

Data assimilation in the geosciences

An overview

Alberto Carrassi^{1,2}, Olivier Talagrand³ and Marc Bocquet⁴

(1): Nansen Environmental and Remote Sensing Center - Norway

(2): Geophysical Institute, University of Bergen - Norway

(3): LMD, École Normale Supérieure, IPSL, France

(4): CEREVA, joint lab École des Ponts ParisTech and EdF R&D, IPSL, France



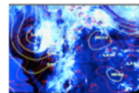
European Geosciences Union
General Assembly 2018

Data Assimilation

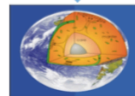


An on-going rapid expansion from **Weather Science (NWP)** into **Climate Science/Geophysics** in general:

- Oceanography
- Climate Prediction
- Climate Assessment
- Hydrology
- Geology
- Climatology
- Detection & Attribution
- ... and many more beyond geosciences ...



NWP



Geophysics

Outline of part I

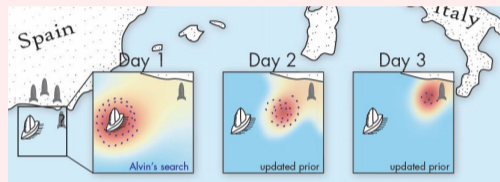
- 1 Posing the problem: sequential inference
- 2 Our ingredients
 - The model: *deterministic* or *stochastic*?
 - The observations
- 3 Bayesian formulation of the inference
- 4 Sequential Bayesian estimate
 - The main result: recursive update
 - *Prediction, filtering* and *smoothing*
- 5 A route to solution: the Gaussian approximation
 - The Kalman filter and smoother
- 6 Essential bibliography Part I

Posing the problem: sequential inference

- ▶ *Inference* is the process of taking a decision based on limited information.
- ▶ Limitations arise by incomplete and noisy data and by an approximate knowledge about the laws (if any) governing the system evolution.
- ▶ Specifically (yet informally): the problem we intend to solve is the estimation of the state of a system, at any arbitrary past, present and/or future times.
- ▶ We will assume that the data are time-ordered and that the system is given under the form of a time-evolving dynamics.
- ▶ *Sequential inference* is the problem of updating our knowledge about the system each time new data becomes available.

Posing the problem

An example: *Palomares (Spain) incident*



From *Berkeley science review* at
<http://berkeleysciencereview.com/article/toolbox/>

- On January 17th, 1966, a US Air Force bomber flying over the south of Spain, with four hydrogen bombs, exploded in midair.
- Three bombs were recovered, undetonated, on land, while the fourth was lost.
- According to a local fisherman, it splashed down somewhere in the Mediterranean Sea.
- How would you go about looking for the bomb in a way that maximizes the chance you'll find it?

► This lecture treats the case of the *discrete-model/discrete-observation* estimation problem (relevant in many practical cases such as climate science, biology or navigation among many others).

► Complete treatments of the *continuous-continuous* and *discrete-continuous* cases can be found in many textbooks on estimation theory (e.g., *Jazwinski*, 1970; *Bain and Crisan*, 2009).

The model: *what we know about the system physical-dynamical laws*

We will assume that a model of the natural process of interest is available as a discrete stochastic-dynamical system,

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k. \quad (1)$$

- ▶ $\mathbf{x}_k \in \mathbb{R}^m$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$ are the model state and parameter vectors respectively.
- ▶ The model parameters may include the external forcings or the boundary conditions.
- ▶ $\mathcal{M}_{k:k-1} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is usually a nonlinear, possibly *chaotic*, function from time t_{k-1} to t_k .
- ▶ $\boldsymbol{\eta}_k \in \mathbb{R}^m$ is the *model error*, represented as a stochastic additive term.

Remark: $\boldsymbol{\eta}_k$ could be included into the parentheses without loss of generality.

Remark: The *stochastic difference model*, Eq. (1), has a continuous-time counterpart ($\Delta t \rightarrow 0$); it is known as the *Itô stochastic differential equation* (see, e.g. Jazwinski, 1970; Reich and Cotter, 2015)

The model: *why stochastic?*

- ▶ **Imperfect model** - $\mathcal{M}_{k:k-1}$ embeds our knowledge about the laws governing the process \implies Such a knowledge is always (in realistic cases) partial and/or incorrect.
- ▶ **Numerical discretization** - $\mathcal{M}_{k:k-1}$ is often a spatio-temporal discretization of physical laws (*e.g.*, the Navier Stokes equations for fluids) expressed as partial differential equations on a continuous media \implies The finite resolution (dictated by the computational power availability) induces errors.
- ▶ **Chaos** - Many natural systems are chaotic and exhibit extreme sensitivity to initial conditions \implies Any (inevitable?) error in the knowledge of the system state at an arbitrary time contaminates the prediction.

The sources of error are accounted for using a stochastic model

The observations and their relation with the quantities of interest

- ▶ Noisy observations, $\mathbf{y}_k \in \mathbb{R}^d$, are available at discrete times and are related to the model state vector through

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k. \quad (2)$$

with $\mathcal{H} : \mathbb{R}^m \rightarrow \mathbb{R}^d$ the, generally nonlinear, *observation operator* mapping from model to observational space.

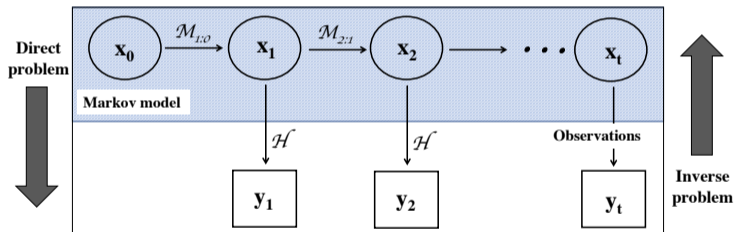
- ▶ Observational error, $\boldsymbol{\epsilon}_k$, is represented as a stochastic term, accounting for the instrumental error, deficiencies in the formulation of \mathcal{H} and the *representativity error*.
- ▶ The latter arises from the presence of unresolved scales and represents their effect on the resolved scales - it is ubiquitous in physical science and inherent to the discretization procedure.

Remark: we assume the observation dimension is constant, so that $d_k = d$ (the generalization is trivial).

Remark: often $d \ll m$, *i.e.*, the amount of available data is insufficient to fully describe the system.

The observations and their relation with the quantities of interest

- The stochastic model dynamics, Eq. (1), together with the stochastic observation model, Eq. (2) define an **Hidden Markov model (HMM)**



- A stochastic model is said *Markov* if its future state depends only on the current state and not on any states the models has attained before.
- We focus here on the **inverse problem** \Leftrightarrow Estimate \mathbf{x} by observing \mathbf{y} .

Bayesian inference for the inverse problem

- ▶ With the two complementary pieces of information in hand, *model* and *data*, we can move forward and formalize their fusion.
- ▶ When making inference we have to decide how much we trust the uncertain information. \Rightarrow We need to **quantify the uncertainty**.
- ▶ Given the stochastic nature of the problem

uncertainty quantification is done using probabilities.

- ▶ The Bayesian approach offers a natural mathematical framework to understand and formalize this problem.
- ▶ In particular, the goal of Bayesian inference is to estimate the uncertainty in \mathbf{x} given $\mathbf{y} \Leftrightarrow$ Compute the conditional **probability density function** (PDF) $p(\mathbf{x}|\mathbf{y})$.

Bayes' theorem

Bayes' theorem

Let \mathbf{x} and \mathbf{y} be jointly distributed random vectors with joint PDF, $p(\mathbf{x}, \mathbf{y})$. Then

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (3)$$

- ▶ An algebraic equation for conditional probabilities \Leftrightarrow The probability that the event, \mathbf{x} , occurs, knowing that another one, \mathbf{y} , has occurred.
- ▶ The output of the estimation process is the **posterior distribution** $p(\mathbf{x}|\mathbf{y})$
- ▶ $p(\mathbf{x})$ is the **prior PDF** that gathers all the knowledge before assimilating the new observations. It is a distinctive feature of the Bayesian approach.

Bayes' theorem

► $p(\mathbf{y}|\mathbf{x})$ is the **likelihood of the data** conditioned on the state \mathbf{x} (*i.e.*, it quantifies the likelihood of observing \mathbf{y} given a particular value of \mathbf{x}).

► $p(\mathbf{y})$ is the marginal distribution of the data, $p(\mathbf{y}) = \int d\mathbf{x} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$. It integrates to one and is treated as a normalization coefficient.

Remark: The factor $p(\mathbf{y})$ is relevant, for instance, in model selection problems and is also referred to as **model evidence**.

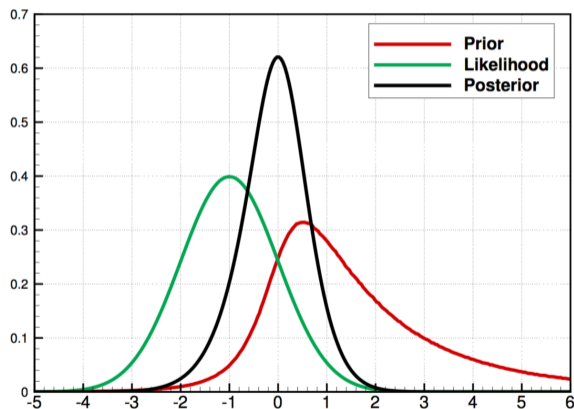


Figure: Courtesy of Geir Evensen

Sequential Bayesian estimate

- Recall our HMM given by the dynamical model, Eq. (1), data model, Eq. (2)

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k, \quad \mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k$$

- The model and the observational errors, $\{\boldsymbol{\eta}_k, \boldsymbol{\epsilon}_k : k = 1, \dots, K\}$ are assumed to be uncorrelated in time, mutually independent, and distributed according to the PDFs $p_{\boldsymbol{\eta}}$ and $p_{\boldsymbol{\epsilon}}$
- Let us define the sequences of system states and observations within the interval $[t_0, t_K]$ as $\mathbf{x}_{K:0} = \{\mathbf{x}_K, \mathbf{x}_{K-1}, \dots, \mathbf{x}_0\}$ and $\mathbf{y}_{K:1} = \{\mathbf{y}_K, \mathbf{y}_{K-1}, \dots, \mathbf{y}_1\}$ respectively.

We wish to estimate the posterior $p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1})$ for any arbitrary, sequentially increasing, t_K .

Using Bayes's law we have

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) \propto p(\mathbf{y}_{K:1}|\mathbf{x}_{K:0})p(\mathbf{x}_{K:0}) \quad (4)$$

Sequential Bayesian estimate

► Since the observational errors are assumed to be uncorrelated in time we have $p(\mathbf{y}_k | \mathbf{x}_{K:0}) = p(\mathbf{y}_k | \mathbf{x}_k)$ and we can split the products of the probabilities

$$p(\mathbf{y}_{K:1} | \mathbf{x}_{K:0}) = \prod_{k=1}^K p(\mathbf{y}_k | \mathbf{x}_k) = \prod_{k=1}^K p_{\epsilon}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)], \quad (5)$$

so that the mutual likelihood of all the observations in the interval $t_K - t_0$ is the product of the individual likelihoods at each time.

► Also, in virtue of the Markovian character we have $p(\mathbf{x}_{k+1} | \mathbf{x}_{k:0}) = p(\mathbf{x}_{k+1} | \mathbf{x}_k)$ (prediction at time t_{k+1} only depends on the state at t_k), and we can split the prior PDF as

$$p(\mathbf{x}_{K:0}) = p(\mathbf{x}_0) \prod_{k=1}^K p(\mathbf{x}_k | \mathbf{x}_{k-1}) = p(\mathbf{x}_0) \prod_{k=1}^K p_{\eta}[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})]. \quad (6)$$

Sequential Bayesian estimate - The main result

By combining Eq. (5) and (6) using Bayes' rule we get the posterior distribution

The recursive update

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:1}) \propto p(\mathbf{x}_0) \prod_{k=1}^K p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p(\mathbf{x}_0) \prod_{k=1}^K p_{\epsilon}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]p_{\eta}[\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})]. \quad (7)$$

► Equation (7) is of central importance: it states that **a new update can be obtained as soon as new data is available.**

► Sequential inference can be obtained by recursively estimating $p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})$.

Prediction, filtering and smoothing

Depending on which time period is needed for state estimation, it is possible to define *three estimation problems*:

- 1 Prediction:** estimate $p(\mathbf{x}_l | \mathbf{y}_{k:1})$ with $l > k$.
- 2 Filtering:** estimate $p(\mathbf{x}_k | \mathbf{y}_{k:1})$.
- 3 Smoothing:** estimate $p(\mathbf{x}_{K:0} | \mathbf{y}_{K:1})$, or selected marginals of this PDF, such as $p(\mathbf{x}_l | \mathbf{y}_{K:1})$, with $0 \leq l < K$.

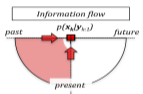
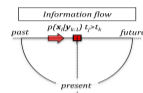
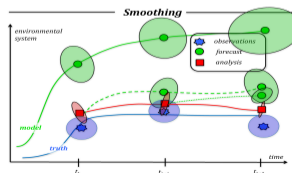
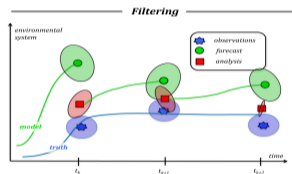
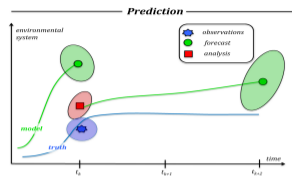


Figure from Carrasi et al., 2018

Formal Bayesian solutions - Prediction

► Given the conditional PDF $p(\mathbf{x}_k|\mathbf{y}_{k:0})$, we seek a law to propagate it forward in time under the effect of the model dynamics, Eq. (1).

This leads to the **Chapman-Kolmogorov equation** for the propagation of a PDF under the model dynamics, Eq. (1), as

$$p(\mathbf{x}_l|\mathbf{y}_{k:1}) = \int_{\mathbb{R}^m} d\mathbf{x}_k p(\mathbf{x}_l|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{k:1}), \quad (8)$$

where the initial PDF at t_k is given by $p(\mathbf{x}_k|\mathbf{y}_{k:1})$, and $p(\mathbf{x}_l|\mathbf{x}_k) = p_{\boldsymbol{\eta}}[\mathbf{x}_l - \mathcal{M}_{l:k}(\mathbf{x}_k)]$ in our HMM model.

► The **Prediction** problem is addressed by solving the Chapman-Kolmogorov equation (8), given the conditional PDF at time t_k , $p(\mathbf{x}_k|\mathbf{y}_{k:1})$.

Remark: In case of model dynamics given as a stochastic differential equation, instead of a stochastic difference equation as in Eq. (1), the Chapman-Kolmogorov equation becomes the *Fokker-Planck* equation.

Formal Bayesian solutions - Filtering

- **Filtering problem** is the most common in applications, and is characterized by sequential processing, in which measurements are utilized as they become available.
- An **analysis step**, in which the conditional PDF $p(\mathbf{x}_k|\mathbf{y}_{k:1})$ is updated using the latest observation, \mathbf{y}_k ,

Analysis

$$p(\mathbf{x}_k|\mathbf{y}_{k:1}) \propto p_{\epsilon}[\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]p(\mathbf{x}_k|\mathbf{y}_{k-1:1}), \quad (9)$$

- alternates with a **forecast step** which propagates this PDF, using Chapman-Kolmogorov equation, forward until the time of a new observation,

Forecast

$$p(\mathbf{x}_{k+1}|\mathbf{y}_{k:1}) = \int_{\mathbb{R}^m} d\mathbf{x}_k p_{\eta}[\mathbf{x}_{k+1} - \mathcal{M}_{k+1:k}(\mathbf{x}_k)]p(\mathbf{x}_k|\mathbf{y}_{k:1}), \quad (10)$$

to get $p(\mathbf{x}_{k+1}|\mathbf{y}_{k:1})$.

Formal Bayesian solutions - Smoothing

- ▶ Smoothing is relevant when, for instance, one is interested in a retrospective analysis after the observations have been collected.
- ▶ The goal is to estimate the conditional PDF, $p(\mathbf{x}_k | \mathbf{y}_{K:1})$ of the state at any time t_k , $0 \leq k \leq K$, based on all observations (past, present and future).
- ▶ First write the smoothing PDF at time t_k by marginalizing over \mathbf{x}_{k+1}

$$p(\mathbf{x}_k | \mathbf{y}_{K:1}) = \int_{\mathbb{R}^m} d\mathbf{x}_{k+1} p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{K:1}) p(\mathbf{x}_{k+1} | \mathbf{y}_{K:1}). \quad (11)$$

Note that, using Bayes' rule, the integrand in Eq. (11) can be written as

$$p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{K:1}) = p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{k:1}) \propto p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{k:1}), \quad (12)$$

given that the observations $\{\mathbf{y}_{k+1}, \dots, \mathbf{y}_K\}$ are independent of \mathbf{x}_k when \mathbf{x}_{k+1} is known.

- ▶ Finally remark that $p(\mathbf{x}_k | \mathbf{y}_{k:1})$ in Eq. (12) is the filter solution at time t_k .

Formal Bayesian solutions - Smoothing

- ▶ This implies that the smoothing PDFs can be obtained using a

forward-backward recursive algorithm.

- ▶ *Forward phase* - Start from $p(\mathbf{x}_0)$ then from $k = 1$ to $k = K$

- Estimate and store the filter PDFs, $p(\mathbf{x}_k|\mathbf{y}_{k:1})$.

- ▶ *Backward phase* - From $k = K - 1$ to $k = 1$

- Compute $p(\mathbf{x}_k|\mathbf{x}_{k+1}, \mathbf{y}_{k:1})$ with Eq. (12) using the stored filter PDFs, $p(\mathbf{x}_k|\mathbf{y}_{k:1})$ and $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$.

- Obtain the smoothing PDFs, $p(\mathbf{x}_k|\mathbf{y}_{K:1})$, from Eq. (11) making use of the smoothing PDF at time t_{k+1} , $p(\mathbf{x}_{k+1}|\mathbf{y}_{K:1})$ estimated at the previous iteration.

Comments on Bayesian filter and smoother

- ▶ The filter solution at an arbitrary t_k ($0 \leq k \leq K$) is obtained by sequential updating until $t_k \Rightarrow$ it thus accounts for all observations until t_k .
- ▶ In contrast, the smoothing solution at the same time t_k also accounts for future observations until t_K \Rightarrow it is thus generally more accurate than the filtering one.
- ▶ At the final time t_K both solutions have incorporated the same amount of data \Rightarrow in the absence of approximations, they will coincide.
- ▶ In general the filtering and smoothing do not possess a analytical solutions.
- ▶ However, when the dynamical and observational model are linear and all error PDFs are Gaussian analytic solutions exists: the famous **Kalman filter** and **Kalman smoother**.

The Gaussian and linear approximation

- ▶ The huge dimension of models and datasets hampers the use of a fully Bayesian approach in geosciences (see Part III for *Particle Filters*).
- ▶ It is usually assumed that observation and model error are Gaussian distributed \implies PDFs can be described completely in terms of their mean and covariance.
- ▶ The Gaussian approximation is at the core of most of the DA procedures successfully used in the geosciences.

Let assume that the dynamical and observational models are both linear

$$\mathbf{x}_k = \mathbf{M}_{k:k-1}\mathbf{x}_{k-1} + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \quad (13)$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad (14)$$

with $\mathbf{M}_{k:k-1}$ and \mathbf{H}_k being matrices in $\mathbb{R}^{m \times m}$ and $\mathbb{R}^{d \times m}$, respectively.

- ▶ The observational and model noises are assumed to be white-in-time, unbiased, and Gaussian distributed with covariances $\mathbf{R}_k \in \mathbb{R}^{d \times d}$ and $\mathbf{Q}_k \in \mathbb{R}^{m \times m}$ respectively.

The Kalman filter

The **KF recursive equations** reads

$$\text{Forecast Step} \quad \mathbf{x}_k^f = \mathbf{M}_{k:k-1} \mathbf{x}_{k-1}^a, \quad (15)$$

$$\mathbf{P}_k^f = \mathbf{M}_{k:k-1} \mathbf{P}_{k-1}^a \mathbf{M}_{k:k-1}^T + \mathbf{Q}_k. \quad (16)$$

$$\text{Analysis step} \quad \mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (17)$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f), \quad (18)$$

$$\mathbf{P}_k^a = (\mathbf{I}_k - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f. \quad (19)$$

► Given \mathbf{Q}_k , \mathbf{R}_k , \mathbf{H}_k and \mathbf{M}_k , for $k \geq 1$, and initial condition for the mean, $\mathbf{x}_0^a = \mathbf{x}_0$, and error covariance, $\mathbf{P}_0^a = \mathbf{P}_0$, Eqs. (15)–(19) estimate sequentially the state and the associated error covariance at any future time $k > 1$.

► The matrix $\mathbf{K}_k \in \mathbb{R}^{m \times d}$ is the Kalman gain and contains the coefficients of the linear combination between the forecast \mathbf{x}_k^f , and the observations.

► The resulting state estimate, the analysis \mathbf{x}_k^a , has minimum variance and is unbiased.

The Kalman smoother

- ▶ A recursive estimate of the PDFs $p(\mathbf{x}_k | \mathbf{y}_{K:1})$, can be obtained using a forward and backward recursions, in which a forward-in-time filter is followed by a backward-in-time smoother.
- ▶ Assume that a forward in time KF has been implemented and the forecast and analysis means and covariances, $\mathbf{x}_k^{f/a}$ and $\mathbf{P}_k^{f/a}$, have been computed and stored.
- ▶ We can then run the **KS recursion** backward in time, for $k = K - 1, \dots, 1$, to compute the smoothing mean and covariance, \mathbf{x}_k^{sm} and \mathbf{P}_k^{sm} , according to

$$\mathbf{S}_k = \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T (\mathbf{M}_{k+1:k} \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T + \mathbf{Q}_{k+1})^{-1} = \mathbf{P}_k^a \mathbf{M}_{k+1:k}^T \mathbf{P}_{k+1}^{-f}, \quad (20)$$

$$\text{KS Mean} \quad \mathbf{x}_k^{\text{sm}} = \mathbf{x}_k^a + \mathbf{S}_k (\mathbf{x}_{k+1}^{\text{sm}} - \mathbf{x}_{k+1}^f), \quad (21)$$

$$\text{KS Covariance} \quad \mathbf{P}_k^{\text{sm}} = \mathbf{P}_k^a + \mathbf{S}_k (\mathbf{P}_{k+1}^{\text{sm}} - \mathbf{P}_{k+1}^f) \mathbf{S}_k^T \quad (22)$$

Some properties of the Kalman filter and smoother

- ▶ **Time dependent prior** - The KF/KS recursions provide a time-dependent estimate of the prior that is highly desirable in chaotic systems, so that \mathbf{P}_k^f is itself strongly time-dependent.
- ▶ **Filter divergence** - It happens when the solution of the KF deviates dramatically from the true signal. When $\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T \ll \mathbf{R}_k$ the filter solution may start to ignore the observations (see Part III for countermeasures).
- ▶ **A diagnostic tool** - The innovation vector sequence, $\mathbf{v}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^f$, is Gaussian and white-in-time: one can thus keep monitoring the innovations to assess the KF optimality and, possibly, to implement corrections.
- ▶ **Bias and covariance estimation** - It is possible to estimate the error covariance and means along with the system states.
- ▶ **Dependency on the initial condition** - The criticality of the choice of the initial error covariance, \mathbf{P}_0 , is related to the filter's stability. If the filter is unstable one must properly initialize it to avoid divergence by correcting the error in the unstable directions.

The extended Kalman filter

- ▶ The **extended Kalman filter** (EKF) is a first-order expansion of the KF to nonlinear dynamics.
- ▶ The mean state is propagated by the full nonlinear model, but the error covariance evolution is approximated using the tangent linear one.
- ▶ The linearization is taken around the nonlinear model solution, so that the Jacobian of the model is evaluated upon it and it is state dependent.
- ▶ The EKF assumes Gaussian errors, but under the action of the nonlinear dynamics, even an initial Gaussian error covariance will not stay Gaussian, and the EKF will only provide an approximate description of the actual estimation error distribution.
- ▶ The EKF has been successful in a number of pioneering applications of DA for meteorology (*Ghil et al.*, 1981; *Dee et al.*, 1985) and oceanography (*Ghil and Malanotte-Rizzoli*, 1991).

Essential bibliography for Part I

- Asch, M., M. Bocquet, and M. Nodet, *Data Assimilation: Methods, Algorithms, and Applications*, Fundamentals of Algorithms, SIAM, Philadelphia, 2016. (Chapter 3).
- Carrassi, A., M. Bocquet, L. Bertino and G. Evensen. *Data assimilation in the geosciences - An overview on methods, issues and perspectives*. WIREs Climate Change, 9:e535, 2018.
- Jazwinski, A.H., *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970. (Chapters 2 and 3).
- Reich, S., and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, Cambridge, 2015. (Chapters 2, 4 and 5).
- Wilke and Berliner, *A Bayesian tutorial for data assimilation*. Physica D, 203, 1-16, 2007.