j} = (\mathbf{y}_i, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_j)$ for $i \leq j$ that denotes conditioning on all the observations in the interval.

where the expression for $\mathbf{P}^{\mathrm{f}}_{k+1}$ is obtained from the dynamics equation and the definition of the model noise covariance, $\mathbf{Q}$.

KF - corrector/analysis step



- At time t_{k+1}, the pdf $p(\mathbf{x}^{\mathrm{f}}_{k+1} \mid \mathbf{y}^{\mathrm{o}}_{1:k})$ is known, thanks to the mean $\mathbf{x}^{\mathrm{f}}_{k+1}$ and covariance matrix $\mathbf{P}^{\mathrm{f}}_{k+1}$ just calculated, as well as the assumption of a Gaussian distribution.

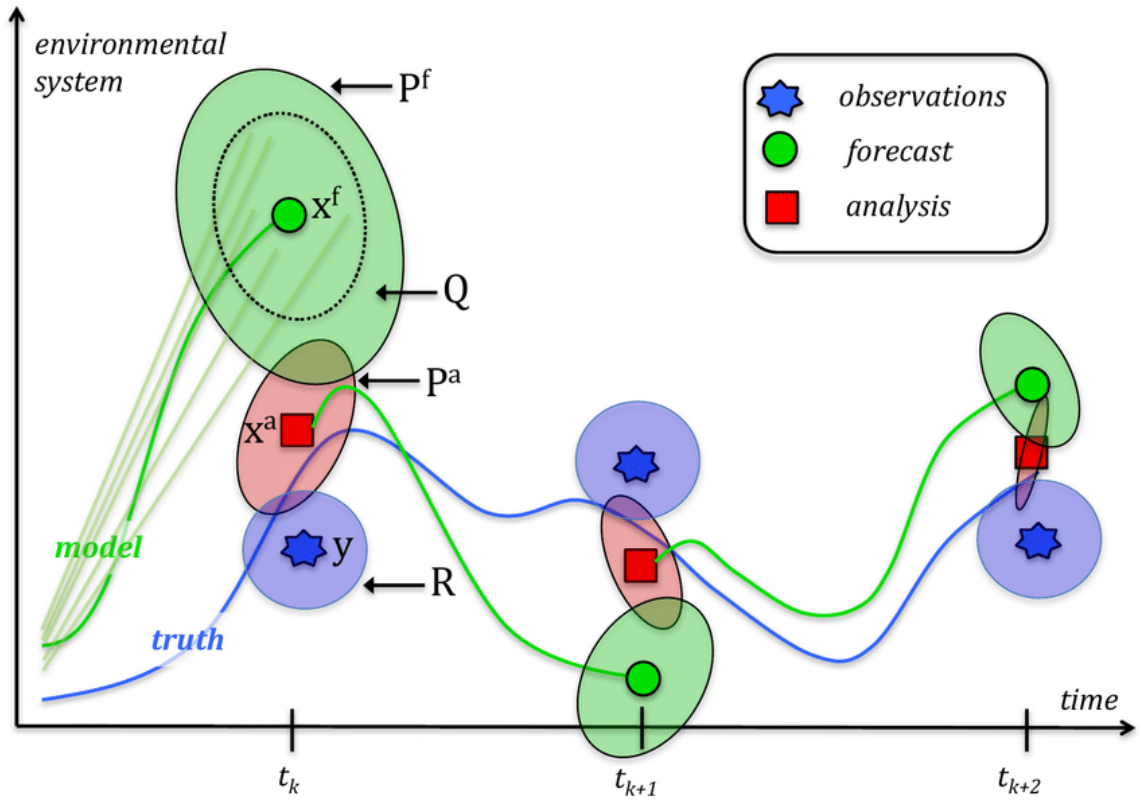
- The analysis step then consists of correcting this pdf using the observation available at time t_{k+1} in order to compute $p(\mathbf{x}^{\mathrm{a}}_{k+1} \mid \mathbf{y}^{\mathrm{o}}_{1:k+1})$. This comes from the BLUE in the dynamical context and gives

$$\mathbf{K}_{k+1} = \mathbf{P}^{\mathrm{f}}_{k+1}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}_{k+1}\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{k+1}\right)^{-1}, \quad (8)$$

$$\mathbf{x}^{\mathrm{a}}_{k+1} = \mathbf{x}^{\mathrm{f}}_{k+1} + \mathbf{K}_{k+1}\left(\mathbf{y}_{k+1} - \mathbf{H}\mathbf{x}^{\mathrm{f}}_{k+1}\right), \quad (9)$$

$$\mathbf{P}^{\mathrm{a}}_{k+1} = \left(\mathbf{I} - \mathbf{K}_{k+1}\mathbf{H}\right)\mathbf{P}^{\mathrm{f}}_{k+1}. \quad (10)$$

KF - Overall Picture



Principle: as we move forward in time, the uncertainty of the analysis is reduced, and the forecast is improved.

KF - Relation Between Bayes and BLUE

- If we know that the *a priori* and the observation data are both Gaussian, Bayes' rule can be readily applied to compute the *a posteriori* pdf.

  $\Rightarrow$ The *a posteriori* pdf is then Gaussian, and its parameters are given by the BLUE equations.

- Hence with Gaussian pdfs and a linear observation operator, there is no need to use Bayes' rule.

  $\Rightarrow$ The BLUE equations can be used instead to compute the parameters of the resulting pdf.
  $\Rightarrow$ Since the BLUE provides the same result as Bayes' rule, it is the best estimator of all.

- In addition one can recognize the 3D-Var cost function.

$\Rightarrow$ By optimizing this cost function, 3D-Var finds the MAP (maximum a posteriori) estimate of the Gaussian pdf, which is equivalent to the MV (minimum variance) estimate found by the BLUE.

ENSEMBLE KALMAN FILTERS

Ensemble Kalman Filter - EnKF

- The ensemble Kalman filter (EnKF) is an elegant approach that avoids

  $\Rightarrow$ the steps of linearization in the classical Kalman Filter,

  $\Rightarrow$ and the need for adjoints in the variational approach.

- It is still based on a Kalman filter, but an ensemble of realizations is used to compute an estimate of the population mean and variance, thus avoiding the need to compute inverses of potentially large matrices to obtain the posterior covariance, as was the case above in equations (8) and (10).

- The EnKF and its variants have been successfully developed and implemented in meteorology and oceanography, including in operational weather forecasting systems. Because the method is simple

to implement, it has been widely used in these fields.

- But it has spread out to other geoscience disciplines and beyond. For instance, to name a few domains, it has been applied in greenhouse gas inverse modeling, air quality forecasting, extra-terrestrial atmosphere forecasting , detection and attribution in climate sciences, geomagnetism re-analysis , and ice-sheet parameter estimation and forecasting. It has also been used in petroleum reservoir estimation, in adaptive optics for extra large telescopes, and highway traffic estimation.

- More recently, the idea was proposed to exploit the EnKF as a universal approach for all inverse problems. The term EKI, Ensemble Kalman Inversion, is used to describe this approach.

Principle of the EnKF

- The EnKF was originally proposed by G. Evensen in 1994 and amended in [Evenson2009].

Definition 1. The ensemble Kalman filter (EnKF) is a Kalman filter that uses an ensemble of realizations to compute estimates of the population mean and covariance.

- Since it is based on Gaussian statistics (mean and covariance) it does not solve the Bayesian filtering problem in the limit of a large number of particles, as opposed to the more general *particle filter*—seeAdvanced Course. Nonetheless, it turns out to be an excellent approximate algorithm for the filtering problem.

- As in the particle filter, the EnKF is based on the concept of particles, a collection of state vectors, which are called the members of the ensemble.

⇒ Rather than propagating huge covariance matrices, the errors are emulated by scattered particles, a collection of state vectors whose variability is meant to be representative of the uncertainty of the system's state resulting from the forecaster's ignorance.

⇒ Just like the particle filter, the members are propagated by the nonlinear model, without any linearization. Not only does this avoid the derivation of the tangent linear model, but it also circumvents the approximate linearization.

⇒ Finally, as opposed to the particle filter, the EnKF does not irremediably suffer from the curse of dimensionality.

- To sum up, here are the important remarks:

⇒ the EnKF avoids the linearization step of the KF;

⇒ the EnKF avoids the inversion of potentially large matrices;

⇒ the EnKF does not require any adjoint, as in variational assimilation;

$\Rightarrow$ the EnKF has been applied to a vast number of real-world problems.

EnKF - the Three Steps

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

- Please see the Algorithm bleow for details of each step.

- 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.

EnKF - Analysis Step

- The EnKF seeks to mimic the analysis step of the Kalman filter but with an ensemble of limited size in place of the unwieldy covariance matrices.

- The goal is to perform for each member of the ensemble an analysis of the form,

$$\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{f}} + \mathbf{K}\left[\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i^{\mathrm{f}})\right], \qquad (11)$$

  where

  $\Rightarrow$ $i = 1, \ldots, m$ is the member index in the ensemble,

  $\Rightarrow$ $\mathbf{x}_i^{\mathrm{f}}$ is the forecast state vector i, which represents a background state or prior at the analysis time.

- To mimic the Kalman filter, $\mathbf{K}$ must be identified with the Kalman gain

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}^{-1}, \qquad (12)$$

that we wish to estimate from the ensemble statistics.

$\Rightarrow$ First of all, we can estimate the forecast error covariance matrix as a sum over the ensemble,

$$\mathbf{P}^{\mathrm{f}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}} \right) \left(\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}} \right)^{\mathrm{T}},$$

with

$$\overline{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i^{\mathrm{f}}.$$

$\Rightarrow$ The forecast error covariance matrix can be factorized into

$$\mathbf{P}^{\mathrm{f}} = \mathbf{X}_{\mathrm{f}} \mathbf{X}_{\mathrm{f}}^{\mathrm{T}},$$

where $\mathbf{X}_{\mathrm{f}}$ is a $n \times m$ matrix whose columns are the *normalized anomalies* or *normalized perturbations* , i.e. for $i = 1, \ldots, m$

$$\left[\mathbf{X}_{\mathrm{f}} \right]_i = \frac{\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}}{\sqrt{m-1}}.$$

- We can now obtain from (11) a posterior ensemble $\{\mathbf{x}_i^{\mathrm{a}}\}_{i=1,\ldots,m}$ from which we can compute the posterior statistics.

- Hence, the posterior state and an ensemble of posterior perturbations can be estimated from

$$\overline{\mathbf{x}}^{\mathrm{a}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^{\mathrm{a}}\,, \quad [\mathbf{X}_{\mathrm{a}}]_i = \frac{\mathbf{x}_i^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}}}{\sqrt{m-1}}.$$

- Since $\mathbf{y}_i \equiv \mathbf{y}$ was assumed, the normalized anomalies, $\mathbf{X}_i^{\mathrm{a}} \equiv [\mathbf{X}_{\mathrm{a}}]_i$, i.e. the normalized deviations of the ensemble members from the mean are obtained from (11) minus the mean update,

$$\mathbf{X}_i^{\mathrm{a}} = \mathbf{X}_i^{\mathrm{f}} + \mathbf{K}\left(\mathbf{0} - \mathbf{H}\mathbf{X}_i^{\mathrm{f}}\right) = \left(\mathbf{I}_n - \mathbf{K}\mathbf{H}\right)\mathbf{X}_i^{\mathrm{f}}, \tag{13}$$

$\Rightarrow$ where $\mathbf{X}_i^{\mathrm{f}} \equiv [\mathbf{X}_{\mathrm{f}}]_i$, which yields the analysis

error covariance matrix,

$$\begin{aligned}\mathbf{P}^{\mathrm{a}} &= \mathbf{X}_{\mathrm{a}}\mathbf{X}_{\mathrm{a}}^{\mathrm{T}} \\ &= (\mathbf{I}_n - \mathbf{KH})\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}} \\ &= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}}.\end{aligned}$$

- Note that such a computation is never carried out in practice. However, theoretically, in order to mimic the best linear unbiased estimator (BLUE) analysis of the Kalman filter, we should have obtained

$$\begin{aligned}\mathbf{P}^{\mathrm{a}} &= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}} \\ &= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}.\end{aligned}$$

  $\Rightarrow$ Therefore, the error covariances are underestimated since the second positive term, related to the observation errors, is ignored, which is likely to lead to the divergence of the EnKF when the scheme is cycled.

- An elegant solution around this problem is to perturb the observation vector for each member:

$\mathbf{y}_i = \mathbf{y} + \mathbf{u}_i$, where $\mathbf{u}_i$ is drawn from the Gaussian distribution $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{R})$.

$\Rightarrow$ Let us define $\overline{\mathbf{u}}$ the mean of the sampled $\mathbf{u}_i$, and the innovation perturbations

$$[\mathbf{Y}_{\mathrm{f}}]_i = \frac{\mathbf{H}\mathbf{x}_i^{\mathrm{f}} - \mathbf{u}_i - \mathbf{H}\overline{\mathbf{x}}^{\mathrm{f}} + \overline{\mathbf{u}}}{\sqrt{m-1}}. \qquad (14)$$

$\Rightarrow$ The posterior anomalies are modified accordingly,

$$\mathbf{X}_i^{\mathrm{a}} = \mathbf{X}_i^{\mathrm{f}} - \mathbf{K}\mathbf{Y}_i^{\mathrm{f}} = (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{X}_i^{\mathrm{f}} + \frac{\mathbf{K}(\mathbf{u}_i - \overline{\mathbf{u}})}{\sqrt{m-1}}. \qquad (15)$$

- These anomalies yield the analysis error covariance

matrix,

$$\begin{aligned}\mathbf{P}^{\mathrm{a}} =&(\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}} \\ &+ \mathbf{K}\left[\frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{u}_i - \overline{\mathbf{u}})(\mathbf{u}_i - \overline{\mathbf{u}})^{\mathrm{T}}\right]\mathbf{K}^{\mathrm{T}} \\ &+ \frac{1}{\sqrt{m-1}}(\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{u}_i - \overline{\mathbf{u}})^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} \\ &+ \frac{1}{\sqrt{m-1}}\mathbf{K}(\mathbf{u}_i - \overline{\mathbf{u}})\mathbf{P}^{\mathrm{f}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}},\end{aligned}$$

whose expectation over the random noise gives the proper expected posterior covariances,

$$\begin{aligned}\mathrm{E}\left[\mathbf{P}^{\mathrm{a}}\right] &= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}} \\ &\quad + \mathbf{K}\mathrm{E}\left[\frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{u}_i - \overline{\mathbf{u}})(\mathbf{u}_i - \overline{\mathbf{u}})^{\mathrm{T}}\right]\mathbf{K}^{\mathrm{T}} \\ &= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}(\mathbf{I}_n - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK} \\ &= (\mathbf{I}_n - \mathbf{KH})\mathbf{P}^{\mathrm{f}}.\end{aligned}$$

- Note that the gain can be formulated in terms of the anomaly matrices only,

$$\mathbf{K} = \mathbf{X}_\mathrm{f} \mathbf{Y}_\mathrm{f}^\mathrm{T} \left(\mathbf{Y}_\mathrm{f} \mathbf{Y}_\mathrm{f}^\mathrm{T} \right)^{-1}, \qquad (16)$$

since

$\Rightarrow$ $\mathbf{X}_\mathrm{f} \mathbf{Y}_\mathrm{f}^\mathrm{T}$ is a sample estimate for $\mathbf{P}^\mathrm{f} \mathbf{H}^\mathrm{T}$ and
$\Rightarrow$ $\mathbf{Y}_\mathrm{f} \mathbf{Y}_\mathrm{f}^\mathrm{T}$ is a sample estimate for $\mathbf{H} \mathbf{P}^\mathrm{f} \mathbf{H}^\mathrm{T} + \mathbf{R}$.

- In this form, it is striking that the updated perturbations are linear combinations of the forecast perturbations. The new perturbations are sought within the ensemble subspace of the initial perturbations.

- Similarly, the state analysis is sought within the affine space $\overline{\mathbf{x}}^\mathrm{f} + \mathrm{Vec}\left(\mathbf{X}_1^\mathrm{f}, \mathbf{X}_2^\mathrm{f}, \ldots, \mathbf{X}_m^\mathrm{f} \right)$.

EnKF - Forecast Step

- In the forecast step, the updated ensemble obtained at the analysis step is propagated by the model over a time step,

$$\text{for} \quad i = 1, \ldots, m \quad \mathbf{x}^{\mathrm{f}}_{i,k+1} = \mathcal{M}_{k+1}(\mathbf{x}^{\mathrm{a}}_{i,k}).$$

- A forecast can be computed from the mean of the forecast ensemble, while the forecast error covariances can be estimated from the forecast perturbations.

- Notyes:

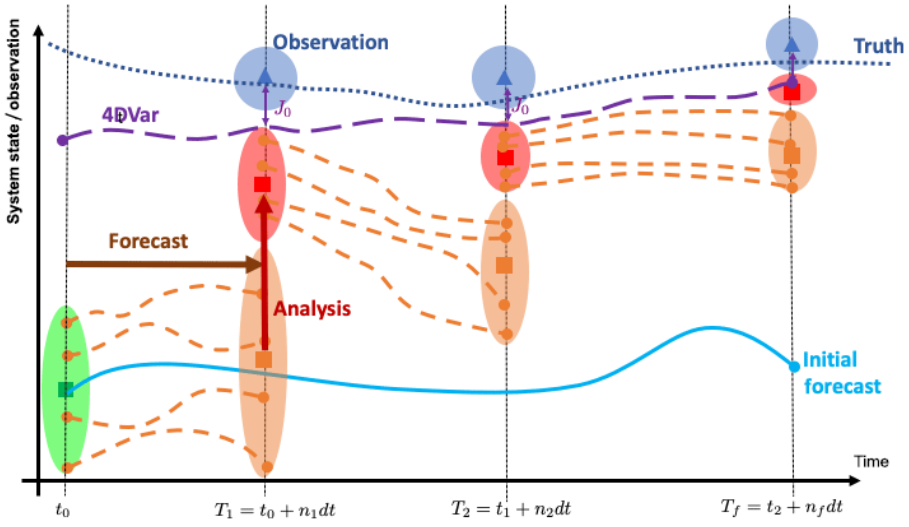
  $\Rightarrow$ These are only optional diagnostics in the scheme and they are not required in the cycling of the EnKF.

  $\Rightarrow$ It is important to observe that using the tangent linear model (TLM) operator, or any linearization thereof, was avoided.

$\Rightarrow$ This difference should particularly matter in a significantly nonlinear regime.

$\Rightarrow$ However, as we shall see in the Advanced Course Lectures, in strongly nonlinear regimes, the EnKF is largely dominated by schemes known as the iterative EnKF and the iterative ensemble Kalman smoother

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 m 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 analyszd 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 to T_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).

EnKF - the Algorithm

Given: For $k = 0, \ldots, K$, observation error cov. matrices

$\mathbf{R}_k$, observation models $\mathcal{H}_k$, forward models $\mathcal{M}_k$.

Compute: the ensemble forecast $\left\{\mathbf{x}_{i,k}^{\mathrm{f}}\right\}_{i=1,\ldots,m,\, k=1,\ldots,K}$

$\left(\mathbf{x}_{i,0}^{\mathrm{f}}\right)_{i=1,\ldots,m}$ #Initialize the ensemble

for $k = 0$ to K do #Loop over time

  for $i = 1$ to m do #Draw a stat. consistent obs. set

    $\mathbf{u}_i \sim \mathcal{N}(0, \mathbf{R}_k)$

    $\mathbf{y}_{i,k} = \mathbf{y}_k + \mathbf{u}_i$

  end for

#Compute the ensemble means

  $\overline{\mathbf{x}}_k^{\mathrm{f}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_{i,k}^{\mathrm{f}}$, $\overline{\mathbf{u}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{u}_i$

  $\left[\mathbf{X}_{\mathrm{f}}\right]_{i,k} = \frac{\mathbf{x}_{i,k}^{\mathrm{f}} - \overline{\mathbf{x}}_k^{\mathrm{f}}}{\sqrt{m-1}}$, #Compute the normalized anomalies

  $\left[\mathbf{Y}_{\mathrm{f}}\right]_{i,k} = \frac{\mathbf{H}_k \mathbf{x}_{i,k}^{\mathrm{f}} - \mathbf{u}_i - \mathbf{H}_k \overline{\mathbf{x}}_k^{\mathrm{f}} + \overline{\mathbf{u}}}{\sqrt{m-1}}$

  $K_k = \mathbf{X}_k^{\mathrm{f}} \left(\mathbf{Y}_k^{\mathrm{f}}\right)^{\mathrm{T}} \left(\mathbf{Y}_k^{\mathrm{f}} \left(\mathbf{Y}_k^{\mathrm{f}}\right)^{\mathrm{T}}\right)^{-1}$ #Compute the gain

  for $i = 1$ to m do #Update the ensemble

    $\mathbf{x}_{i,k}^{\mathrm{a}} = \mathbf{x}_{i,k}^{\mathrm{f}} + K_k \left(\mathbf{y}_{i,k} - \mathcal{H}_k \left(\mathbf{x}_{i,k}^{\mathrm{f}}\right)\right)$

    $\mathbf{x}_{i,k+1}^{\mathrm{f}} = \mathcal{M}_{k+1} \left(\mathbf{x}_{i,k}^{\mathrm{a}}\right)$ #Compute the ensemble forecast

  end for

end for

Localization and Inflation

- We have traded the extended Kalman filter for a seemingly considerably cheaper filter meant to achieve similar performances.

- But this comes with significant drawbacks.

  $\Rightarrow$ Fundamentally, one cannot hope to represent the full error covariance matrix of a complex high-dimensional system with only a few modes $m \ll n$, usually from a few dozens to a few hundreds.

  $\Rightarrow$ This implies large *sampling errors*, meaning that the error covariance matrix is only sampled by a limited number of modes.

  $\Rightarrow$ This rank-deficiency is accompanied by spurious correlations at long distances that strongly affect the filter performance.

  $\Rightarrow$ Even though the unstable degrees of freedom of dynamical systems that we wish to control with

a filter are usually far fewer than the dimension of the system, they often still represent a substantial fraction of the total degrees of freedom. Forecasting an ensemble of such size is usually not affordable.

- The consequence of this issue always is the divergence of the filter.

- Hence, the EnKF is useful on the condition that efficient fixes are applied.

  $\Rightarrow$ To make it a viable algorithm, one first needs to cope with the rank-deficiency of the filter and with its manifestations, i.e. sampling errors.

  $\Rightarrow$ Fortunately, there are clever tricks to overcome this major issue, known as *localization* and *inflation*, which explains, ultimately, the broad success of the EnKF in geosciences and engineering.

Localization

- This idea is that for many systems with some geographic spread, distant observables are weakly correlated.

- In other words, two distant parts of the system are almost independent at least for short time scales.

- It is possible to exploit this relative independence and spatially localize the analysis. This has been naturally termed *localization*.

- There are two types of localization:

  $\Rightarrow$ Domain localization, where instead of performing a global analysis valid at any location in the domain, we perform a local analysis to update the local state variables using local observations.

  $\Rightarrow$ Covariance localization focuses on the forecast error covariance matrix. It is based on the

remark that the forecast error covariance matrix $\mathbf{P}_{\mathrm{f}}$ is of low rank, at most $m-1$, and that this rank-deficiency could be cured by filtering these empirical covariances.

- For implementation details, please consult [Asch2016].

Inflation

- Even when the analysis is made local, the error covariance matrices are still evaluated with an ensemble of limited size.

  $\Rightarrow$ This often leads to sampling errors and spurious correlations.

  $\Rightarrow$ With a proper localization scheme, they might be significantly reduced.

  $\Rightarrow$ However small are the residual errors, they will accumulate and they will carry over to the next cycles of the sequential EnKF scheme. As a consequence, there is always a risk that the filter may ultimately diverge.

- One way around is to inflate the error covariance matrix by a factor λ^2 slightly greater than 1 before or after the analysis.

$\Rightarrow$ For instance, after the analysis,

$$\mathbf{P}^{\mathrm{a}} \longrightarrow \lambda^2 \mathbf{P}^{\mathrm{a}}.$$

$\Rightarrow$ Another way to achieve this is to inflate the ensemble,

$$\mathbf{x}_i^{\mathrm{a}} \longrightarrow \overline{\mathbf{x}}^{\mathrm{a}} + \lambda \mathbf{x}_i^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}},$$

which can alternatively be enforced on the prior (forecast) ensemble. This type of inflation is called *multiplicative inflation.*

- For implementation details, please consult [Asch2016].

EnKF -Variants

- Many variants of the EnKF algorithm have been proposed to overcome some of its weaknesses. We will just mention some of them, and refer the reader to [Asch2016, Evensen2009] for full details and further references.

  $\Rightarrow$ The *ensemble square root*, or deterministic ensemble Kalman filter, which does not perturb the observations.

  $\Rightarrow$ The *local ensemble* Kalman filter that remedies so-called divergence of the EnKF due to the rank deficiency of its approximated covariance matrix.

  $\Rightarrow$ The *maximum likelihood* ensemble filter that generalizes the BLUE update to nonlinear observation operators.

  $\Rightarrow$ *Hierarchical* EnKF based on the use of a Bayesian statistical hierarchy.

EXAMPLES

Comparison of KF, 4D-Var, and 3D-Var

- As in the previous Lecture, we consider the same scalar 4D-Var example, but this time apply the Kalman filter to it.

- We take the most simple linear forecast model,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\alpha x,$$

  with α a known positive constant.

- We assume the same discrete dynamics considered in with a single observation at time step 3.

- The stochastic system (1)-(2) is

$$x_{k+1}^{\mathrm{t}} = M(x_k^{\mathrm{t}}) + w_k,$$
$$y_{k+1} = x_k^{\mathrm{t}} + v_k,$$

where $w_k \sim \mathcal{N}(0, \sigma_Q^2)$, $v_k \sim \mathcal{N}(0, \sigma_R^2)$ and $x_0^{\mathrm{t}} - x_0^{\mathrm{b}} \sim \mathcal{N}(0, \sigma_B^2)$.

- The Kalman filter steps are

Forecast:

$$x_{k+1}^{\mathrm{f}} = M(x_k^{\mathrm{a}}) = \gamma x_k,$$
$$P_{k+1}^{\mathrm{f}} = \gamma^2 P_k^{\mathrm{a}} + \sigma_Q^2.$$

Analysis:

$$K_{k+1} = P_{k+1}^{\mathrm{f}} H \left(H^2 P_{k+1}^{\mathrm{f}} + \sigma_R^2\right)^{-1},$$
$$x_{k+1}^{\mathrm{a}} = x_{k+1}^{\mathrm{f}} + K_{k+1}(x_{k+1}^{\mathrm{o}} - H x_{k+1}^{\mathrm{f}}),$$
$$P_{k+1}^{\mathrm{a}} = (1 - K_{k+1}H) P_{k+1}^{\mathrm{f}} = \left(\frac{1}{P_{k+1}^{\mathrm{f}}} + \frac{1}{\sigma_R^2}\right)^{-1}, \quad H = 1.$$

Initialization:

$$x_0^{\mathrm{a}} = x_0^{\mathrm{b}},$$
$$P_0^{\mathrm{a}} = \sigma_B^2.$$

- We start with the initial state, at time step $k = 0$. The initial conditions are as above. The forecast is

$$x_1^{\mathrm{f}} = M(x_0^{\mathrm{a}}) = \gamma x_0^{\mathrm{b}},$$
$$P_1^{\mathrm{f}} = \gamma^2 \sigma_B^2 + \sigma_Q^2.$$

- Since there is no observation available, $H = 0$, and the analysis gives,

$$K_1 = 0,$$
$$x_1^{\mathrm{a}} = x_1^{\mathrm{f}} = \gamma x_0^{\mathrm{b}},$$
$$P_1^{\mathrm{a}} = P_1^{\mathrm{f}} = \gamma^2 \sigma_B^2 + \sigma_Q^2.$$

- At the next time step, $k = 1$, and the forecast

gives

$$x_2^{\mathrm{f}} = M(x_1^{\mathrm{a}}) = \gamma^2 x_0^{\mathrm{b}},$$
$$P_2^{\mathrm{f}} = \gamma^2 P_1^{\mathrm{a}} + \sigma_Q^2 = \gamma^4 \sigma_B^2 + (\gamma^2 + 1)\sigma_Q^2.$$

- Once again there is no observation available, $H = 0$, and the analysis yields

$$K_2 = 0,$$
$$x_2^{\mathrm{a}} = x_2^{\mathrm{f}} = \gamma^2 x_0^{\mathrm{b}},$$
$$P_2^{\mathrm{a}} = P_2^{\mathrm{f}} = \gamma^4 \sigma_B^2 + (\gamma^2 + 1)\sigma_Q^2.$$

- Moving on to $k = 2$, we have the new forecast,

$$x_3^{\mathrm{f}} = M(x_2^{\mathrm{a}}) = \gamma^3 x_0^{\mathrm{b}},$$
$$P_3^{\mathrm{f}} = \gamma^2 P_2^{\mathrm{a}} + \sigma_Q^2 = \gamma^6 \sigma_B^2 + (\gamma^4 + \gamma^2 + 1)\sigma_Q^2.$$

- Now there is an observation, x_3^{o}, available, so

$H = 1$ and the analysis is

$$
\begin{aligned}
K_3 &= P_3^{\mathrm{f}} \left(P_3^{\mathrm{f}} + \sigma_R^2 \right)^{-1}, \\
x_3^{\mathrm{a}} &= x_3^{\mathrm{f}} + K_3 (x_3^{\mathrm{o}} - x_3^{\mathrm{f}}), \\
P_3^{\mathrm{a}} &= (1 - K_3) P_3^{\mathrm{f}}.
\end{aligned}
$$

- Substituting and simplifying, we find

$$
x_3^{\mathrm{a}} = \gamma^3 x_0^{\mathrm{b}} + \frac{\gamma^6 \sigma_B^2 + (\gamma^4 + \gamma^2 + 1)\sigma_Q^2}{\sigma_R^2 + \gamma^6 \sigma_B^2 + (\gamma^4 + \gamma^2 + 1)\sigma_Q^2} \left(x_3^{\mathrm{o}} - \gamma^3 x_0^{\mathrm{b}} \right). \tag{17}
$$

Case 1: Assume we have a perfect model, then $\sigma_Q^2 = 0$ and the Kalman filter state (17) becomes

$$
x_3^{\mathrm{a}} = \gamma^3 x_0^{\mathrm{b}} + \frac{\gamma^6 \sigma_B^2}{\sigma_R^2 + \gamma^6 \sigma_B^2} \left(x_3^{\mathrm{o}} - \gamma^3 x_0^{\mathrm{b}} \right),
$$

which is precisely the 4D-Var expression obtained before.

Case 2: When the parameter α tends to zero, then γ tends to one, the model is stationary and the Kalman filter state (17) becomes

$$x_3^{\mathrm{a}} = x_0^{\mathrm{b}} + \frac{\sigma_B^2 + 3\sigma_Q^2}{\sigma_R^2 + \sigma_B^2 + 3\sigma_Q^2} \left(x_3^{\mathrm{o}} - x_0^{\mathrm{b}} \right),$$

which, when $\sigma_Q^2 = 0$, reduces to the 3D-Var solution,

$$x_3^{\mathrm{a}} = x_0^{\mathrm{b}} + \frac{\sigma_B^2}{\sigma_R^2 + \sigma_B^2} \left(x_3^{\mathrm{o}} - x_0^{\mathrm{b}} \right),$$

that was obtained before.

Case 3: When α tends to infinity, then γ goes to zero, and we are in the case where there is no longer any memory with

$$x_3^{\mathrm{a}} = \frac{\sigma_Q^2}{\sigma_R^2 + \sigma_Q^2} x_3^{\mathrm{o}}.$$

Then, if the model is perfect, $\sigma_Q^2 = 0$ and $x_3^{\mathrm{a}} = 0$. If the observation is perfect, $\sigma_R^2 = 0$ and $x_3^{\mathrm{a}} = x_3^{\mathrm{o}}$.

- This example shows the complete chain, from the Kalman filter solution, through the 4D-Var, and finally reaching the 3D-Var one.

- Hopefully this clarifies the relationship between the three and demonstrates why the Kalman filter provides the most general solution possible.

PRACTICAL GUIDELINES

General Guidelines

We briefly point out some important practical considerations. It should now be clear that there are four basic ingredients in any inverse or data assimilation problem:

1. Observation or measured data.

2. A forward or direct model of the real-world context.

3. A backwards or adjoint model, in the variational case. A probabilistic framework, in the statistical case.

4. An optimization cycle.

But where does one start?

- The traditional approach, often employed in mathematical and numerical modeling, is to begin with some simplified, or at least well-known, situation.

- Once the above four items have been successfully implemented and tested on this instance, we then proceed to take into account more and more reality in the form of real data, more realistic models, more robust optimization procedures, etc.

- In other words, we introduce uncertainty, but into a system where we at least control some of the aspects.

Twin Experiments

Twin experiments, or synthetic runs, are a basic and indispensable tool for all inverse problems. In order to evaluate the performance of a data assimilation system we invariably begin with the following methodology.

1. Fix all parameters and unknowns and define a reference trajectory, obtained from a run of the direct model—call this the "truth".

2. Derive a set of (synthetic) measurements, or background data, from this "true" run.

3. Optionally, perturb these observations in order to generate a more realistic observed state.

4. Run the data assimilation or inverse problem algorithm, starting from an initial guess (different

from the "true" initial state used above), using the synthetic observations.

5. Evaluate the performance, modify the model/algorithm/observa and cycle back to step 1.

- Twin experiments thus provide a well-structured methodological framework.

- Within this framework we can perform different "stress tests" of our system.

  $\Rightarrow$ We can modify the observation network,
  $\Rightarrow$ increase or decrease (even switch off) the uncertainty,
  $\Rightarrow$ test the robustness of the optimization method,
  $\Rightarrow$ even modify the model.

- In fact, these experiments can be performed on the full physical model, or on some simpler (or reduced-order) model.

Toy Models

Toy models are, by definition, simplified models that we can play with. Yes, but these are of course "serious games." In certain complex physical contexts, of which meteorology is a famous example, we have well-established toy models, often of increasing complexity. These can be substituted for the real model, whose computational complexity is often too large, and provide a cheaper test-bed.

Some well-known examples of toy models are:

- Lorenz models that are used as an avatar for weather simulations.
- Various harmonic oscillators that are used to simulate dynamic systems.
- Other well-known models are the Ising model in physics, the Lotka-Volterra model in life sciences, and the Schelling model in social sciences.

Machine Learning

Machine Learning (ML) is becoming more and more present in our daily lives, and in scientific research. The use of ML in DA and Inverse modeling will be dealt with in the Advanced Course, where we will consider:

- ML-based Surrogate Models.
- Scientific ML.
- Bias and Ethics of ML.

Kalman Filter - extensions

- There are many variants, extensions and generalizations of the Kalman Filter.

- In the Advanced Course, we will study in more detail:

  $\Rightarrow$ ensemble Kalman Filters
  $\Rightarrow$ Bayesian and nonlinear Kalman Filters: extended, unscented
  $\Rightarrow$ particle filters

Choosing a Filter

One usually has to choose between

- linear Kalman filters
- ensemble Kalman filters
- nonlinear filters
- hybrid variational-filter methods.

These questions will be addressed in the Advanced Course.

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. S. Sarkka. *Bayesian Filtering and Smoothing.* Cambridge University Press, 2013.

3. S. Sarkka, A. Solin. Applied Stochastic Differential Equations. Cambridge University Press, 2019.

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

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

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

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

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

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

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