Introduction to

# Data Assimilation

and the

# Ensemble Kalman Filter

Based on PhD thesis of Patrick N. Raanes

Edition 3
NORCE, February 2019

# Contents


**1 Introduction** **1**
- 1.1 Data assimilation . . . 1
- 1.2 Principles of sequential inference . . . 2

**2 The ensemble Kalman filter** **6**
- 2.1 The Kalman filter . . . 6
- 2.2 Ensemble preliminaries . . . 7
- 2.3 The EnKF algorithm . . . 8
- 2.4 Properties in the linear-Gaussian case . . . 10
- 2.5 The effects of a small ensemble size . . . 13
- 2.6 Summary and discussion . . . 15

**3 Numerical twin experiments** **16**
- 3.1 RMSE averages . . . 17
- 3.2 Considerations on the metric . . . 18
- 3.3 Linear advection . . . 19
- 3.4 Lorenz-63 . . . 20
- 3.5 Lorenz-96 . . . 21

**4 The square root analysis update: The ETKF** **22**
- 4.1 Method . . . 22
- 4.2 The symmetric square root . . . 23
- 4.3 Efficient computation . . . 24
- 4.4 Random rotations . . . 24

**A The Kalman filter in detail** **25**
- A.1 Matrix inversion identities . . . 25
- A.2 Kalman filter derivation . . . 26
- A.3 Interpreting the scalar KF . . . 29

**B The SVD, linear inverse problems, and the pseudoinverse** **31**
- B.1 The SVD . . . 31
- B.2 Linear inverse problems . . . 32
- B.3 The pseudoinverse . . . 33

**C Notation** **36**

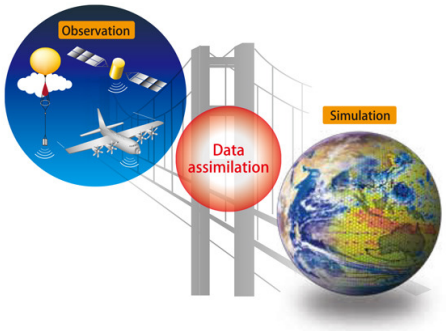
**Bibliography** **38**

# Chapter 1

# Introduction

## 1.1 Data assimilation

Data assimilation (DA) is the process of combining model forecasts with observation data; it is a set of techniques specialized for sequential, statistical inference on dynamic model variables, in addition to static parameters [133]. The typical application of DA is forecast initialization, but it is also used for reanalysis, i.e. the estimation of historical states.



DA "bridges" data and models.[1]

Modern DA builds on "state estimation" techniques developed in control systems engineering. For example, the Kalman filter (section 2.1) was originally developed to steer the Apollo mission spacecraft [49, 85]. In this case, the state variables to be estimated are the location and the (linear and angular) velocity and acceleration; the dynamics are the kinematic equations; and the observations come from accelerometers, gyroscopes, and sextant measurements. In DA, the aim is to condition the estimate on all of the available data; indeed, if the velocity (resp. acceleration) of the spacecraft example is assumed to be a fixed but unknown parameter, then the Kalman filter is the recursive, on-line solution to the first (resp. second) degree polynomial linear regression problem for the trajectory of the spacecraft [114].

In contrast to the above example, in geoscience, the state variables are often the result of the discretization of a physical field, and may number in the billions [95], and the dynamics are typically nonlinear. Examples of geoscientific DA include

- "History matching" in reservoir engineering for oil production planning and optimization [1, 30, 39, 100, 106]. Here, the state variables include pressure and saturation fields; fixed, but unknown, parameters include permeability and porosity fields; the dynamics are modelled using a combination of Darcy's law, material balance, and thermodynamic equations of state.
- Meteorology and weather forecast initialization [10, 75, 78]. Here, the state variables include velocity, pressure, and temperature fields; the dynamics consist of the Navier-Stokes equations, the first law of thermodynamics and the ideal gas law; observations come from diverse satellite instruments, weather stations, and buoys.
- Epidemiology analysis and forecast initialization [8, 59, 94, 115]. Here, the state variable is the number count of people susceptible, infected, and recovered; the dynamics are modelled by a set of coupled differential equations, which may be spatially distributed using agent-based modelling or diffusion; observation include compiled reports from government health agencies and "Google Flu Trends".

[1] Source and permission: Data Assimilation Research Team, `www.aics.riken.jp`.

Other application areas of DA include oceanography [13, 14, 111, 135], climatology on Earth [9] and Mars [58], hydrology [105, 126], atmospheric chemistry and air quality analysis [53], and forest fire prediction [86].

## 1.2 Principles of sequential inference

Several approaches to the DA problem can be classified as Bayesian, sequential inference. Its main principles are introduced here.

### 1.2.1 The hidden Markov model

Following Jazwinski [66], suppose the state and observation, $\boldsymbol{x}_t \in \mathbb{R}^M$ and $\boldsymbol{y}_t \in \mathbb{R}^P$ respectively, are generated for sequentially increasing $t$, by a dynamical model, $\mathcal{M} : \mathbb{R}^M \to \mathbb{R}^M$, and a measurement model, $\mathcal{H} : \mathbb{R}^M \to \mathbb{R}^P$, as follows:

$$\boldsymbol{x}_{t+1} = \mathcal{M}(\boldsymbol{x}_t) + \boldsymbol{q}_t\,, \qquad t = 0, 1, \ldots\,, \tag{1.1}$$

$$\boldsymbol{y}_t = \mathcal{H}(\boldsymbol{x}_t) + \boldsymbol{r}_t\,, \qquad t = 1, 2, \ldots\,, \tag{1.2}$$

where the Gaussian white noise processes $\{\boldsymbol{q}_t \ ; \ t = 0, 1, \ldots\}$ and $\{\boldsymbol{r}_t \ ; \ t = 1, 2, \ldots\}$, and the initial condition, $\boldsymbol{x}_0$, are specified by:

$$\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{P}_0)\,, \qquad \boldsymbol{q}_t \sim \mathcal{N}(0, \mathbf{Q})\,, \qquad \boldsymbol{r}_t \sim \mathcal{N}(0, \mathbf{R})\,, \tag{1.3}$$

where $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ is the multivariate Gaussian probability law with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. Note that we do not use upper and lowercase to distinguish between random variables and their realizations. Generalizations to time-dependent $\mathbf{Q}, \mathbf{R}, \mathcal{M}$, and $\mathcal{H}$ are straightforward for all of the theory developed in this text. The models and parameters, $\mathcal{M}$, $\mathcal{H}$, $\boldsymbol{\mu}_0$, $\mathbf{P}_0, \mathbf{Q}$ and $\mathbf{R}$, are all assumed known.

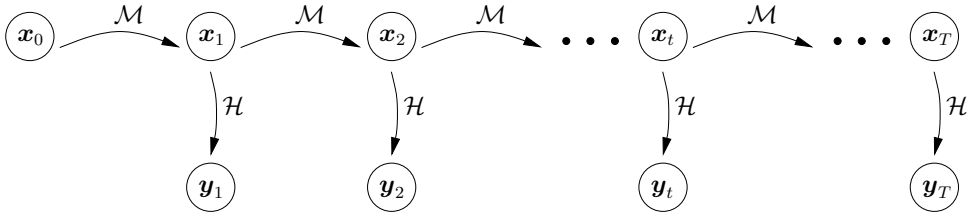



**Figure 1.1:** Diagram of the HMM of eqns. (1.1) and (1.2). The arrows indicate causality.

The process $\{(\boldsymbol{x}_t, \boldsymbol{y}_t) \ ; \ t = 1, 2, \ldots\}$, illustrated in Fig. 1.1, constitutes a hidden Markov model (HMM): a sequence of (possibly) hidden states linked together by a forecast law, such as eqn. (1.1), which are only observed through an observation law, such as eqn. (1.2). For generic random variables (possibly vectors) $\boldsymbol{u}$ and $\boldsymbol{v}$, denote $p(\boldsymbol{u}|\boldsymbol{v})$ the probability density function (pdf) of $\boldsymbol{u}$ conditioned on $\boldsymbol{v}$. The fundamental properties of a HMM, which explain its name, and can be derived from eqns. (1.1) to (1.3), are the independence relations

$$p(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_{0:t}) = p(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t)\,, \tag{1.4}$$

$$p(\boldsymbol{y}_t \mid \boldsymbol{x}_{0:T}) = p(\boldsymbol{y}_t \mid \boldsymbol{x}_t)\,, \tag{1.5}$$

for any two time indices $t \leq T$, where the colon is used to indicate a sequence, e.g. $\boldsymbol{x}_{0:T} = \{\boldsymbol{x}_t \ ; \ t = 0, 1, \ldots, T\}$. Being very general, but possessive of the above properties (and more [22]), the HMM is a useful abstraction for building inference techniques for dynamical systems [91, §10.2.2].

An advantage of ensemble DA methods is that they are non-invasive; indeed, it is not uncommon that the dynamical model, $\mathcal{M}$, is only available as a binary executable. Nevertheless, though $\mathcal{M}$ is then treated as a generic "black-box", the dynamics of the forecast law, eqn. (1.1), are typically a discretization of an underlying time-continuous, physical system. For example, if $\mathcal{M}(\boldsymbol{x}) = \boldsymbol{x} + \Delta t\, u(\boldsymbol{x})$, for some function $u$, and if $\mathbf{Q}$ scales with $\sqrt{\Delta t}$, then the stochastic difference equation, eqn. (1.1), may be regarded as the Euler-Maruyama discretization of the time-continuous, $M$-vector, Itô stochastic differential equation known as the (simplified) Langevin equation,

$$d\boldsymbol{x} = u(\boldsymbol{x}, t)\, dt + d\boldsymbol{q}, \quad t > 0\,. \tag{1.6}$$

Conditions for the existence of solutions to eqn. (1.6) are discussed by Jazwinski [66, §4.4], while Gardiner [45, §10] provides a discussion of the convergence of eqn. (1.1) to eqn. (1.6) as $\Delta t \to 0$. Although $\boldsymbol{x}(t)$ is a random variable (process), its pdf, $p(\boldsymbol{x}; t)$, evolves *deterministically* in time, according to the Fokker-Planck equation [66, §4.4],

$$\frac{\partial p}{\partial t} = -\nabla.(up) + \frac{1}{2}\nabla.(\mathbf{Q}\nabla p)\,, \tag{1.7}$$

with far-field boundary conditions $p(\boldsymbol{x}; t) \to 0$ as $\|\boldsymbol{x}\|_2 \to \infty$ for all $t > 0$. An example solution to eqn. (1.7) with a scalar $\boldsymbol{x}$ is illustrated in Fig. 1.2.

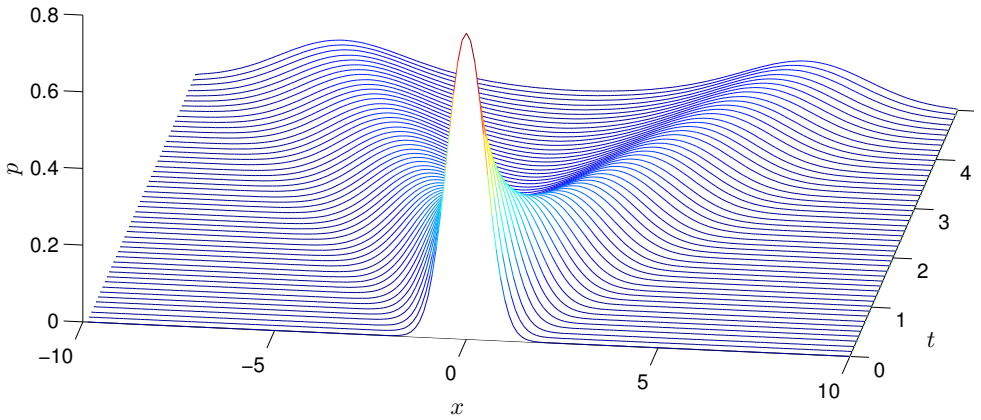
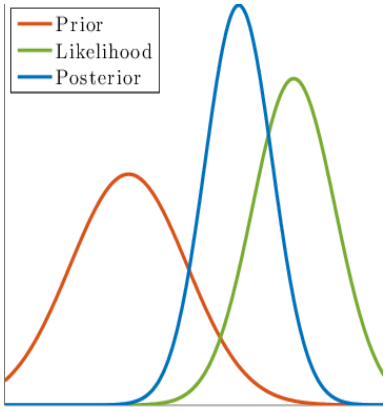



**Figure 1.2:** Illustration of the integration of the Fokker-Planck equation (1.7) with $u(x) = \arctan(x)$, $\mathbf{Q} = 0.5$, and $p(x) = \mathcal{N}(x|0, 1/4)$ at $t = 0$.

### 1.2.2 Bayesian data assimilation

**(a)** P.S. Laplace, 1749 – 1827. Pioneer of inference.

**(b)** C.F. Gauss, 1777 – 1855. Pioneer of inference.




**(c)** Illustration of Bayes' rule, eqn. (1.8), with Gaussian distributions.

DA is now formulated as statistical inference on a HMM. The principal objective in Bayesian inference is to compute $p(\boldsymbol{x}|\boldsymbol{y})$, resulting from Bayes' rule:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})\,, \tag{1.8}$$

where $\boldsymbol{y}$ is any data available, and $\boldsymbol{x}$ is the unknown that we wish to estimate. The constant of proportionality is $1/\,p(\boldsymbol{y})$. In other words, the proportionality is "with respect to $\boldsymbol{x}$". The pdf $p(\boldsymbol{x}|\boldsymbol{y})$ is known as the *posterior*, resulting from the pointwise multiplication of the *likelihood*, $p(\boldsymbol{y}|\boldsymbol{x})$, our knowledge from the observations, and the *prior*, $p(\boldsymbol{x})$, a pdf in $\boldsymbol{x}$ quantifying our prior information about $\boldsymbol{x}$. Note the usage of "overloading" notation: $p$ is the pdf of the random variable identified by the symbol of its input argument.

In DA, the unknowns are the state vectors of the dynamics, $\boldsymbol{x}_{0:T}$, and the data are the observation vectors, $\boldsymbol{y}_{1:T}$. The objective is then to compute the pdf, $p(\boldsymbol{x}_{0:t}|\boldsymbol{y}_{1:T})$, though some of these time indices may be dropped, depending on the application of the estimation. Suppose we are only interested in the marginal pdf, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:T})$; then the estimation problem is called

- smoothing if $t < T$;
- filtering if $t = T$;
- forecasting (or prediction) if $t > T$.

The nomenclature is also used when the aim is to compute a joint pdf (i.e. where there are multiple time indices on $\boldsymbol{x}$).

Before deriving the filter recursions, suppose we want to obtain the joint posterior, $p(\boldsymbol{x}_{0:T}|\boldsymbol{y}_{1:T})$, for sequentially increasing time indices, $T$. Using Bayes' rule (1.8),

$$p(\boldsymbol{x}_{0:T}|\boldsymbol{y}_{1:T}) \propto p(\boldsymbol{x}_{0:T})\, p(\boldsymbol{y}_{1:T}|\boldsymbol{x}_{0:T})\,. \tag{1.9}$$

Next, eqns. (1.4) and (1.5) imply that

$$p(\boldsymbol{x}_{0:T}|\boldsymbol{y}_{1:T}) \propto p(\boldsymbol{x}_0) \prod_{t=1}^{T} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, p(\boldsymbol{y}_t|\boldsymbol{x}_t)\,, \tag{1.10}$$

inducing the recursion

$$p(\boldsymbol{x}_{0:t}|\boldsymbol{y}_{1:t}) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t)\, p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, p(\boldsymbol{x}_{0:t-1}|\boldsymbol{y}_{1:t-1})\,. \tag{1.11}$$

Thus joint, sequential inference can be performed by recursively appending the factor $p(\boldsymbol{y}_t|\boldsymbol{x}_t)\, p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ to the previous posterior.

### 1.2.3 Filtering

Filtering is the procedure whereby one computes the pdf $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$, typically in order to initialize a probabilistic forecast. The equations may be derived by marginalization of the joint inference equations (1.11), but is abbreviated here by deliberately seeking a recursion in the marginal pdf. To that end, assume that $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ is available, or $p(\boldsymbol{x}_0)$ if $t = 1$. Then, by the Markov property, eqn. (1.4), the "forecast" pdf is given by the Chapman-Kolmogorov equation:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})\, d\boldsymbol{x}_{t-1}\,, \tag{1.12}$$

where the integral is over the state space, $\mathbb{R}^M$. Next, by virtue of the other conditional independence property of the HMM, eqn. (1.5), and Bayes' rule, the filtered, posterior, or "analysis", pdf is given by:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t)\, p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})\,. \tag{1.13}$$

These two equations constitute a cycle that can be repeated sequentially in time to obtain the filtered pdfs.

As mentioned in section 1.2.2, filtering is the archetypical sequential inference procedure. Indeed, in the linear-Gaussian case, the recursions of eqns. (1.12) and (1.13) reduce to a famous set of matrix formulae known as the Kalman filter. In geoscientific DA, however, $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ may be very high-dimensional, and $\mathcal{M}$ and $\mathcal{H}$ may be highly nonlinear (section 1.1). Therefore, only approximate solution methods are possible. The family of techniques deriving from the ensemble Kalman filter (EnKF) provides a set of such methods.

# Chapter 2

# The ensemble Kalman filter

The recursions established in section 1.2.3 avoid repeating the calculations in full after each time step. The recursions represent the fundamental exploitation of the time-sequential structure of the estimation problem on HMMs, thus reducing the problem's dimensionality and complexity. Nevertheless, the size and nonlinearity of the systems remain challenging, necessitating the use of approximate solution methods. One approximate solution method is the ensemble Kalman filter (EnKF).

The EnKF builds on the Kalman filter (KF), which is developed in section 2.1. Sections 2.2 and 2.3 develop the basic theory of the EnKF. Sections 2.4 and 2.5 discuss its properties, including bias and consistency. An initial survey of related efforts for solving the filtering problem is provided in section 2.6.

## 2.1 The Kalman filter

The KF is the closed-form solution to the Bayesian filtering equations (1.12) and (1.13) obtained in the linear, Gaussian case. This section provides a summary of the KF; a complete derivation can be found in section A.2. Suppose the HMM, eqns. (1.1) and (1.2), is linear, i.e. that $\mathcal{M}(\boldsymbol{x}) = \mathbf{F}\boldsymbol{x}$ and $\mathcal{H}(\boldsymbol{x}) = \mathbf{H}\boldsymbol{x}$, where $\mathbf{F} \in \mathbb{R}^{M^2}$ and $\mathbf{H} \in \mathbb{R}^{P \times M}$. It can then be shown that the pdfs involved remain Gaussian throughout the filtering process. Thus, the pdfs are fully characterized by their first two moments, labelled as follows:

$$\boldsymbol{x}^{\mathrm{f}} = \mathbb{E}(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1})\,, \qquad \mathbf{P}^{\mathrm{f}} = \mathbb{V}\mathrm{ar}(\boldsymbol{x}_t | \boldsymbol{y}_{1:t-1})\,, \tag{2.1}$$

$$\boldsymbol{x}^{\mathrm{a}} = \mathbb{E}(\boldsymbol{x}_t | \boldsymbol{y}_{1:t})\,, \qquad \mathbf{P}^{\mathrm{a}} = \mathbb{V}\mathrm{ar}(\boldsymbol{x}_t | \boldsymbol{y}_{1:t})\,. \tag{2.2}$$

where $\mathbb{E}(.)$ and $\mathbb{V}\mathrm{ar}(.)$ are the multivariate expectation and variance operators, respectively (i.e. $\mathbb{V}\mathrm{ar}(.)$ outputs a square *matrix*). As is convention [65], the superscripts $\mathcal{M}$ or $a$ are used to distinguish between forecast and analysis objects, respectively. The explicit time indices are dropped because the focus will typically be restricted to a single index anyway.

Assume initially that $p(\boldsymbol{x}_{t-1} \mid \boldsymbol{y}_{1:t-1}) = \mathcal{N}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}})$, where $\boldsymbol{x}^{\mathrm{a}}$ and $\mathbf{P}^{\mathrm{a}}$ are associated with the time index, $t-1$. As shown in section A.2, the forecast equation (1.12) yields $p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t-1}) = \mathcal{N}(\boldsymbol{x}_t \mid \boldsymbol{x}^{\mathrm{f}}, \mathbf{P}^{\mathrm{f}})$ with

$$\boldsymbol{x}^{\mathrm{f}} = \mathbf{F}\boldsymbol{x}^{\mathrm{a}}\,, \tag{2.3}$$

$$\mathbf{P}^{\mathrm{f}} = \mathbf{F}\mathbf{P}^{\mathrm{a}}\mathbf{F}^{\mathsf{T}} + \mathbf{Q}\,, \tag{2.4}$$

where the time interval between the *indices* $t$ and $t-1$ should be included as a scaling of $\mathbf{Q}$. Next, the analysis equation (1.13) implies that $p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}) = \mathcal{N}(\boldsymbol{x}_t \mid \boldsymbol{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}})$, where $\boldsymbol{x}^{\mathrm{a}}$ and $\mathbf{P}^{\mathrm{a}}$

are now associated with $t$, and are given by

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{f}} + \mathbf{K}[\boldsymbol{y}_t - \mathbf{H}\boldsymbol{x}^{\mathrm{f}}]\,, \tag{2.5}$$

$$\mathbf{P}^{\mathrm{a}} = [\mathbf{I}_m - \mathbf{K}\mathbf{H}]\mathbf{P}^{\mathrm{f}}\,, \tag{2.6}$$

where the "Kalman gain matrix", $\mathbf{K} \in \mathbb{R}^{M\times P}$, is

$$\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\right)^{-1}\,. \tag{2.7}$$

The KF thus consists in repeating the matrix computations of eqns. (2.3) to (2.6) for sequentially increasing time indices $t$.

The Kalman gain can also be derived from the perspective of optimal point estimation. Assuming Gaussianity as before, any convex loss function yields the posterior mean (or median or mode) as its optimum [116], and thereby the KF equations are recovered. Alternatively, without assuming Gaussianity, it can be derived as the minimum square error *linear estimator* [3, Ths. 2.1,2.3]. Derivations under the heading of orthogonal projections or recursive least squares [66, §7.3] can be formulated through one of the above [3, Th. 2.5].

The applicability of the KF equations can be extended to the case of nonlinear models by linearization. That is, while the formulae (2.3) and (2.5) are computed using the full nonlinear models, formulae (2.4), (2.6) and (2.7) are computed with $\mathbf{F}$ and $\mathbf{H}$ being the linearization of $\mathcal{M}$ and $\mathcal{H}$ respectively. This first order approximation to the exact solution is known as the extended Kalman filter.

## 2.2 Ensemble preliminaries

The EnKF is an algorithm that approximately generates an ensemble, i.e. an iid sample, $\boldsymbol{x}_{1:N} = \{\boldsymbol{x}_n\ ;\ n \in 1{:}N\}$, from the Bayesian filtering distributions (1.12) and (1.13), recursively in time, for sequentially increasing time indices, $t$. More vaguely, it is also sometimes helpful to think of the ensemble as a "cloud" in the phase space of $\boldsymbol{x}$ representing its pdf.

The positive integer $N$ is used to denote ensemble size, while $M$ and $P$ are used to denote state and observation vector sizes, and a colon is used to indicate an integer sequence. For convenience, we concatenate all of the state realizations, or "ensemble members", into the "ensemble matrix":

$$\mathbf{E} = \Big[\boldsymbol{x}_1,\ \ \ldots\ \ \boldsymbol{x}_n,\ \ \ldots\ \ \boldsymbol{x}_N\Big] \in \mathbb{R}^{M\times N}\,. \tag{2.8}$$

The overhead bar is used to designate the ensemble-estimate counterparts to the exact mean and covariance matrix of eqns. (2.1) and (2.2);

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n\,, \qquad\qquad \bar{\mathbf{P}} = \frac{1}{N-1}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^{\mathsf{T}}\,. \tag{2.9}$$

These are the canonical, unbiased [e.g. 90, §3.1] sample estimators. However, in EnKF contexts they should be regarded mainly as a conventional notational tool; indeed, a significant body of research (e.g. Bocquet et al. [20]) deals with adjusting these estimates. For a given ensemble, its "anomalies" are defined as

$$\mathbf{X} = \Big[\boldsymbol{x}_1 - \bar{\boldsymbol{x}},\ \ \ldots\ \ \boldsymbol{x}_n - \bar{\boldsymbol{x}},\ \ \ldots\ \ \boldsymbol{x}_N - \bar{\boldsymbol{x}}\Big]\,. \tag{2.10}$$

Note that the ensemble mean, anomalies, and covariance matrix can be conveniently expressed and computed as

$$\bar{\boldsymbol{x}} = \frac{1}{N}\mathbf{E}\mathbb{1}\,, \qquad \mathbf{X} = \mathbf{E}(\mathbf{I}_N - \mathbb{1}\mathbb{1}^{\mathsf{T}}/N)\,, \qquad (N-1)\bar{\mathbf{P}} = \mathbf{X}\mathbf{X}^{\mathsf{T}}\,, \tag{2.11}$$

where $\mathbb{1}$ is the vector of ones of length $N$, and $(\mathbf{I}_N - \mathbb{1}\mathbb{1}^\mathsf{T}/N) \in \mathbb{R}^{N^2}$ should be recognized as $\mathbf{\Pi}_{\mathbb{1}}^{\perp}$, the orthogonal projector[1] onto $\mathsf{range}(\mathbb{1})^{\perp}$, the orthogonal complement space to $\mathsf{range}(\mathbb{1})$.

As will be seen, these relations are not just helpful because they (i) abbreviate eqns. (2.9) and (2.10) and avoid the member index, $n$; they also (ii) highlight linearity aspects of the operations involved; (iii) give insight into subspace rank issues; and (iv) emphasize that sometimes the ensemble may be seen as a deterministic parameterization of the pdf.

## 2.3 The EnKF algorithm

As with the KF, the EnKF consists of the recursive application of a forecast step and an analysis step. This section follows the traditional [35] template, presenting the EnKF as the "the Monte Carlo version of the KF where the state covariance is estimated by the ensemble covariance". It is not obvious that this postulated method should work; indeed, it is only justified upon inspection of its properties, deferred to section 2.4. An improved *derivation* is given in Raanes [103, §. 6.2].

As in eqn. (2.1), the time indices of the state and conditioning are implied by the superscript $\mathcal{M}$ or $a$ for the ensemble. This indicates that $\boldsymbol{x}^{\mathrm{f}}_{1:N}$ (resp. $\boldsymbol{x}^{\mathrm{a}}_{1:N}$) is a forecast (resp. analysis) ensemble, and is also used for the derivative objects, $\mathbf{E}, \mathbf{X}, \bar{\boldsymbol{x}}, \mathbf{P}$.

### 2.3.1 Forecast step

For a given, implicit, time index, $t$, assume $\boldsymbol{x}^{\mathrm{a}}_{1:N}$ is an iid sample from $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$, which is not necessarily Gaussian. The forecast step of the EnKF consists of a Monte Carlo simulation of eqn. (1.1): for each $n \in 1{:}N$, $\boldsymbol{x}^{\mathrm{a}}_n$ is propagated in time through the forecast, dynamical model, $\mathcal{M}$, and a simulated noise realization, $\boldsymbol{q}_n$ is added to it

$$\forall n, \quad \boldsymbol{x}^{\mathrm{f}}_n = \mathcal{M}(\boldsymbol{x}^{\mathrm{a}}_n) + \boldsymbol{q}_n \,, \tag{2.12}$$

$$\text{or,} \quad \mathbf{E}^{\mathrm{f}} = \mathcal{M}(\mathbf{E}^{\mathrm{a}}) + \mathbf{D}_{\mathrm{mod}} \,, \tag{2.13}$$

where $\mathcal{M}$ is applied column-wise to $\mathbf{E}^{\mathrm{a}}$, and the columns of $\mathbf{D}_{\mathrm{mod}}$ are sampled iid from $\mathcal{N}(0, \mathbf{Q})$. The ensemble, $\boldsymbol{x}^{\mathrm{f}}_{1:N}$, is then an iid sample from the forecast pdf, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$.

**Lemma 2.1.**
*Assuming the ensemble (forecast or analysis) is iid drawn from a non-degenerate pdf, then, almost surely,*

$$\mathsf{rank}(\mathbf{E}) = \mathsf{min}(M, N) \,, \qquad \mathsf{rank}(\mathbf{X}) = \mathsf{min}(M, N-1) \,. \tag{2.14}$$

*Proof.* The size of $\mathbf{E}$ and $\mathbf{X}$ imply that their ranks are bounded by $M$ and $N$. For $N \leq M$, any strict subset of the ensemble members spans a strict subspace of $\mathbb{R}^M$, which is of (probability) measure zero. This can be used inductively to prove almost sure linear independence between the ensemble members. However, the fact that $\mathbf{X}\mathbb{1} = 0$, as seen from eqn. (2.11), reduces the rank of $\mathbf{X}$ by one. More details are given by Gupta and Nagar [50, Th. 3.2.1] and Muirhead [90, Th. 3.1.4]. □

Because of possible nonlinear dynamics, and finite precision, the assumption of Lemma 2.1 is frequently not satisfied in practice. But it provides useful upper bounds on the ranks that are convenient to keep in mind when manipulating the ensemble matrices.

[1] A matrix $\mathbf{\Pi}$ is an orthogonal projection matrix if it satisfies $\mathbf{\Pi}^2 = \mathbf{\Pi}$ and $\mathbf{\Pi} = \mathbf{\Pi}^\mathsf{T}$.

### 2.3.2 Analysis step

The analysis update of the ensemble is given by:

$$\forall n, \quad \boldsymbol{x}_n^{\mathrm{a}} = \boldsymbol{x}_n^{\mathrm{f}} + \bar{\mathbf{K}} \left\{ \boldsymbol{y} - \boldsymbol{r}_n - \mathcal{H}(\boldsymbol{x}_n^{\mathrm{f}}) \right\} , \tag{2.15}$$

$$\text{or,} \quad \mathbf{E}^{\mathrm{a}} = \mathbf{E}^{\mathrm{f}} + \bar{\mathbf{K}} \left\{ \boldsymbol{y}\mathbb{1}^{\mathsf{T}} - \mathbf{D}_{\mathrm{obs}} - \mathcal{H}(\mathbf{E}^{\mathrm{f}}) \right\} , \tag{2.16}$$

where the "observation perturbations", $\boldsymbol{r}_n$, are sampled iid from $\mathcal{N}(0, \mathbf{R})$ and form the columns of $\mathbf{D}_{\mathrm{obs}}$, and $\mathcal{H}$ is applied column-wise to $\mathbf{E}^{\mathrm{f}}$. If the forecast distribution were Gaussian, $\mathcal{H}$ linear,

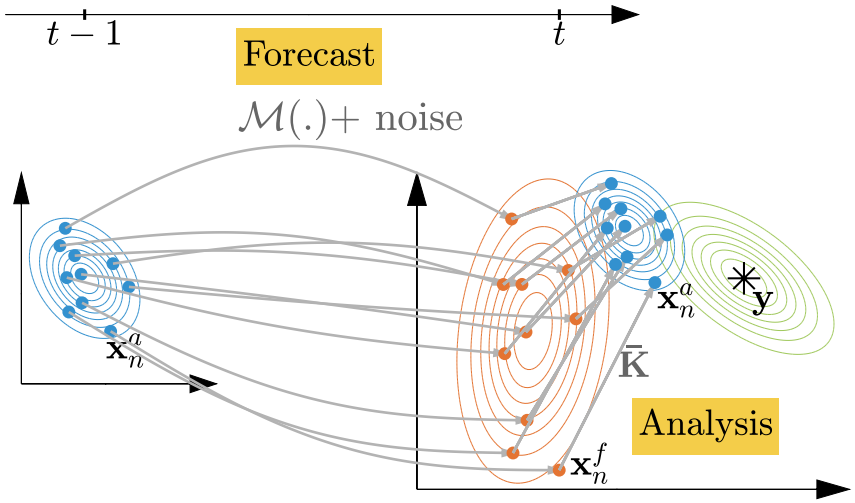



**Figure 2.1:** Illustration of one assimilation cycle of the EnKF. The ensemble members are shown propagated by the dynamics, eqn. (2.13), and then updated by eqn. (2.16). The colouring corresponds to figure (c) on page 4. The truth (black) is assumed drawn from the same distribution as the ensemble.

and if $\bar{\mathbf{K}}$ (detailed below) were the actual Kalman gain, $\mathbf{K}$ of eqn. (2.7), then the columns of $\mathbf{E}^{\mathrm{a}}$ would provide an updated ensemble, $\boldsymbol{x}_{1:N}^{\mathrm{a}}$, sampled iid from the analysis distribution, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$. This is demonstrated along the lines of section 2.4.1 by showing that the analysis ensemble members maintain Gaussianity and that $\mathbb{E}(\boldsymbol{x}_n^{\mathrm{a}}) = \boldsymbol{x}^{\mathrm{a}}$ and $\mathbb{V}\mathrm{ar}(\boldsymbol{x}_n^{\mathrm{a}}) = \mathbf{P}^{\mathrm{a}}$ [96]. In any case, $\bar{\mathbf{K}}$ is not exact, but rather the ensemble estimate obtained by substituting $\bar{\mathbf{P}}^{\mathrm{f}}$ for $\mathbf{P}^{\mathrm{f}}$ in eqn. (2.7), and therefore the analysis ensemble is only approximately drawn from $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$. If $\mathbf{H}$ is defined, then the ensemble-estimated Kalman gain is obtained by substituting $\bar{\mathbf{P}}^{\mathrm{f}}$ for $\mathbf{P}^{\mathrm{f}}$ in eqn. (2.7)

$$\bar{\mathbf{K}} = \bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}(\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}} + \mathbf{R})^{-1} . \tag{2.17}$$

However, $\bar{\mathbf{K}}$ can also be estimated even if $\mathcal{H}$ is nonlinear, and $\mathbf{H}$ undefined; instead of substituting for $\mathbf{P}^{\mathrm{f}}$, the ensemble is used to individually estimate (i) $\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}$: the cross-covariance between $\boldsymbol{x}^{\mathrm{f}}$ and $\mathcal{H}(\boldsymbol{x}^{\mathrm{f}})$, and (ii) $\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}$: the covariance matrix of $\mathcal{H}(\boldsymbol{x}^{\mathrm{f}})$. Using the canonical estimators, as with $\mathbf{P}^{\mathrm{f}}$ of eqn. (2.9), yields

$$\bar{\mathbf{K}} = \mathbf{X}\mathbf{Y}^{\mathsf{T}}(\mathbf{Y}\mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R})^{-1} , \tag{2.18}$$

where $\mathbf{Y} = \mathcal{H}(\mathbf{E}^{\mathrm{f}})(\mathbf{I}_N - \mathbb{1}\mathbb{1}^{\mathsf{T}}/N) \in \mathbb{R}^{P \times N}$ are the anomalies of the *observed* ensemble. When eqn. (2.17) is defined, eqn. (2.18) agrees with it.

Note that $\bar{\mathbf{P}}^{\mathrm{f}}$ is not explicitly computed by eqn. (2.18). Thus, even if $\mathbf{H}$ is defined, eqn. (2.18) saves significant computation compared to eqn. (2.17) if $P < M$. As described in section 4.1, there is also an ensemble-space formulation of $\bar{\mathbf{K}}$, requiring the inversion of an $N \times N$ matrix. If $\mathbf{R}$ is readily invertible [21] this can be significantly more computationally cheap and stable, compared to formula (2.18). Unless $P$ is significantly larger than $N$, the time savings might not be crucial, however, because the forecast model integration typically takes up much more time.

The connection between eqn. (2.17) and eqn. (2.18) indicates that the ensemble somehow performs statistical *linearization* of $\mathcal{H}$. The connection to linear regression was recognized by Anderson [5], and its understanding is important for iterative methods [31]. However, it is worth noting that the nonlinearities of $\mathcal{H}$ can also be concealed by an augmented forecast model, yielding a state vector that includes the observations. The resulting observation operator then becomes merely the matrix that selects the observation components of the augmented state vector. The analysis formulae of this linear, augmented system, can then be dismantled into blocks, whereupon eqn. (2.18) is recovered. This text is not particularly concerned with nonlinear observation operators, $\mathcal{H}$. However, all of the developments employ statistical linearization, and are thus fully applicable in the nonlinear case.

## 2.4 Properties in the linear-Gaussian case

Given the assumptions of the EnKF, it is desirable that $\bar{\mathbf{P}}^{\mathrm{f/a}} = \mathbf{P}^{\mathrm{f/a}}$ and $\bar{\boldsymbol{x}}^{\mathrm{f/a}} = \boldsymbol{x}^{\mathrm{f/a}}$ throughout the DA process, at least in the linear-Gaussian case, where the KF solves the Bayesian filtering equations exactly. This section provides the justification for the EnKF analysis update, eqn. (2.16), by showing that it satisfies these conditions in the expected sense, where the expectation is with respect to $\mathbf{D}_{\mathrm{obs}}$. Related concepts are also discussed. The subscript on $\mathbf{D}$ is dropped in this section.

### 2.4.1 Conformality and unbiasedness

By "conformality" we mean that the ensemble estimates satisfy the relations obtained by *individually* replacing, in the KF equations, the exact moments of the forecast distribution by their ensemble estimates. It is not a particularly distinguished property in itself, but is used as a lemma for showing other results such as unbiasedness.

**Proposition 2.1** – Unbiasedness of the EnKF in the mean**.**
*If the column-mean of* $\mathbf{D}$ *is zero, i.e.*

$$\mathbf{D}\mathbb{1} = 0\,, \tag{2.19}$$

*then the mean EnKF update conforms to the KF mean update. Furthermore, with expectation over* $\mathbf{D}$*, the EnKF analysis update is unbiased in the mean.*

*Proof.* If eqn. (2.19) holds, then eqn. (2.16) yields

$$\bar{\boldsymbol{x}}^{\mathrm{a}} = \frac{1}{N}\mathbf{E}^{\mathrm{a}}\mathbb{1} = \frac{1}{N}\mathbf{E}^{\mathrm{f}}\mathbb{1} + \frac{1}{N}\bar{\mathbf{K}}\left(\boldsymbol{y}\mathbb{1}^{\mathsf{T}} - \mathbf{D} - \mathbf{H}\mathbf{E}^{\mathrm{f}}\right)\mathbb{1} = \bar{\boldsymbol{x}}^{\mathrm{f}} + \bar{\mathbf{K}}\left[\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}^{\mathrm{f}}\right]\,, \tag{2.20}$$

and the mean EnKF update conforms to the KF mean update (2.5). By virtue of the linearity of eqn. (2.20) with respect to eqn. (2.19), which holds in the expected sense, the mean EnKF update is unbiased with respect to $\mathbf{D}$. □

**Proposition 2.2** – Unbiasedness of the EnKF in the covariance**.**
*If, in addition to the assumption of Proposition 2.1, the sample variance of* $\mathbf{D}$ *is* exactly $\mathbf{R}$*, and its sample cross-covariance with the anomalies zero, i.e.*

$$\frac{1}{N-1}\mathbf{D}\mathbf{D}^{\mathsf{T}} = \mathbf{R}\,, \tag{2.21a}$$

$$\mathbf{X}^{\mathrm{f}}\mathbf{D}^{\mathsf{T}} = 0\,, \tag{2.21b}$$

*then the resulting covariance update of the EnKF conforms with the KF covariance update. Furthermore, with expectation over* $\mathbf{D}$*, the EnKF analysis update is unbiased in the covariance.*

*Proof.* First, compute the updated anomalies, $\mathbf{X}^{\mathrm{a}}$, by inserting eqn. (2.16) for $\mathbf{E}^{\mathrm{a}}$:

$$\mathbf{X}^{\mathrm{a}} = \mathbf{E}^{\mathrm{a}}(\mathbf{I}_N - \mathbb{1}\mathbb{1}^{\mathsf{T}}/N) \tag{2.22}$$

$$= \mathbf{X}^{\mathrm{f}} - \bar{\mathbf{K}}\left[\mathbf{D} + \mathbf{H}\mathbf{X}^{\mathrm{f}}\right], \tag{2.23}$$

where eqn. (2.19) has been used. Hence, the updated ensemble covariance matrix is

$$\bar{\mathbf{P}}^{\mathrm{a}} = \frac{1}{N-1}\mathbf{X}^{\mathrm{a}}\mathbf{X}^{\mathrm{a}\mathsf{T}} \tag{2.24}$$

$$= \frac{1}{N-1}\left(\mathbf{X}^{\mathrm{f}} - \bar{\mathbf{K}}\left[\mathbf{D} + \mathbf{H}\mathbf{X}^{\mathrm{f}}\right]\right)\left(\mathbf{X}^{\mathrm{f}} - \bar{\mathbf{K}}\left[\mathbf{D} + \mathbf{H}\mathbf{X}^{\mathrm{f}}\right]\right)^{\mathsf{T}} \tag{2.25}$$

$$= \bar{\mathbf{P}}^{\mathrm{f}} + \bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}} \tag{2.26}$$

$$- \frac{1}{N-1}\left[\mathbf{X}^{\mathrm{f}}\mathbf{D}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} + \bar{\mathbf{K}}\mathbf{D}\mathbf{X}^{\mathrm{f}\mathsf{T}} - \bar{\mathbf{K}}\mathbf{D}\mathbf{X}^{\mathrm{f}\mathsf{T}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{K}}\mathbf{H}\mathbf{X}^{\mathrm{f}}\mathbf{D}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{K}}\mathbf{D}\mathbf{D}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}}\right]$$

$$= (\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\bar{\mathbf{P}}^{\mathrm{f}}(\mathbf{I}_m - \mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}}) \tag{2.27}$$

$$- \frac{1}{N-1}\left[(\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\mathbf{X}^{\mathrm{f}}\mathbf{D}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} + \bar{\mathbf{K}}\mathbf{D}\mathbf{X}^{\mathrm{f}\mathsf{T}}(\mathbf{I}_m - \mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}}) - \bar{\mathbf{K}}\mathbf{D}\mathbf{D}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}}\right],$$

where the identity $(N-1)^{-1}\mathbf{X}^{\mathrm{f}}\mathbf{X}^{\mathrm{f}\mathsf{T}} = \bar{\mathbf{P}}^{\mathrm{f}}$ has been used. Substituting eqns. (2.21) into eqn. (2.26) yields

$$\bar{\mathbf{P}}^{\mathrm{a}} = \bar{\mathbf{P}}^{\mathrm{f}} + \bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}} + \bar{\mathbf{K}}\mathbf{R}\bar{\mathbf{K}}^{\mathsf{T}} \tag{2.28}$$

$$= (\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\bar{\mathbf{P}}^{\mathrm{f}} + \bar{\mathbf{K}}(\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}} + \mathbf{R})\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} \tag{2.29}$$

$$= (\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\bar{\mathbf{P}}^{\mathrm{f}} + \bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} - \bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} \tag{2.30}$$

$$= (\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\bar{\mathbf{P}}^{\mathrm{f}}, \tag{2.31}$$

and thus the covariance of the EnKF update conforms to the KF covariance update (2.6). By virtue of the linearity of eqn. (2.26) with respect to eqns. (2.21), which hold in the expected sense, the covariance of the EnKF update is unbiased with respect to $\mathbf{D}$. □

For completeness it is worth mentioning that before [25, 61] the original EnKF [35] used the analysis update scheme

$$\mathbf{E}^{\mathrm{a}} = \mathbf{E}^{\mathrm{f}} + \bar{\mathbf{K}}\left(\boldsymbol{y}\mathbb{1}^{\mathsf{T}} - \mathbf{H}\mathbf{E}^{\mathrm{f}}\right), \tag{2.32}$$

which lacks the perturbations, $\mathbf{D}$ of eqn. (2.16), and has the effect on eqn. (2.27) of removing its entire second line. Thus $(\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})$ is effectively applied twice on the covariance, meaning that the reduction in the spread of the ensemble is too significant.

The Deterministic EnKF, proposed by Sakov and Oke [109], also uses no perturbations, i.e. $\mathbf{D} = 0$. However, instead of operating on the ensemble, as eqns. (2.16) and (2.32), it updates the ensemble mean, $\bar{\boldsymbol{x}}$ and anomalies, $\mathbf{X}$, separately. This allows it to shift the anomalies only halfway compared to eqn. (2.23), implying

$$\bar{\mathbf{P}}^{\mathrm{a}} = \left(\mathbf{I}_m - \frac{1}{2}\bar{\mathbf{K}}\mathbf{H}\right)\bar{\mathbf{P}}^{\mathrm{f}}\left(\mathbf{I}_m - \frac{1}{2}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}}\right) \tag{2.33}$$

$$= \bar{\mathbf{P}}^{\mathrm{f}} - \frac{1}{2}\bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}} - \frac{1}{2}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} + \bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}} \tag{2.34}$$

$$= (\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\bar{\mathbf{P}}^{\mathrm{f}} + \bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathsf{T}}\bar{\mathbf{K}}^{\mathsf{T}}, \tag{2.35}$$

where the symmetry of $\bar{\mathbf{K}}\mathbf{H}\bar{\mathbf{P}}^{\mathrm{f}}$ can be shown by inserting eqn. (2.18). Thus, if $\bar{\mathbf{K}}\mathbf{H}$ is small (loosely speaking, if the observation uncertainty is larger than the state uncertainty) then the

term that is quadratic in $\bar{\mathbf{K}}\mathbf{H}$ is dominated by the linear term, and hence $\bar{\mathbf{P}}^{\mathrm{a}}$ is only slightly larger than its desired value, $(\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H})\mathbf{P}^{\mathrm{f}}$.

### 2.4.2 Convergence

Propositions 2.1 and 2.2 and Slutsky's theorem [87] suffice to show that

$$\lim_{N\to\infty} \bar{\boldsymbol{x}}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{a}}\,, \tag{2.36a}$$

$$\lim_{N\to\infty} \bar{\mathbf{P}}^{\mathrm{a}} = \mathbf{P}^{\mathrm{a}}\,, \tag{2.36b}$$

with convergence "in probability", provided the same holds for the forecast moments. In other words, the EnKF analysis update is consistent.

At first sight, in the linear-Gaussian case, it would appear that the convergence of the EnKF moments to those of the KF in eqns. (2.36) can be inductively cascaded through time so as to conclude that the EnKF converges to the exact posterior for any time index. Note, however, that $\bar{\mathbf{K}}$ is computed from the ensemble, which is then updated by the same $\bar{\mathbf{K}}$. This destroys the linearity with respect to $\mathbf{E}^{\mathrm{f}}$ of the analysis update (2.16) (that would exist with the true $\mathbf{K}$ and a linear $\mathcal{H}$), and thus the apparent Gaussianity and independence of the resulting ensemble members, $\mathbf{E}^{\mathrm{a}}$. This should give pause before drawing conclusions across multiple time steps. Nevertheless, Le Gland et al. [72] showed that (i) under suitable continuity conditions on $\mathcal{M}$, moment conditions on $p(\boldsymbol{x}_0)$, and with a linear observation operator, the moments of the empirical distribution of the ensemble do converge, and (ii) in the linear-Gaussian case, this limit is that of the actual posterior. The latter was also shown by Mandel et al. [87], whose proof depends on the notion of exchangeability of the ensemble members: as $N \to \infty$, the irrelevance of the ordering of the ensemble is sufficient for a particular law of large numbers to hold, which, in turn, yields convergence.

### 2.4.3 Sampling error and bias

The conditions of eqns. (2.19), (2.21a) and (2.21b) are never realized for finite $N$. The same applies for the ensemble estimates of the forecast moments, $\bar{\boldsymbol{x}}^{\mathrm{f}}$ and $\bar{\mathbf{P}}^{\mathrm{f}}$ (as compared to $\boldsymbol{x}^{\mathrm{f}}$ and $\mathbf{P}^{\mathrm{f}}$). In the EnKF literature, this discrepancy is known as sampling error.

The sampling error in the forecast moments communicates to the moments of the ensemble updated by the analysis, eqn. (2.16). Moreover, in contrast to consistency, expectation (and hence unbiasedness) does not commute with nonlinear operations. In other words, individually inserting expected values into a formula, such as eqn. (2.17), does not yield the expected value of the formula. The nonlinearity in $\bar{\mathbf{P}}^{\mathrm{f}}$ of eqn. (2.26) therefore induces Proposition 2.3. Note that it does not contradict Propositions 2.1 and 2.2: the expectation referred to by Propositions 2.1 and 2.2 is only with respect to $\mathbf{D}$, while Proposition 2.3 also averages over the forecast ensemble.

**Proposition 2.3** – Bias of the EnKF update**.**
*Even if the forecast moments are unbiased, i.e.* $\mathbb{E}(\bar{\boldsymbol{x}}^{\mathrm{f}}) = \boldsymbol{x}^{\mathrm{f}}$ *and* $\mathbb{E}(\bar{\mathbf{P}}^{\mathrm{f}}) = \mathbf{P}^{\mathrm{f}}$*, the same does not hold for the analysis estimates,* $\bar{\boldsymbol{x}}^{\mathrm{a}}$ *and* $\bar{\mathbf{P}}^{\mathrm{a}}$*. More specifically,*

$$\mathbb{E}(\mathsf{tr}(\bar{\mathbf{P}}^{\mathrm{a}})) < \mathsf{tr}(\mathbf{P}^{\mathrm{a}})\,. \tag{2.37}$$

*Proof.* The proof is adapted from Furrer and Bengtsson [44]. It only considers a simplified, scalar case. More details on and additional results are provided therein and by [107] and [124].

Suppose $\mathbf{H} = 1$, $\mathbf{R} = 1$, and assume that $\mathbf{P}^{\mathrm{f}}$ is known. Let $g(u) = u/(1+u)$, so that the true and estimated Kalman gains, eqns. (2.7) and (2.17), are $\mathbf{K} = g(\mathbf{P}^{\mathrm{f}})$ and $\bar{\mathbf{K}} = g(\bar{\mathbf{P}}^{\mathrm{f}})$. Then, by Jensen's inequality [73, Th. 7.5], $\mathbb{E}(\bar{\mathbf{K}}) < \mathbf{K}$, since $g$ is a concave function and since $\mathbb{E}(\bar{\mathbf{P}}^{\mathrm{f}}) =$

$\mathbf{P}^{\mathrm{f}}$. Furthermore, $\mathbf{P}^{\mathrm{a}} = \mathbf{K}$ (for these particular values of $\mathbf{R}$ and $\mathbf{H}$) and, as can be shown from eqn. (2.31), $\mathbb{E}(\bar{\mathbf{P}}^{\mathrm{a}}) = \mathbb{E}(\bar{\mathbf{K}})$, where the expectation is *also* over $\mathbf{D}$. Hence $\mathbb{E}(\bar{\mathbf{P}}^{\mathrm{a}}) = \mathbb{E}(\bar{\mathbf{K}}) < \mathbf{K} = \mathbf{P}^{\mathrm{a}}$. □

Proposition 2.3 means that the sampling error in the forecast moments yields a systematic underestimation of $\bar{\mathbf{P}}^{\mathrm{a}}$. This may be compensated for by multiplicative "inflation", either manually tuned, or estimated on-line [20]. Another possibility is to split the ensemble into smaller sub-ensembles [61].

## 2.5 The effects of a small ensemble size

Rank deficiency refers to the fact that $N$ might be a lot smaller than $M$, and that therefore the low-order "parameterization" of the pdf by the ensemble cannot represent the *directions* of uncertainty of the distribution (as measured by the second order moment, the covariance matrix), nor its growth through the models. Conceptually, at least, this should be distinguished from sampling error, which affects the *accuracy* of the pdf representations and the model linearizations. Whether due to rank deficiency or sampling error, the smaller the ensemble size, $N$, the worse the ensemble estimates. However, even for large geoscientific models, experiments indicate that an ensemble size in the range of 10 to 100 is often sufficient [2, 7, 37, 62, 89, 132]. How can this be? The following subsections provide a semi-heuristic explanation.

### 2.5.1 Rank deficiency

With respect to the forecast step of the EnKF, it may be explained by observing that the dynamics typically only have a relatively small number of fast-growing "modes" [97, 125]. The growth in uncertainty can therefore be emulated by the growth in spread of the ensemble, if this is big enough [98, 101]. The number of growing modes of the dynamics can be estimated by the number of positive Lyapunov exponents [70], which is therefore sometimes used to estimate the required size of the ensemble [51, 63, 99].

Note that the EnKF analysis step, eqn. (2.16), does not explicitly change the subspace (i.e. directions) spanned by the ensemble. Instead, it is the propagation through the dynamical forecast model that is trusted with tracking the growing modes. As such, it is hoped that the ensemble constitutes an adaptive set of vectors that span the subspace of uncertainty (outside of which the uncertainty is close to zero), and that its left singular vectors form a "sparse" basis thereof [26].

This ability of the ensemble has enabled recommendations [29, 64] not to overly reduce the parameterization of the models.[2] This not only reduces the required preparation for DA, but may also reduce the propensity of the ensemble analysis update to produce dynamical inconsistencies [97].

With respect to the analysis step, it is helpful to think of the EnKF as trading bias for variance [119]. More specifically, the EnKF analysis step builds on the KF, which is a linear estimator. Although the inflexibility due to the linearity of the KF estimator yields a bias in the non-Gaussian setting, it also protects against large, random variations [16, §3.2]. The connection to rank deficiency is that the parameterization of the $k$-th order moment requires on the order of $M^k$ degrees of freedom. So although the ensemble may not be rank deficient compared to the covariance, it still is with respect to the third or forth order moments. In this view, the EnKF's "refusal to acknowledge" its rank deficiency (by neglecting higher order moments) is what guards it from the "curse of dimensionality" that so afflicts the particle filter [125].

[2] Reduced parameterizations restricts the DA process to a select subset of the state variables and parameters. A central concern is which basis is more adept for this selection. Sometimes simplified dynamics are also used [121]. Reduced parameterizations is also sometimes used for the EnKF [40].

### 2.5.2 Spurious correlations and localization

The limited ensemble size also causes problems through an effect known as spurious correlations. Recall that the ensemble is subject to stochastic variation due to the initial sampling, and the sampling of $\mathbf{D}_{\text{obs}}$ and $\mathbf{D}_{\text{mod}}$. Model nonlinearity may also be included in this reckoning [20]. This means that the forecast moments are subject to stochastic error, or sampling error, as defined in section 2.4.3. The errors *off* the diagonal of $\bar{\mathbf{P}}^{\text{f}}$ are known as spurious correlations.

To illustrate their effect, suppose that, based on physical arguments, one knows that two state variables that are far apart[3] have zero correlation. However, the variance of the estimate of this correlation is of the order of $1/N$, i.e. not zero. This, in turn, reduces the size of $\bar{\mathbf{P}}^{\text{a}}$, as summarized by Proposition 2.3 (which also concerns the effects of errors *on* the diagonal of $\bar{\mathbf{P}}^{\text{f}}$). It was already mentioned that inflation may be used to compensate for the systematic underestimation of $\bar{\mathbf{P}}^{\text{a}}$. By contrast, "localization" attempts to prevent spurious correlations.

There are two types of localization: local analysis and covariance tapering. Both techniques are outlined below, for completeness, but neither one is employed in any of the theory or experiments of this text.

#### 2.5.2.1 Local analysis

The local analysis approach is to perform the EnKF analysis update separately for different regions of a distance-based partition of the state vector, omitting observations that are beyond the area of influence of the region under consideration. The omission can be effectuated smoothly by multiplying the observation error variance (i.e. $\mathbf{R}$) of distant locations by factors larger than 1. Spurious correlations are diminished by omitting observation subsets that are known to carry little information on the region under consideration.

Furthermore, unlike the global update, a local analysis allows the analysis ensemble to consist of different linear combinations of the ensemble members in different regions. Hence the composite global analysis is not confined to the $N$-dimensional ensemble subspace, but is free to explore a much higher-dimensional space [63]. Thus, local analysis also constitutes a remedy to the issue of rank deficiency.

#### 2.5.2.2 Covariance tapering

The covariance tapering localization technique consists of the Schur (element-wise) product $\bar{\mathbf{P}}^{\text{f}} \mapsto \rho \circ \bar{\mathbf{P}}^{\text{f}}$ where $\rho$ is a distance-based $M \times M$ correlation matrix. The effect of applying $\rho$ is to taper the off-diagonal elements of $\bar{\mathbf{P}}^{\text{f}}$, thus making it more banded, and increasing its rank.

Sakov and Bertino [108] showed that the effects of the two localization approaches are highly similar. However, covariance tapering has the theoretical benefit that it is similar to Tikhonov regularization [55], which in this case would consist in *adding* a banded matrix to $\bar{\mathbf{P}}^{\text{f}}$. Since Tikhonov regularization can be seen as a way of including prior information [41], the link to Tikhonov regularization provides some formalism for localization [20], whose justification, as given above, is otherwise rather ad hoc.

[3] Distance is the simplest example, but more advanced considerations may also be used.

## 2.6 Summary and discussion

The principal reason for employing the EnKF as an approximate filtering algorithm is its capacity to deal with nonlinearity despite its low-order ensemble cloud representation of the pdfs [19]. For example, in contrast with the extended KF, the EnKF does not require pre-derived linearizations of $\mathcal{M}$ and $\mathcal{H}$. Furthermore, if the state vector length, $M$, is on the order of $10^9$, as it may be in DA (section 1.1), then working with the state covariance matrix requires at least 8 terabytes of memory. This is currently infeasible, ruling out naive application of the extended KF. On the other hand, as seen from eqns. (2.13), (2.16) and (2.18), the EnKF does not explicitly compute $\bar{\mathbf{P}}^{\mathrm{f/a}}$, but works with matrices of size $M \times N$, $P \times N$, and $P \times P$. The latter can be exchanged for a matrix of size $N \times N$, as shown in chapter 4.

Another advantage of the EnKF is that it is relatively easy to understand and implement, depending only on matrix algebra libraries. Moreover, in contrast to the extended KF and variational methods such as 4D-Var,[4] the EnKF is non-invasive: the forecast and observation models, $\mathcal{M}$ and $\mathcal{H}$, are operated as black boxes, because the ensemble provides approximations to their linearized sensitivities (section 2.3.2). Furthermore, the EnKF is of a Bayesian nature: the background covariance estimate is provided by the forecast and is thus state-dependent, and the ensemble provides multiple possible realizations rather than a single estimate. Nevertheless, in some applications, 4D-Var is still preferred to the EnKF, mainly because its iterative smoothing formulation makes it more accurate for forecast initialization. However, this formulation can also be adapted by the EnKF [18, 112], and significant efforts are currently focused on merging the variational and the ensemble approaches, giving rise to methods such as 4D-EnVar, En-4D-Var [78, 129], and the ensemble "randomized maximum likelihood" method [29].

Another advantage of the EnKF is that it is extensively, and trivially, parallelizable. The model integration for the forecast step can be carried out on individual computers for each ensemble member, with no intercommunication required before the analysis step. If local analysis localization is used, then the analysis can also be distributed to individual computers, each one performing the update for its subset/region of the state vector.

In summary, while it is equivalent to the KF [36] and thus also 4D-Var under idealistic assumptions [19, 42], the main advantages of the EnKF in practice are that it (i) has a natural Bayesian interpretation; (ii) uses state-dependent error estimates in its observation analysis; (iii) does not require an adjoint tangent linear model of the dynamics; (iv) requires moderate working memory storage; and (v) is highly parallelizable.

Two algorithms closely related to the EnKF are the Unscented KF [68] and the reduced-rank square root filter (RRSQRT) [56]. Both of these can be viewed as versions of the EnKF where the ensemble members are re-initialized after each analysis so as to parameterize the filtered pdf according to particular criteria. This is in contrast with the analysis update of the EnKF, where each updated ensemble member is a linear combination of the forecasted ensemble members, which is a relatively benign updating mechanism [103, §5.3] compared to the re-initializations of the Unscented KF and the RRSQRT.

Attempts have been made to construct EnKF-derived filters that converge to the exact distributions also in the non-Gaussian case. One such method is the rank histogram filter [6]. It is only rigorous in one-dimensional DA problems, but attempts have been made to generalize it to higher dimensions [88]. Another possible avenue is to merge the EnKF with the particle filter in some way. For example, the moment-matching EnKF replaces the analysis ensemble mean with that from the particle filter [74, 136]; higher-order corrections are also possible; performance improvements have been shown for large $N$ [104]. Other attempts include particle filters which use the EnKF as a proposal distribution [93, 125], as well as Gaussian mixture filters [119].

[4] Very briefly, 4D-Var [76, 77] can be described as the method of iteratively optimizing (for a single point) $p(\boldsymbol{x}_{t-L} \mid \boldsymbol{y}_{t-L:t})$, for some lag length $L$, using the full nonlinear models, $\mathcal{M}$ and $\mathcal{H}$, and their gradients. The prior, $p(\boldsymbol{x}_{t-L})$, is typically a climatological average.

# Chapter 3

# Numerical twin experiments

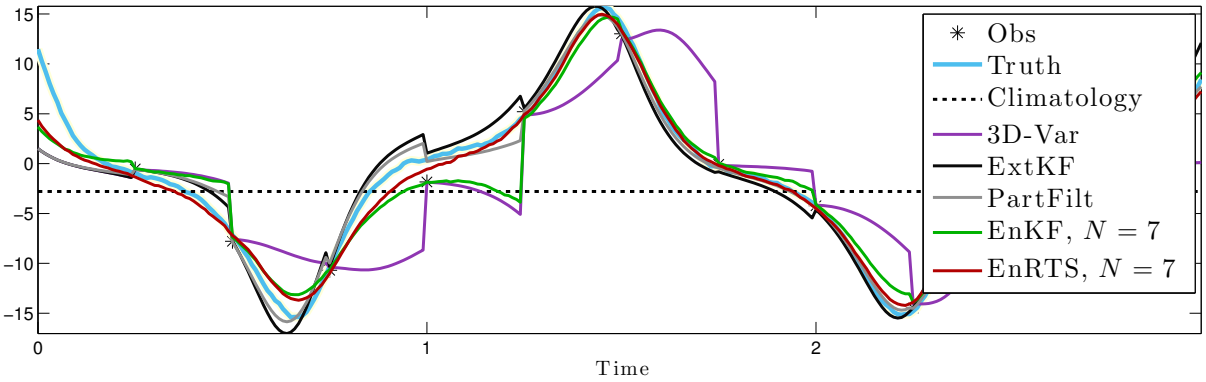



**Figure 3.1:** Illustration of a twin experiment, obtained with the Lorenz-63 system. Each line represents the method's mean estimate of the "$x$" dimension. Also included is the simulated, synthetic truth, and the noisy observations of it.

DA methods may be benchmarked using "twin experiments": an artificial "truth" is simulated, and its trajectory is subsequently estimated by the DA methods. Figure 3.1 gives an illustration. Although tracking the truth is not the formal objective of DA, which is to compute the Bayesian posterior distribution of the truth, the idea behind such twin experiments is intuitive. Furthermore, it is possible that the two objectives can be theoretically related by the connection between the "log score" and the average Kullback-Leibler divergence [130]. More details are given in section 3.2.

In this text, a linear advection model, the Lorenz-63 model, and the Lorenz-96 model, described in sections 3.3 to 3.5, are used to test the performance of the different methods.

In addition to the ensemble methods, the following DA methods are sometimes provided as baselines:

- The extended Kalman Filter (ExtKF): described in section 2.1; for linear systems, it provides the exact posterior.
- The particle filter (PartFilt): see [118, 125]; only applied for the Lorenz-63 model, where it is used with a sufficiently large number of particles ($N = 10^4$) to practically converge to the exact posterior.
- Climatology: the stationary Gaussian distribution whose mean and covariance are the averages of the dynamics over a long period of time.
- 3D-Var: like the extended KF, except that the forecast prior (i.e. $\boldsymbol{x}^{\mathrm{f}}$ and $\mathbf{P}^{\mathrm{f}}$) is taken to be the climatology.

However, these baselines will not draw too much of our attention; their purpose is to help identify the experiments that are relevant. For example, it is not very interesting to evaluate improvements to the EnKF under conditions (i.e. small $N$) where the EnKF is largely outperformed by 3D-Var. The baselines also help in appraising the significance of the differences between different versions of the EnKF.

## 3.1 RMSE averages

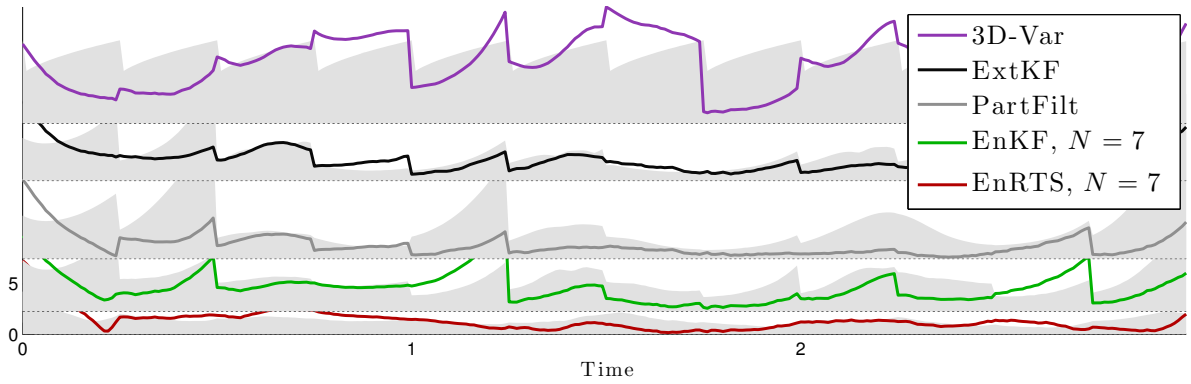



**Figure 3.2:** Illustration of the RMSE in a twin experiment. The vertical scale is indicated for the bottom plot only, due to the axes overlap. The shaded areas show the "root-mean variance" of the corresponding method, such as $\mathsf{tr}(\bar{\mathbf{P}}_t)/M$ for the ensemble methods.

The performance of the ensemble methods is measured by their accuracy, as quantified by the root-mean-square error (RMSE) metric, between the ensemble mean and truth. For a given experiment with truth trajectory $\{\boldsymbol{x}_t \ ; \ t = 0, 1, \ldots\}$, at a specific time, $t$, the RMSE of the ensemble mean, $\bar{\boldsymbol{x}}_t$, is

$$\text{RMSE} = \sqrt{\frac{1}{M} \|\bar{\boldsymbol{x}}_t - \boldsymbol{x}_t\|_2^2} \,. \tag{3.1}$$

Figure 3.2 shows the RMSE values and the root mean variance values corresponding to Fig. 3.1. Plots such as Figs. 3.1 and 3.2 can be invaluable as diagnostic tools, for debugging, and to gain understanding. However, time series plots of individual or even domain-averaged variables, statistics, or ensemble trajectories should be interpreted with caution, as they are always subject to chance occurrences. The same can be said for instantaneous snapshots of the state amplitudes/fields or scatter plots of ensembles.

The RMSE is therefore also averaged across time, with the exclusion of an initial transitory period whose duration is established beforehand by studying the time series. By convention, the RMSE is measured (only) immediately following each analysis update. This choice (as opposed to measuring *before* analysis, for example) was generally found to have little impact on the relative scores of the different EnKF methods tested. Each experiment is repeated 16 times with different initial random seeds. The empirical variance of the experiment RMSEs is computed and checked to ensure satisfying convergence.

By comparing with two of the baselines mentioned on page 16, the RMSE averages may be normalized to "skill scores" between 0 and 1 [134, §7.1.4]. We have chosen to not do so here so that the absolute RMSE values can serve as some measure of the difficulty of the DA problem of a given figure.

## 3.2 Considerations on the metric

The RMSE only directly evaluates the performance of the ensemble mean, $\bar{\boldsymbol{x}}$. However, the ensemble approach produces probabilistic forecast distributions, and not just point estimates. Therefore, it would be more appropriate [67] to use probabilistic scoring rules, provided they are proper.[1] One example is the log score, which evaluates the logarithm of the estimated pdf at the point of the truth. However, the computation of probabilistic scores for ensembles typically requires some form of density estimation, which can be complicated in high dimensions. Possible approximations include marginal totals [119] and low-order moment truncations [33].

Another set of metrics that are more suitable for probabilistic forecasts are the distance measures for distributions, including the Kolmogorov-Smirnov distance [92], and the Kullback-Leibler divergence [32]. However, these metrics require the actual posterior for reference, which is typically too costly to compute. Some medium-scale examples exist [34, 92], but are burdened by the question of their serendipity. Moreover, this metric requires density estimation across the entire state domain. Therefore, attempts have been made to derive cloud-to-cloud formulations, but this introduces additional variance [127].

By comparison to scoring rules and distribution distances, the RMSE score is very simple. Yet it can be argued that the average RMSE score should be quite reliable in DA for assessing the skill of an ensemble system in a twin experiment. This is because, by the sequentiality of DA, it indirectly assesses other aspects than just the ensemble mean. For example, while the spread of the ensemble does not directly impact the concurrent RMSE score, it will impact how the observation is weighted vis-a-vis the prior at the *next* analysis step, and hence the next RMSE score. Additionally, an ensemble that poorly represents the actual distribution will undergo different, nonlinear propagation than an accurately representative ensemble, hence also impacting the mean and the RMSE. Lastly, the RMSE is a highly standard metric in the EnKF literature, and its use therefore facilitates comparison and reproducibility.

The following sections describe the models used in the experiments.

[1] A "proper" scoring rule cannot be "gamed": a forecaster always best served by listing his or her actual beliefs [46], rather than erring on the side of caution, optimism, or in some other way. Additionally, all proper scores can be understood as a sum of measures of uncertainty, reliability, and resolution [24]. Famous examples include the Brier score [23] and the log score [15]. Beyond propriety, Bickel [15] showed that the log score has some useful advantages compared to other proper scores. Shuford Jr. et al. [117] showed that it is the only "local" proper score. Benedetti [12] recognizes this as a fundamental scientific "likelihood" principle, and shows that the log score can be derived from this result. Benedetti [12], Weijs et al. [130] show that the Brier score is the second-order approximation to the log score.

## 3.3 Linear advection

Upon discretization with the first-order upwind scheme, the nondimensionalized, one-dimensional advection equation yields

$$[\boldsymbol{x}_{t+1}]_i = \Delta t\,[\boldsymbol{x}_t]_{i-1} + (1 - \Delta t\,)[\boldsymbol{x}_t]_i\,. \tag{3.2}$$

The customary [e.g. 38] model time step is used, namely $\Delta t = 1$, coinciding with the CFL limit [e.g. 122, §4]. Additionally, to counteract the growth due to the additive model noise, the state is multiplied by a dissipation factor of 0.98 after each time step. Equation (3.2) therefore becomes

$$[\boldsymbol{x}_{t+1}]_i = 0.98[\boldsymbol{x}_t]_{i-1}\,, \tag{3.3}$$

which is run for $t = 0, \ldots, 2000$, $i \in 1{:}m$, with $M = 1000$, and periodic boundary conditions. It is illustrated in Fig. 3.3. Direct observations of the truth are taken at $P = 40$ equidistant locations, with $\mathbf{R} = 0.01\mathbf{I}_p$, every fifth time step.

Similarly to [110], the initial ensemble members, $\{\boldsymbol{x}_{0,n}\ ;\ n \in 1{:}N\}$, as well as the truth, $\boldsymbol{x}_0$, are generated as a sum of 25 sinusoids of random amplitude and phase,

$$[\boldsymbol{x}_{0,n}]_i = \frac{1}{c_n}\sum_{k=1}^{25} a_n^k \sin\left(2\pi k\left(i/M + \varphi_n^k\right)\right)\,, \tag{3.4}$$

where $a_n^k$ and $\varphi_n^k$ are drawn independently and uniformly from the interval $(0, 1)$ for each $n$ and $k$, and the normalization constant, $c_n$, is such that the standard deviation of each $\boldsymbol{x}_{0,n}$ is 1. Note that the (spatial) mean of each realization of eqn. (3.4) is zero.

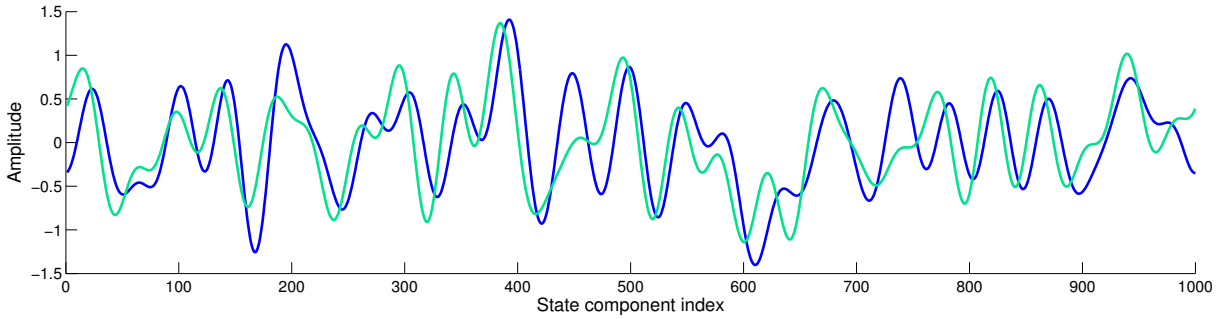



**Figure 3.3:** Snapshots of amplitudes during a free run of the linear advection system of eqns. (3.3) and (3.4). The first snapshot (turquoise) is taken at $t = 140$, The last (blue) snapshot is taken 10 time steps later, at $t = 150$. Because of model noise, the advection is not a pure translation.

## 3.4 Lorenz-63

The Lorenz-63 model [81, 113], is a three-dimensional model that has become ubiquitous in the field of nonlinear dynamics. Its equations,

$$\dot{x} = \sigma(y - x)\,, \tag{3.5}$$
$$\dot{y} = rx - y - xz\,, \tag{3.6}$$
$$\dot{z} = xy - bz\,, \tag{3.7}$$

are a leading-order approximation to Rayleigh-Bénard flow [57, §C], where an incompressible fluid of Rayleigh number $r$ and Prandtl number $\sigma$ circulates in a box of aspect ratio $b$ due to a temperature gradient. The variables represent flow strength, $x$; temperature perturbation strength, $y$; and deviation from a nonlinear temperature profile, $z$. We use the common parameter settings of $r = 28$, $\sigma = 10$, and $b = 8/3$, in which case the system is nonlinear with a maximum Lyapunov exponent of 0.9 [e.g. 27]. In this case there are no stable fixed points or limit cycles; instead the system has a "strange attractor" and the solutions never repeat, but bear resemblance to a butterfly when plotted over time, as illustrated in Fig. 3.4.

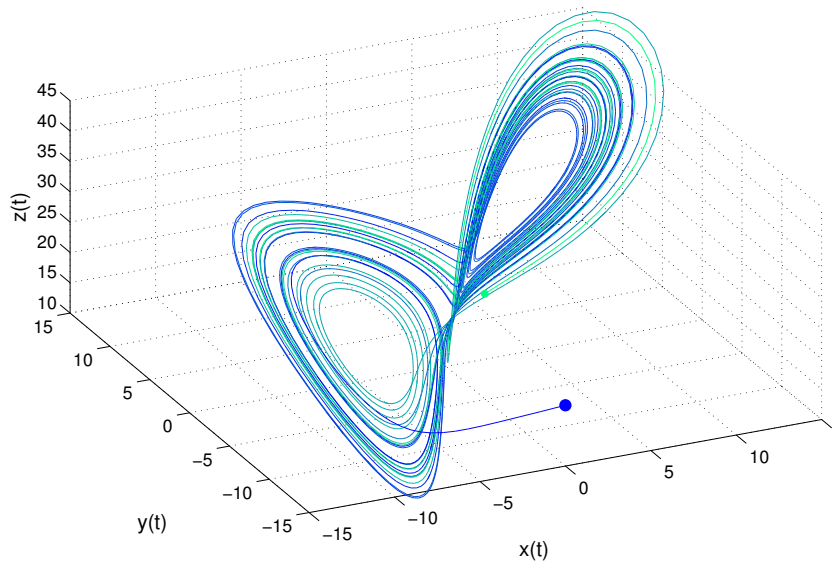



**Figure 3.4:** Trajectory of a free run of the Lorenz-63 system (3.5) in phase space. The trajectory traces a run from $t = 0$ to $t = 40$, with $\boldsymbol{x}_0 = [0, -15, 20]^\mathsf{T}$.

Equation (3.5) is integrated using the fourth-order Runge-Kutta numerical time stepping scheme with a time step of $\Delta t = 0.01$ for $T = 50\,000$ time steps. Unless otherwise stated, direct observations of the entire state vector are taken $\Delta t_{\text{obs}} = 0.25$ apart, with error covariance $\mathbf{R} = 2\mathbf{I}_3$.

## 3.5 Lorenz-96

The Lorenz-96 model [82], given by the coupled set of ordinary differential equations,

$$\frac{d[\boldsymbol{x}]_i}{dt} = ([\boldsymbol{x}]_{i+1} - [\boldsymbol{x}]_{i-2})\,[\boldsymbol{x}]_{i-1} - [\boldsymbol{x}]_i + F\,, \tag{3.8}$$

applied for $t > 0$, and $i \in 1{:}m$, with $M = 40$ and periodic boundary conditions. It is a nonlinear, chaotic model that mimics the atmosphere at a certain latitude circle. Unlike its spiritual predecessor, the Lorenz-63 model, it was not derived by truncating a more comprehensive set of meteorological equations. It was *designed* [83] as a simplistic, symmetric system where the variables can be thought of as a some equidistant discretization of a scalar, meteorological quantity such as temperature or vorticity, and (i) the nonlinear terms, intended to simulate advection, are quadratic and together conserve $\|\boldsymbol{x}\|_2^2$, the total energy; (ii) the linear terms, representing mechanical or thermal dissipation, decrease the total energy; and (iii) the constant terms, representing external forcing, prevent the total energy from decaying to zero.

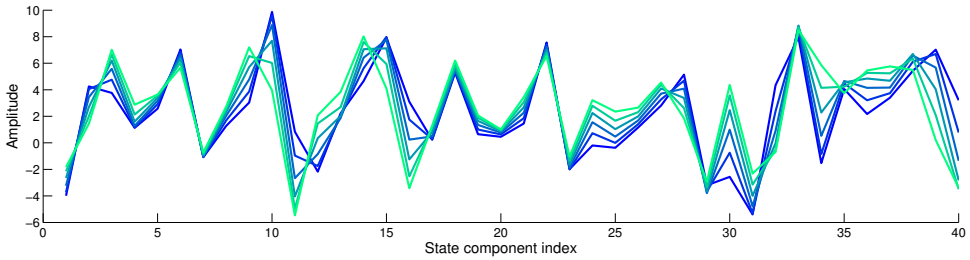



**Figure 3.5:** Snapshots of amplitudes during a free run of the Lorenz-96 system (3.8). Although the initial sample is sampled from $\mathcal{N}(0, \mathbf{I}_{40})$, the first snapshot (dark blue) is taken at $t = 10$, so that the transient effects have had time to expire. The last (turquoise) snapshot is at $t = 10.20$, corresponding to the analogous atmospheric time span of 1 day. Although wavenumber 8 appears prominently in the spectrum, the individual "highs" and "lows" alter their shapes and intensities rather irregularly, as they progress slowly and not invariably westward (towards lower dimension indices).

With $M = 40$, the steady-state solution, $[\boldsymbol{x}]_1 = \ldots = [\boldsymbol{x}]_m = F$, is stable for $F < 0.895$ [84]. For $0.895 < F < 4$ the system instead gradually settles into a wavenumber 8 periodic solution travelling in the direction of decreasing $i$. For $F > 4$ the system is chaotic, although wavenumber 8 still dominates its spectrum (Fig. 3.5). In this case the time-average mean is 2.3 and the standard deviation 3.6 (for all of the state variables). Furthermore, the system has 13 positive Lyapunov exponents, which Bocquet and Sakov [18] connected to the minimum useful ensemble size for the EnKF without localization, and the Kaplan-Yorke dimension of its fractal attractor is 27.05. The leading Lyapunov exponent corresponds to a doubling time of 0.42 time units; equating this to the doubling time of global atmospheric circulation models, estimated at 2.1 days by Lorenz [80], a unit of non-dimensional time can be identified with 5 days. Using this identification means that the leading exponent is 0.336 $(\text{days})^{-1}$ [28].

Equation (3.8) is integrated using the fourth-order Runge-Kutta scheme with a time step of $\Delta t = 0.05$ for $T = 10\,000$ time steps. Unless otherwise stated, direct observations of the entire state vector are taken $\Delta t_{\text{obs}} = 0.05$ apart (representative of 6 hours) with error covariance $\mathbf{R} = \mathbf{I}_{40}$.

# Chapter 4

# The square root analysis update: The ETKF

Square root KFs [102] primarily improve on the numerical stability of the KF [69]. They inherently preserve the symmetry and non-negativity of the state covariance matrix, and the condition number of the matrices involved is approximately the square root of that of the standard KF [e.g. 3, 75].

Introduced to the analysis step of the EnKF by references [4, 17, 131], the square root analysis update was soon recognized and connected to the original square root KFs [120]. However, for the EnKF, the main purpose is not numerical stability, but rather to do away with the stochasticity and the accompanying sampling errors (non-conformality) of the perturbed-observation analysis update of the original EnKF.

Recall, in particular, that $\mathbf{D}_{\text{obs}} \in \mathbb{R}^{P\times N}$ is a matrix whose columns are drawn independently from $\mathcal{N}(0, \mathbf{R})$. Unfortunately, as was shown in section 2.4.1, the perturbed-observation analysis update of section 2.3.2,

$$\mathbf{E}^{\text{a}} = \mathbf{E}^{\text{f}} + \bar{\mathbf{K}}\left\{\boldsymbol{y}\mathbb{1}^{\mathsf{T}} - \mathbf{D}_{\text{obs}} - \mathcal{H}(\mathbf{E}^{\text{f}})\right\}, \tag{4.1}$$

only yields the intended, conformal covariance, eqn. (2.31), on average:[1][2]

$$\mathbb{E}(\bar{\mathbf{P}}^{\text{a}}) = [\mathbf{I}_m - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{P}}^{\text{f}}. \tag{4.2}$$

## 4.1 Method

By contrast, the square root analysis update, described immediately below, satisfies eqn. (2.31) exactly. This is possible because $\mathbf{X}^{\text{f}}$ factorizes out from the right hand side of eqn. (2.31). Define the notation $\overline{\mathcal{H}(\mathbf{E}^{\text{f}})} = \frac{1}{N}\mathcal{H}(\mathbf{E}^{\text{f}})\mathbb{1}$, and let

$$\bar{\boldsymbol{\delta}} = \boldsymbol{y} - \overline{\mathcal{H}(\mathbf{E}^{\text{f}})} \qquad \in \mathbb{R}^{P}, \tag{4.3}$$

$$\mathbf{Y} = \mathcal{H}(\mathbf{E}^{\text{f}})(\mathbf{I}_N - \mathbb{1}\mathbb{1}^{\mathsf{T}}/N) \in \mathbb{R}^{P\times N}, \tag{4.4}$$

[1] In accordance with Propositions 2.1 and 2.2, the expectation, $\mathbb{E}$, is taken with respect to $\mathbf{D}_{\text{obs}}$, but not with respect to the forecast ensemble, which is considered fixed in this chapter.

[2] Nonlinearity is a separate issue, and not the topic of this chapter. If $\mathcal{H}$ is nonlinear, this discussion and eqn. (4.2) should be interpreted either with $\mathbf{H}$ as in the state-observation augmentation trick, or by replacing $\mathbf{H}\bar{\mathbf{P}}^{\text{f}}$ by $\frac{1}{N-1}\mathbf{Y}\mathbf{X}^{\text{f}\mathsf{T}}$, as outlined in section 2.3.2.

be the mean "innovation" and the anomalies of the *observed* ensemble, respectively. Using eqn. (2.11), it may be shown that eqns. (2.20) and (2.31) are satisfied if

$$\bar{\boldsymbol{x}}^{\mathrm{a}} = \bar{\boldsymbol{x}}^{\mathrm{f}} + \mathbf{X}^{\mathrm{f}}\mathbf{Y}^{\mathsf{T}}\left(\mathbf{Y}\mathbf{Y}^{\mathsf{T}} + (N{-}1)\mathbf{R}\right)^{-1}\bar{\boldsymbol{\delta}}\,, \tag{4.5}$$

$$\mathbf{X}^{\mathrm{a}}\mathbf{X}^{\mathrm{a}\mathsf{T}} = \mathbf{X}^{\mathrm{f}}\mathbf{G}^{\mathrm{a}}\mathbf{X}^{\mathrm{f}\mathsf{T}}\,, \tag{4.6}$$

$$\text{where} \quad \mathbf{G}^{\mathrm{a}} = \mathbf{I}_N - \mathbf{Y}^{\mathsf{T}}(\mathbf{Y}\mathbf{Y}^{\mathsf{T}} + (N{-}1)\mathbf{R})^{-1}\mathbf{Y}\,. \tag{4.7}$$

Corollaries A.1 and A.2 may be used to rewrite eqns. (4.5) and (4.7) as

$$\mathbf{G}^{\mathrm{a}} = \left(\tfrac{1}{N-1}\mathbf{Y}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{Y} + \mathbf{I}_N\right)^{-1}\,, \tag{4.8}$$

$$\bar{\boldsymbol{x}}^{\mathrm{a}} = \bar{\boldsymbol{x}}^{\mathrm{f}} + \tfrac{1}{N-1}\mathbf{X}^{\mathrm{f}}\mathbf{G}^{\mathrm{a}}\mathbf{Y}^{\mathsf{T}}\mathbf{R}^{-1}\bar{\boldsymbol{\delta}}\,, \tag{4.9}$$

the calculation of which is computationally cheaper if $N < P$ and $\mathbf{R}$ is readily inverted.

Therefore, if $\mathbf{X}^{\mathrm{a}}$ is computed by

$$\mathbf{X}^{\mathrm{a}} = \mathbf{X}^{\mathrm{f}}\mathbf{T}^{\mathrm{a}}\,, \tag{4.10}$$

with $\mathbf{T}^{\mathrm{a}}$ being a matrix square root of $\mathbf{G}^{\mathrm{a}}$, then $\mathbf{X}^{\mathrm{a}}$ satisfies eqn. (2.31) exactly. Note, however, as indicated by eqn. (4.2), that the square root update only addresses the issue of sampling error in $\mathbf{D}_{\mathrm{obs}}$. That is, the EnKF with the square root analysis method is still afflicted by the issue of Proposition 2.3.

The ensemble is obtained by recombining the anomalies and the mean:

$$\mathbf{E}^{\mathrm{a}} = \bar{\boldsymbol{x}}^{\mathrm{a}}\mathbb{1}^{\mathsf{T}} + \mathbf{X}^{\mathrm{a}}\,. \tag{4.11}$$

The term "square root update" is henceforth used to refer to any update of the anomalies through the right-multiplication of a transform matrix, as in eqn. (4.10).

## 4.2 The symmetric square root

Equation (4.8) implies that $\mathbf{G}^{\mathrm{a}}$ is symmetric, positive-definite (SPD). The matrix $\mathbf{T}^{\mathrm{a}}$ is a square root of $\mathbf{G}^{\mathrm{a}}$ if it satisfies

$$\mathbf{G}^{\mathrm{a}} = \mathbf{T}^{\mathrm{a}}\mathbf{T}^{\mathrm{a}\mathsf{T}}\,. \tag{4.12}$$

However, by substitution into eqn. (4.12) it is clear that $\mathbf{T}^{\mathrm{a}}\boldsymbol{\Omega}$ is also a square root of $\mathbf{G}^{\mathrm{a}}$, for any orthogonal matrix $\boldsymbol{\Omega}$. There are therefore infinitely many square roots. Nevertheless, some have properties that make them unique. For example, the Cholesky factor is unique as the only triangular square root with positive diagonal entries.

Here, however, the square root of most interest is the symmetric one, $\mathbf{T}^{\mathrm{a}}_{\mathrm{sym}} = \mathbf{V}\boldsymbol{\Sigma}^{1/2}\mathbf{V}^{\mathsf{T}}$, where $\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}} = \mathbf{G}^{\mathrm{a}}$ is the eigendecomposition of $\mathbf{G}^{\mathrm{a}}$, and $\boldsymbol{\Sigma}^{1/2}$ is defined as the entry-wise *positive* square root of $\boldsymbol{\Sigma}$ [60, Th. 7.2.6]. Its existence follows from the spectral theorem, and its uniqueness from that of the eigendecomposition. Note its distinction by the "sym" subscript.

It was gradually discovered that the symmetric square root choice has several advantageous properties for its use in eqn. (4.10), one of which is that it does not affect the ensemble mean [e.g. 36, 128], which is updated by eqn. (4.5) apart from the anomalies. Further advantages are surveyed in Raanes [103, §5.3], providing strong justification for choosing the symmetric square root, and strong motivation to extend the square root approach to the forecast step.

## 4.3 Efficient computation

It is necessary to compute $\mathbf{G}^{\mathrm{a}}$ for the mean update, eqn. (4.9), and $\mathbf{T}^{\mathrm{a}}_{\mathrm{sym}}$, for the analysis update, eqn. (4.10). Fortunately, $\mathbf{T}^{\mathrm{a}}_{\mathrm{sym}}$ comes at practically no additional cost: with $\mathbf{V} \in \mathbb{R}^{N^2}$ as the *right* singular vectors of $((N-1)\mathbf{R})^{-1/2}\mathbf{Y}$, and $\mathbf{\Sigma} \in \mathbb{R}^{P\times N}$ as the diagonal matrix of singular values,

$$\mathbf{G}^{\mathrm{a}} = \mathbf{V}(\mathbf{I}_N + \mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma})^{-1}\mathbf{V}^{\mathsf{T}}\,, \tag{4.13}$$

$$\mathbf{T}^{\mathrm{a}}_{\mathrm{sym}} = \mathbf{V}(\mathbf{I}_N + \mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma})^{-1/2}\mathbf{V}^{\mathsf{T}}\,. \tag{4.14}$$

Evidently, $\mathbf{V}$ is also the matrix of eigenvectors of $\mathbf{G}^{\mathrm{a}}$ and $\mathbf{T}^{\mathrm{a}}_{\mathrm{sym}}$, while $(\mathbf{I}_N + \mathbf{\Sigma}^{\mathsf{T}}\mathbf{\Sigma})$ contains the eigenvalues. However, as long as only the reduced SVD (Definition B.2) is computed, it does not matter much which of these paths is used to obtain $\mathbf{V}$.

## 4.4 Random rotations

Although Sakov and Oke [110] found the symmetric square root to be superior to other square root choices, it was also noted that for sufficiently large ensemble sizes, $N$, the RMSE performance may actually deteriorate with increasing $N$. This counter-intuitive phenomenon was attributed to a tendency of the symmetric square root choice to build up "outlier" ensemble members. In response, it was found that it is sometimes beneficial to "scramble" the ensemble from time to time by drawing a random $\mathbf{\Omega}$ and right-multiplying the anomalies by it after the update of eqn. (4.10). As in section 4.2, $\mathbf{\Omega}$ should be orthogonal, and (in order to preserve the mean) should have $\mathbb{1}$ as an eigenvector. As noted by Pham [101], $\mathbf{\Omega}$ can be generated by sampling an $(N-1)\times(N-1)$ noise sample, where each is element is drawn independently from $\mathcal{N}(0,1)$, and then computing how the sample is rotated compared to some reference.[3]

The deterioration and its remedy are still somewhat mysterious, nevertheless, because it remains unexplained why the build-up of outliers occurs. A possible explanation is proposed in the following. As detailed in Raanes [103, §5.3.3], the symmetric choice has the property that it minimizes the "transport" of the ensemble through a square root update. This is not only beneficial for the dynamical consistency of the ensemble, but also for its statistical properties, because it preserves higher order moments of the prior ensemble beyond its covariance [36]. However, there is such a thing as too much preservation: after all, the likelihood also contains higher order moment information (typically they are all assumed zero, by Gaussianity). The outliers may be a reflection of this over-preservation.

[3] For example, the angles may be defined by how the singular vectors of the sample compare to the coordinate vectors [38]. However, it is less costly to use the orthogonal factor from the QR decomposition. Moreover, if such scrambling is used, then it is of no consequence which square root choice is used for $\mathbf{T}^{\mathrm{a}}$ [43, Lemma 4.5.5]. In terms of "flops" and standard matrix decomposition algorithms, it is then computationally cheaper [by a factor of 4, 123] to use the Cholesky factor rather than the symmetric square root.

# Appendix A

# The Kalman filter in detail

This appendix complements section 2.1 by filling in the details of the derivation of the KF, as well as providing better understanding of its significance.

## A.1 Matrix inversion identities

Matrix algebra is complicated by the fact that matrix multiplication is not commutative, and matrices are not generally invertible. It is therefore necessary to catalogue some identities.

**Lemma A.1** – Woodbury identity, also known as the matrix inversion lemma.**.**
*For any invertible* $\mathbf{B}$ *and* $\mathbf{R}$*, and any* $\mathbf{V}, \mathbf{U}$ *such that* $\mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U} + \mathbf{B}^{-1}$ *is defined,*

$$\left(\mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U} + \mathbf{B}^{-1}\right)^{-1} = \mathbf{B} - \mathbf{B}\mathbf{V}^\mathsf{T}\left(\mathbf{R} + \mathbf{U}\mathbf{B}\mathbf{V}^\mathsf{T}\right)^{-1}\mathbf{U}\mathbf{B}\,. \tag{A.1}$$

*Proof.* As a direct proof, pre-multiply the right hand side by the inverse of the left:

$$\begin{aligned}
&\left(\mathbf{B}^{-1} + \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}\right)\left[\mathbf{B} - \mathbf{B}\mathbf{V}^\mathsf{T}\left(\mathbf{R} + \mathbf{U}\mathbf{B}\mathbf{V}^\mathsf{T}\right)^{-1}\mathbf{U}\mathbf{B}\right] \\
&\quad = \mathbf{I}_m + \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}\mathbf{B} - (\mathbf{V}^\mathsf{T} + \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}\mathbf{B}\mathbf{V}^\mathsf{T})(\mathbf{R} + \mathbf{U}\mathbf{B}\mathbf{V}^\mathsf{T})^{-1}\mathbf{U}\mathbf{B} \\
&\quad = \mathbf{I}_m + \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}\mathbf{B} - \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}(\mathbf{R} + \mathbf{U}\mathbf{B}\mathbf{V}^\mathsf{T})(\mathbf{R} + \mathbf{U}\mathbf{B}\mathbf{V}^\mathsf{T})^{-1}\mathbf{U}\mathbf{B} \\
&\quad = \mathbf{I}_m + \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}\mathbf{B} - \mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}\mathbf{B} = \mathbf{I}_m
\end{aligned}$$

□

The significance of this identity is evident when $\mathbf{U}, \mathbf{V}$ are rectangular matrices (e.g., vectors) such that $\mathbf{V}^\mathsf{T}\mathbf{R}^{-1}\mathbf{U}$ constitutes a lower-rank "update" (addition) to $\mathbf{B}^{-1}$. The lemma then says that, assuming the inverse ($\mathbf{B}$) of $\mathbf{B}^{-1}$ is already known, computing the inverse of the updated matrix only requires an inversion of the size of $\mathbf{R}$. While the identity had been derived earlier through the inversion of a block matrix, this significance was only fully recognized when the identity was obtained as a generalization on rank-1 updated inverses [52]. Here, Lemma A.1 is used through Corollaries A.1 and A.2, both of which embody the useful change of dimensionality of the inversion.

**Corollary A.1.**
*For any symmetric, positive-definite* $\mathbf{R}$ *and* $\mathbf{B}$*, and* $\mathbf{U} = \mathbf{V} = \mathbf{H}$*,*

$$\left(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1} = \mathbf{B} - \mathbf{B}\mathbf{H}^\mathsf{T}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\right)^{-1}\mathbf{H}\mathbf{B} \tag{A.2}$$

*Proof.* Need to show the existence of the left hand side. For all $\boldsymbol{x} \in \mathbb{R}^M$, $\boldsymbol{x}^\mathsf{T}\mathbf{B}^{-1}\boldsymbol{x} > 0$. Thus $\boldsymbol{x}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}\boldsymbol{x} \geq 0$ implying that the left hand side is symmetric positive-definite: $\boldsymbol{x}^\mathsf{T}(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})\boldsymbol{x} > 0$, and hence invertible. □

**Corollary A.2.**
*For the same matrices as in Corollary A.1,*

$$\left(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}+\mathbf{B}^{-1}\right)^{-1}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}=\mathbf{B}\mathbf{H}^\mathsf{T}\left(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\right)^{-1}. \tag{A.3}$$

*Proof.* A straightforward validation of eqn. (A.3) is obtained by "cross-multiplying" out the inverses, but a more satisfying exercise is to derive it from eqn. (A.2), starting by right-multiplying by $\mathbf{H}^\mathsf{T}$. □

## A.2 Kalman filter derivation

Our derivation of the KF consists in applying the linearity assumptions of section 2.1 to the Bayesian filtering equations, (1.12) and (1.13). Focusing on one particular time index, $t$, assume that initially $\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1}\sim\mathcal{N}(\boldsymbol{x}^\mathrm{a},\mathbf{P}^\mathrm{a})$, where the moments, $\boldsymbol{x}^\mathrm{a}$ and $\mathbf{P}^\mathrm{a}$ are known.

### A.2.1 Forecast step

The linear forecast process, $\boldsymbol{x}_t=\mathbf{F}\boldsymbol{x}_{t-1}+\boldsymbol{q}_{t-1}$, amounts to a summation of Gaussian random variables. It therefore maintains the Gaussianity, and we can write

$$\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}\sim\mathcal{N}(\boldsymbol{x}^\mathrm{f},\mathbf{P}^\mathrm{f})\,. \tag{A.4}$$

where the forecast moments, $\boldsymbol{x}^\mathrm{f}$ and $\mathbf{P}^\mathrm{f}$, remain to be derived. This may be done without any reference to Gaussianity; using the law of total expectation, the forecast step of the KF becomes:

$$\begin{aligned}\boldsymbol{x}^\mathrm{f}&=\mathbb{E}\left(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}\right)=\mathbb{E}\big(\,\mathbb{E}(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1},\boldsymbol{x}_{t-1})|\boldsymbol{y}_{1:t-1}\big) &\text{(A.5)}\\ &=\mathbb{E}\big(\,\mathbb{E}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})|\boldsymbol{y}_{1:t-1}\big)=\mathbb{E}(\mathbf{F}\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})=\mathbf{F}\,\mathbb{E}(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})=\mathbf{F}\boldsymbol{x}^\mathrm{a}\,. &\text{(A.6)}\end{aligned}$$

Similarly, the law of total variance is used in the following:

$$\begin{aligned}\mathbf{P}^\mathrm{f}&=\mathbb{V}\mathrm{ar}\left(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}\right)=\mathbb{V}\mathrm{ar}\big(\,\mathbb{E}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})|\boldsymbol{y}_{1:t-1}\big)+\mathbb{E}\big(\,\mathbb{V}\mathrm{ar}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})|\boldsymbol{y}_{1:t-1}\big) &\text{(A.7)}\\ &=\mathbb{V}\mathrm{ar}(\mathbf{F}\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})+\mathbb{E}(\mathbf{Q}|\boldsymbol{y}_{1:t-1})=\mathbf{F}\mathbf{P}^\mathrm{a}\mathbf{F}^\mathsf{T}+\mathbf{Q}\,. &\text{(A.8)}\end{aligned}$$

Alternatively eqn. (A.4) can be derived by inserting the Gaussian pdfs $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ and $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ in the Chapman-Kolmogorov equation (1.12).

### A.2.2 Analysis step

As was done with the forecast step, it can be shown that the linear observation process, $\boldsymbol{y}_t=\mathbf{H}\boldsymbol{x}_t+\boldsymbol{r}_t$, means that $\boldsymbol{y}_t|\boldsymbol{x}_t\sim\mathcal{N}(\mathbf{H}\boldsymbol{x}_t,\mathbf{R})$. Inserting this and eqn. (A.4) into Bayes' rule (1.13) yields

$$\begin{aligned}-2\log p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})&=\|\mathbf{H}\boldsymbol{x}_t-\boldsymbol{y}_t\|^2_\mathbf{R}+\left\|\boldsymbol{x}_t-\boldsymbol{x}^\mathrm{f}\right\|^2_{\mathbf{P}^\mathrm{f}} &\text{(A.9)}\\ &=\|\boldsymbol{x}_t-\boldsymbol{x}^\mathrm{a}\|^2_{\mathbf{P}^\mathrm{a}}+\text{const.} &\text{(A.10)}\end{aligned}$$

where the second line, including the specifications of $\boldsymbol{x}^\mathrm{a}$, and $\mathbf{P}^\mathrm{a}$, is given by Lemma A.2, and the constant is with respect to $\boldsymbol{x}_t$. Thus,

$$\boldsymbol{x}_t|\boldsymbol{y}_{1:t}\sim\mathcal{N}(\boldsymbol{x}^\mathrm{a},\mathbf{P}^\mathrm{a})\,, \tag{A.11}$$

and the forecast-analysis cycle can start over again. Lemma A.2 gives the moments, $\boldsymbol{x}^{\mathrm{a}}$ and $\mathbf{P}^{\mathrm{a}}$, in several forms, each with a particular utility. Note that both Corollaries A.1 and A.2 are employed. Since Lemma A.2 is pure matrix algebra, the notation is simplified to

$$\boldsymbol{x}_t = \boldsymbol{x} \qquad \boldsymbol{y}_t = \boldsymbol{y} \qquad \boldsymbol{x}^{\mathrm{f}} = \boldsymbol{b} \qquad \mathbf{P}^{\mathrm{f}} = \mathbf{B} \qquad \boldsymbol{x}^{\mathrm{a}} = \boldsymbol{\mu} \qquad \mathbf{P}^{\mathrm{a}} = \mathbf{P}\,. \tag{A.12}$$

**Lemma A.2** – Completing the square**.**
*For any* $\mathbf{R}$*,* $\mathbf{B}$*, and* $\mathbf{H}$ *as in Corollary A.1,*[1] *and* $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^M$ *and* $\boldsymbol{y} \in \mathbb{R}^P$*,*

$$J = \|\mathbf{H}\boldsymbol{x} - \boldsymbol{y}\|^2_{\mathbf{R}} + \|\boldsymbol{x} - \boldsymbol{b}\|^2_{\mathbf{B}} \;=\; \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} + \|\boldsymbol{y} - \mathbf{H}\boldsymbol{b}\|^2_{\mathbf{C}}\,. \tag{A.13}$$

*where* $\boldsymbol{\mu}$*,* $\mathbf{P}$*, and* $\mathbf{C}$ *are given in the derivation below.*

*Proof.* Expanding the squares of the left hand side of eqn. (A.13), gathering terms, and completing the square in $\boldsymbol{x}$ yields

$$J = \boldsymbol{x}^{\mathsf{T}}\left(\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)\boldsymbol{x} - 2\boldsymbol{x}^{\mathsf{T}}\left[\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\boldsymbol{y} + \mathbf{B}^{-1}\boldsymbol{b}\right] + \|\boldsymbol{b}\|^2_{\mathbf{B}} + \|\boldsymbol{y}\|^2_{\mathbf{R}} \tag{A.14}$$

$$= \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} - \|\boldsymbol{\mu}\|^2_{\mathbf{P}} + \|\boldsymbol{b}\|^2_{\mathbf{B}} + \|\boldsymbol{y}\|^2_{\mathbf{R}}\,, \tag{A.15}$$

where we have defined

$$\mathbf{P} = (\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}\,, \tag{A.16a}$$

$$\boldsymbol{\mu} = \mathbf{P}\left[\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\boldsymbol{y} + \mathbf{B}^{-1}\boldsymbol{b}\right]\,, \tag{A.16b}$$

which we can identify as the posterior mean and variance by comparison to eqn. (A.10). Numerically, it is better to compute the moments by

$$\mathbf{P} = (\mathbf{B}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{I}_m)^{-1}\mathbf{B}\,, \tag{A.17a}$$

$$\boldsymbol{\mu} = (\mathbf{B}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{I}_m)^{-1}\left[\mathbf{B}\mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\boldsymbol{y} + \boldsymbol{b}\right]\,, \tag{A.17b}$$

as this avoids the inversion of $\mathbf{B}^{-1}$. However, if $P < M$, then the Kalman gain form, which inverts a $P \times P$ symmetric matrix, is generally preferable; applying Corollary A.1 to eqn. (A.16a) yields eqn. (A.18a)

$$\mathbf{P} = [\mathbf{I}_m - \mathbf{K}\mathbf{H}]\mathbf{B}\,, \tag{A.18a}$$

$$\boldsymbol{\mu} = \boldsymbol{b} + \mathbf{K}\left[\boldsymbol{y} - \mathbf{H}\boldsymbol{b}\right]\,, \tag{A.18b}$$

(eqn. (A.18b) is derived further below) where

$$\mathbf{C} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\,, \tag{A.19a}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathsf{T}}\mathbf{C}^{-1}\,, \tag{A.19b}$$

are the total forecast observation variance and the Kalman gain, respectively. Comparing eqns. A.18 and A.19 to the scalar expressions, eqns. (A.41), it is seen that the Kalman gain is

[1] With regards to the KF, $\mathbf{P}^{\mathrm{f}}$ can be assumed invertible because the prior, $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$, is assumed proper and not degenerate. This will hold if $\mathbf{F}$ is invertible, and $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ is proper and not degenerate, etc. On the other hand, $\mathbf{R}$ can be assumed invertible, because zero eigenvalues would imply perfect observations, which is unrealistic, while infinite eigenvalues would correspond to a component of the observation carrying no information at all which can in any case be circumvented by removing this (known) component from the observation (i.e. redefining the observation operator

essentially a ratio of covariances. Two useful identities for this ratio,

$$\mathbf{K} = \mathbf{B}\mathbf{H}^\mathsf{T}\mathbf{C}^{-1} = \mathbf{P}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\,, \tag{A.20a}$$

$$\mathbf{I}_m - \mathbf{K}\mathbf{H} = \mathbf{P}\mathbf{B}^{-1}\,, \tag{A.20b}$$

follow from Corollary A.2 and eqn. (A.18a) respectively. Eqns. (A.20) applied to eqn. (A.16b) can then be used to derive eqn. (A.18b). See Lewis et al. [75, P. 497] and Anderson and Moore [3, P. 147] for further forms of, and discussions on, the KF formulae.

For the purpose of deriving the KF formulae, it is not necessary to develop eqn. (A.15) any further, because $\|\boldsymbol{\mu}\|^2_{\mathbf{P}} + \|\boldsymbol{b}\|^2_{\mathbf{B}} + \|\boldsymbol{y}\|^2_{\mathbf{R}}$ is constant with respect to $\boldsymbol{x}$. However, for the derivation of the RTS smoother, or the finite-size EnKF, the remainder is still of interest. Substituting eqns. (A.16) into eqn. (A.15), expanding, and gathering terms yields

$$J = \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} - \left\|\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\boldsymbol{y} + \mathbf{B}^{-1}\boldsymbol{b}\right\|^2_{\mathbf{P}^{-1}} + \|\boldsymbol{b}\|^2_{\mathbf{B}} + \|\boldsymbol{y}\|^2_{\mathbf{R}} \tag{A.21}$$

$$= \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} + \boldsymbol{y}^\mathsf{T}\left\{\mathbf{R}^{-1} - (\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})^\mathsf{T}\mathbf{P}(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})\right\}\boldsymbol{y} - 2\boldsymbol{y}^\mathsf{T}(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})^\mathsf{T}\mathbf{P}\mathbf{B}^{-1}\boldsymbol{b} \tag{A.22}$$

$$- \|\boldsymbol{b}\|^2_{\mathbf{B}\mathbf{P}^{-1}\mathbf{B}} + \|\boldsymbol{b}\|^2_{\mathbf{B}}.$$

Completing the square on $\boldsymbol{y}$ yields

$$J = \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} + \|\boldsymbol{y} - \boldsymbol{d}\|^2_{\tilde{\mathbf{C}}} - \|\boldsymbol{d}\|^2_{\tilde{\mathbf{C}}} - \|\boldsymbol{b}\|^2_{\mathbf{B}\mathbf{P}^{-1}\mathbf{B}} + \|\boldsymbol{b}\|^2_{\mathbf{B}}\,, \tag{A.23}$$

with the definitions

$$\tilde{\mathbf{C}} = \left\{\mathbf{R}^{-1} - (\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})^\mathsf{T}\mathbf{P}(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})\right\}^{-1}\,, \tag{A.24a}$$

$$\boldsymbol{d} = \tilde{\mathbf{C}}(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})^\mathsf{T}\mathbf{P}\mathbf{B}^{-1}\boldsymbol{b}\,. \tag{A.24b}$$

Equation (A.20a) and eqn. (A.19a) are used to show what the notation hints at,

$$\tilde{\mathbf{C}} = \left\{\mathbf{R}^{-1} - (\mathbf{H}^\mathsf{T}\mathbf{R}^{-1})^\mathsf{T}\mathbf{B}\mathbf{H}^\mathsf{T}\mathbf{C}^{-1}\right\}^{-1} \tag{A.25}$$

$$= \left\{\mathbf{R}^{-1} - \mathbf{R}^{-1}\left(\mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\right)\mathbf{C}^{-1}\right\}^{-1} \tag{A.26}$$

$$= \mathbf{C}\left\{\mathbf{C} - \mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\right\}^{-1}\mathbf{R} \tag{A.27}$$

$$= \mathbf{C}\mathbf{R}^{-1}\mathbf{R} \tag{A.28}$$

$$= \mathbf{C}\,, \tag{A.29}$$

while $\boldsymbol{d}$ is simplified using eqns. (A.16a) and (A.19a),

$$\boldsymbol{d} = (\mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T} + \mathbf{R})\mathbf{R}^{-1}\mathbf{H}(\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}\boldsymbol{b} \tag{A.30}$$

$$= (\mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{H})(\mathbf{B}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{I}_m)^{-1}\boldsymbol{b} \tag{A.31}$$

$$= \mathbf{H}(\mathbf{B}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{I}_m)(\mathbf{B}\mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{I}_m)^{-1}\boldsymbol{b} \tag{A.32}$$

$$= \mathbf{H}\boldsymbol{b}\,. \tag{A.33}$$

Thus, eqn. (A.23) becomes

$$J = \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} + \|\boldsymbol{y} - \mathbf{H}\boldsymbol{b}\|^2_{\mathbf{C}} - \|\mathbf{H}\boldsymbol{b}\|^2_{\mathbf{C}} - \|\boldsymbol{b}\|^2_{\mathbf{B}\mathbf{P}^{-1}\mathbf{B}} + \|\boldsymbol{b}\|^2_{\mathbf{B}} \tag{A.34}$$

$$= \|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\mathbf{P}} + \|\boldsymbol{y} - \mathbf{H}\boldsymbol{b}\|^2_{\mathbf{C}} + \boldsymbol{b}^\mathsf{T}\mathbf{D}^{-1}\boldsymbol{b}\,, \tag{A.35}$$

with

$$\mathbf{D}^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{P}\mathbf{B}^{-1} + \mathbf{H}^{\mathsf{T}}\mathbf{C}^{-1}\mathbf{H} \tag{A.36}$$
$$= \mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{I}_m - \mathbf{K}\mathbf{H}) + \mathbf{B}^{-1}\mathbf{K}\mathbf{H} \tag{A.37}$$
$$= 0\,, \tag{A.38}$$

where eqn. (A.20b) and eqn. (A.19b) were used on the second line. □

## A.3 Interpreting the scalar KF

Comprehension of matrix identities can often be enhanced by investigating the scalar case. This section interprets the various KF equations by doing so, with $\boldsymbol{x}, \boldsymbol{b}, \boldsymbol{y} \in \mathbb{R}$, $\mathbf{H} = 1$, and $\mathbf{B}, \mathbf{R} > 0$. Equation (A.13) is then

$$\frac{(\boldsymbol{x}-\boldsymbol{b})^2}{\mathbf{B}} + \frac{(\boldsymbol{x}-\boldsymbol{y})^2}{\mathbf{R}} = \frac{(\boldsymbol{x}-\boldsymbol{\mu})^2}{\mathbf{P}} + \frac{(\boldsymbol{y}-\boldsymbol{b})^2}{\mathbf{R}+\mathbf{B}}\,, \tag{A.39}$$

where $\boldsymbol{\mu}$ and $\mathbf{P}$ are given by eqns. (A.16),

$$\mathbf{P} = \frac{1}{1/\mathbf{B} + 1/\mathbf{R}}\,, \tag{A.40a}$$
$$\boldsymbol{\mu} = \mathbf{P}(\boldsymbol{b}/\mathbf{B} + \boldsymbol{y}/\mathbf{R})\,. \tag{A.40b}$$

This formulation shows that the posterior moments are "weighted averages": the posterior mean is a weighted average of the observation and the prior mean, and the posterior precision is the sum of the observation precision and the background precision.

Next, consider the Kalman gain formulation of the posterior moments:

$$\mathbf{K} = \frac{\mathbf{B}}{\mathbf{B}+\mathbf{R}}\,, \tag{A.41a}$$
$$\mathbf{P} = (1-\mathbf{K})\mathbf{B}\,, \tag{A.41b}$$
$$\boldsymbol{\mu} = \boldsymbol{b} + \mathbf{K}(\boldsymbol{y}-\boldsymbol{b})\,. \tag{A.41c}$$

Here, $\boldsymbol{\mu}$ is not obtained as a weighted average, but rather as a linear interpolation between the prior mean and the observations, controlled by the "gain" matrix $\mathbf{K}$. The definition (A.41a) of $\mathbf{K}$ is straightforward: it is the ratio of one variance to the total variance. As seen from eqn. (A.41b), the gain also linearly controls the reduction in variance.

In the multivariate case, the weighted average and the gain formulations are related through the Woodbury corollaries, A.1 and A.2. In the scalar case, however, both of these are just a step away from the trivial identity

$$\frac{1}{1/\mathbf{B} + 1/\mathbf{R}} = \frac{\mathbf{B}\mathbf{R}}{\mathbf{B}+\mathbf{R}}\,. \tag{A.42}$$

Similarly, the identity eqn. (A.20a) becomes just another ratio as well:

$$\mathbf{K} = \mathbf{P}/\mathbf{R} = \frac{1/\mathbf{R}}{1/\mathbf{R} + 1/\mathbf{B}}\,. \tag{A.43}$$

Consider again the right hand side of eqn. (A.39) or (A.13). The second term, $(\boldsymbol{y}-\boldsymbol{b})^2/(\mathbf{R}+\mathbf{B})$, is referred to as "model evidence" [54]. It is the mismatch between the prior mean and the data, normalized by their total variance. Although insignificant to the KF, the term plays a role for the RTS smoother [103, §. 7] it carries useful information for adjusting inflation parameters [20] and it may serve to interpret the stopping criteria for iterative methods [71, 79].

Now consider the first term, $(\boldsymbol{x}-\boldsymbol{\mu})^2/\mathbf{P}$. (i) It gathers all of the dependence in $\boldsymbol{x}$, facilitating the integration over $\boldsymbol{x}$. (ii) It is again quadratic, meaning that the posterior pdf of the KF is Gaussian. (iii) The posterior variance, $\mathbf{P}$, does not depend on the value of the model evidence. The last item is a remarkable [e.g. 3, example 3.6] property of Gaussian variables: the posterior variance is always reduced. This is puzzling; if the prior is very far from the data relative to $\mathbf{B}$, then maybe $\mathbf{B}$ was set too small? It is countered in Bocquet et al. [20] by a hierarchical approach. In contrast to the Gaussian case, in general, only the expected variance is reduced, and similarly, only the expected Shannon information is increased.

# Appendix B

# The SVD, linear inverse problems, and the pseudoinverse

This appendix presents the basic theory of the singular value decomposition (SVD), linear least squares, and the pseudoinverse.

## B.1 The SVD

The SVD is important both as a computational and analytic tool. The proofs for this section are given by Trefethen and Bau [123].

**Definition B.1** – The (full) SVD**.**
*For any positive integers $P$ and $M$, the SVD of any $\mathbf{H} \in \mathbb{R}^{P\times M}$ is the factorization*

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}, \tag{B.1}$$

*where $\mathbf{U} \in \mathbb{R}^{P^2}$ and $\mathbf{V} \in \mathbb{R}^{M^2}$ are orthogonal (i.e. their columns $\{\boldsymbol{u}_j \in \mathbb{R}^P \ ; \ j = 1, \ldots, P\}$ and $\{\boldsymbol{v}_j \in \mathbb{R}^M \ ; \ j = 1, \ldots, M\}$, are both ortho*normal*) and $\mathbf{\Sigma} \in \mathbb{R}^{P\times M}$ is diagonal and* non-negative*. The columns of $\mathbf{U}$ are known as the* left *singular vectors of $\mathbf{H}$. The columns of $\mathbf{V}$ are known as the* right *singular vectors of $\mathbf{H}$. The diagonal entries of $\mathbf{\Sigma}$, $\{\sigma_j \geq 0 \ ; \ j = 1, \ldots, \mathsf{min}(\{P, M\})\}$, are known as the singular values of $\mathbf{H}$, and are ordered in non-increasing order by convention.*

**Proposition B.1** – Existence and uniqueness**.**
*Every matrix $\mathbf{H} \in \mathbb{R}^{P\times M}$ has an SVD. By virtue of the sorting convention, and the non-negativity, the matrix $\mathbf{\Sigma}$ is uniquely determined. By virtue of the restriction that $\mathbf{U}$ and $\mathbf{V}$ be real, $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are uniquely determined up to a multiplication by $-1$ of both, provided that the $j$-th singular value is unique, i.e. $\sigma_{j+1} < \sigma_j < \sigma_{j-1}$).*

Note that, as opposed to the eigenvalue decomposition, the SVD exists for rectangular and/or rank-deficient matrices. However, for SPD matrices, the SVD and the eigenvalue decomposition coincide. Proposition B.2 catalogues some properties.

**Proposition B.2** – SVD properties**.**

- $\|\mathbf{H}\|_2 = \max\limits_{\boldsymbol{x}\in\mathbb{R}^M} \|\mathbf{H}\boldsymbol{x}\|_2 = \sigma_1$ .
- *If $k = \mathsf{rank}(\mathbf{H})$, then $k = \mathsf{max}\{j \ ; \ \sigma_j > 0\}$, and*
- *the range of $\mathbf{H}$ is the span of the $k$ first left singular vectors, $(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\})$.*
- *If $\mathbf{H}$ is square, $M$-by-$M$, then $|\mathbf{H}| = \prod_{j=1}^{M} \sigma_j$ .*

In addition, by virtue of their orthogonality, the singular vectors (and the singular values) can have important physical or statistical interpretations, as illustrated by "empirical orthogonal eigenfunctions", or "principal component analysis" [134].

Defined below, the reduced SVD only computes the singular vectors with non-zero singular values. There may be significant computational cost savings in only calculating the reduced SVD instead of the full SVD. Golub and Van Loan [48, §8.6.4] and [123, §1.4] provide computational details.

**Definition B.2** – The reduced SVD**.**
*Let* $k = \mathsf{rank}(\mathbf{H})$*. The reduced SVD of* $\mathbf{H}$ *is the factorization*

$$\mathbf{H} = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^{\mathsf{T}}\,, \tag{B.2}$$

*where* $\hat{\mathbf{U}} \in \mathbb{R}^{P\times k}$*,* $\hat{\mathbf{V}} \in \mathbb{R}^{M\times k}$ *and* $\hat{\mathbf{\Sigma}} \in \mathbb{R}^{k^2}$ *can be derived from* $\mathbf{U}$*,* $\mathbf{V}$ *and* $\mathbf{\Sigma}$ *by removing the appropriate number of rows and columns corresponding to any zero singular values. By construction, existence and uniqueness follow from the full SVD.*

Note that $\hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathsf{T}} \neq \mathbf{I}_p$, even though $\hat{\mathbf{U}}^{\mathsf{T}}\hat{\mathbf{U}} = \mathbf{I}_k$. I.e., even though its columns are orthonormal, $\hat{\mathbf{U}}$ is not orthogonal. The same applies for $\hat{\mathbf{V}}$.

## B.2 Linear inverse problems

This section compiles ideas from Wunsch [135] and Ben-Israel and Greville [11]. Consider the linear system of equations

$$\mathbf{H}\boldsymbol{x} = \boldsymbol{y}\,, \tag{B.3}$$

where $\mathbf{H} \in \mathbb{R}^{P\times M}$ again. As is typically the case for inverse problems, $\mathbf{H}$ is rectangular, i.e. $P \neq M$, and therefore not invertible. However, it is assumed, for the moment, that $\mathbf{H}$ is full-rank: $k = \mathsf{rank}(\mathbf{H}) = \mathsf{min}(P, M)$.

### B.2.1 The overdetermined case

The system (B.3) is "overdetermined" when $P > M$. In other words, there are more observations in $\boldsymbol{y}$ (and hence equations) than degrees of freedom in $\boldsymbol{x}$, and therefore eqn. (B.3) has no solution. One approach that renders the problem well-posed is to formulate it as a minimization problem on the residual:

$$\underset{\boldsymbol{x}\in\mathbb{R}^M}{\mathsf{minimize}} \quad J(\boldsymbol{x}) = \|\mathbf{H}\boldsymbol{x} - \boldsymbol{y}\|_2^2\,. \tag{B.4}$$

This can be reformulated as the constrained optimization problem

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^M}{\mathsf{minimize}} \quad & J(\boldsymbol{x}, \boldsymbol{r}) = \|\boldsymbol{r}\|_2^2 \\ \mathsf{sub.\,to} \quad & \mathbf{H}\boldsymbol{x} + \boldsymbol{r} = \boldsymbol{y}\,. \end{aligned} \tag{B.5}$$

The problem (B.5) reduces to (B.4) by substitution. But the explicit residual, $\boldsymbol{r}$, serves to make the connection with a Gaussian noise probability (ref. section A.2.2), and thus motivates the constrained optimization approach. Moreover, writing the constraint as $\begin{bmatrix}\mathbf{H}\,, \mathbf{I}_p\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{r}\end{bmatrix} = \boldsymbol{y}$, it is seen that the problem is, in a way, *under*determined. The solution may be found by differentiation of eqn. (B.4), or by the geometric argument that the shortest line between a point and a (hyper)plane (the range of $\mathbf{H}$) is the perpendicular one: $0 = \mathbf{H}^{\mathsf{T}}\boldsymbol{r} = \mathbf{H}^{\mathsf{T}}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}_\star)$, yielding

$$\mathbf{H}^{\mathsf{T}}\mathbf{H}\boldsymbol{x}_\star = \mathbf{H}^{\mathsf{T}}\boldsymbol{y}\,, \tag{B.6}$$

where $\boldsymbol{x}_\star$ denotes the minimizing argument, obtained by $\boldsymbol{x}_\star = \mathbf{H}^L \boldsymbol{y}$, where

$$\mathbf{H}^L = \left(\mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T} \tag{B.7}$$

is the "left-inverse" of $\mathbf{H}$.

### B.2.2 The underdetermined case

The system (B.3) is "underdetermined" when $P < M$, in which case there are infinitely many solutions. A related problem that is well-posed is to find the solution with the minimal norm, i.e.

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^M}{\mathsf{minimize}} \quad J(\boldsymbol{x},\boldsymbol{q}) = \|\boldsymbol{q}\|_2^2 \\ &\mathsf{sub.\,to} \quad \begin{cases} \boldsymbol{x}+\boldsymbol{q}=0 \\ \mathbf{H}\boldsymbol{x}=\boldsymbol{y}\,. \end{cases} \end{aligned} \tag{B.8}$$

In this case, the constraint $\mathbf{H}\boldsymbol{x} = \boldsymbol{y}$ cannot be eliminated by substitution. But, using Lagrangian multipliers, i.e. $\lambda \in \mathbb{R}^P$, the stationarity conditions yield

$$\boldsymbol{x}_\star - \mathbf{H}^\mathsf{T}\lambda = 0\,, \tag{B.9}$$

$$\mathbf{H}\boldsymbol{x}_\star - \boldsymbol{y} = 0\,, \tag{B.10}$$

resulting in $\boldsymbol{x}_\star = \mathbf{H}^R\boldsymbol{y}$, where $\mathbf{H}^R$ is the "right-inverse",

$$\mathbf{H}^R = \mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{H}^\mathsf{T}\right)^{-1}\,, \tag{B.11}$$

to be compared with eqn. (B.7).

## B.3 The pseudoinverse

This section shows that, remarkably, both of the above "generalized inverses", $\mathbf{H}^R$ and $\mathbf{H}^L$, and hence their optimality properties, are unified by the pseudoinverse.

### B.3.1 The SVD pseudoinverse

The formulae (B.7) and (B.11), for the left and right inverses of $\mathbf{H}$ reduce to the same when written in terms of the reduced SVD:

$$\mathbf{H}^+ = \hat{\mathbf{V}}\hat{\mathbf{\Sigma}}^{-1}\hat{\mathbf{U}}^\mathsf{T}\,. \tag{B.12}$$

In this sense, the SVD unifies the left and right inverses. But the formula (B.12) is also defined when $k = \mathsf{rank}(\mathbf{H}) < \mathsf{min}(M,P)$, in which case both $\mathbf{H}\mathbf{H}^\mathsf{T}$ and $\mathbf{H}^\mathsf{T}\mathbf{H}$ are rank deficient. In effect, the matrix rank problems of $\mathbf{H}\mathbf{H}^\mathsf{T}$ and $\mathbf{H}^\mathsf{T}\mathbf{H}$ are automatically handled by the definition of the reduced SVD. This SVD inverse, $\mathbf{H}^+$, is called the pseudoinverse. The following theorem, from [47, Th. 5.5.1], summarizes these findings.

**Theorem B.1** – Pseudoinverse optimality**.**
*$\boldsymbol{x}_\star = \mathbf{H}^+\boldsymbol{y}$ minimizes (i) the 2-norm of the residual, $\|\mathbf{H}\boldsymbol{x}-\boldsymbol{y}\|_2^2$ and (ii) the 2-norm of $\boldsymbol{x}$ itself, and it does so in that order of priority.*

It can be shown that for a diagonal matrix, such as $\mathbf{\Sigma}$, the pseudoinverse is obtained by taking the reciprocal of each non-zero element on the diagonal, leaving the zeros in place. If $\mathbf{\Sigma}$ is rectangular, then $\mathbf{\Sigma}^+$ has the size of $\mathbf{\Sigma}^\mathsf{T}$. Using this, one can derive, from eqn. (B.12), that

$$\mathbf{H}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^\mathsf{T}\,, \tag{B.13}$$

which is useful to prove larger identities on the pseudoinverse.

### B.3.2 The limit pseudoinverse

Another appealing avenue for unifying $\mathbf{H}^R$ and $\mathbf{H}^L$ is to solve the problem

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^M}{\text{minimize}} && J(\boldsymbol{x},\boldsymbol{q},\boldsymbol{r}) = \|\boldsymbol{q}\|_2^2 + \|\boldsymbol{r}\|_2^2 \\ &\text{sub. to} && \begin{cases} \boldsymbol{x}+\boldsymbol{q} = 0 \\ \mathbf{H}\boldsymbol{x}+\boldsymbol{r} = \boldsymbol{y}\,, \end{cases} \end{aligned} \tag{B.14}$$

as a "mix" of the underdetermined and the overdetermined problem. Equivalently,

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^M}{\text{minimize}} && J(\boldsymbol{x},\boldsymbol{q},\boldsymbol{r}) = \|\boldsymbol{q}\|_2^2 + \|\boldsymbol{r}\|_2^2 \\ &\text{sub. to} && \begin{bmatrix}\mathbf{I}_m \\ \mathbf{H}\end{bmatrix}\boldsymbol{x} + \begin{bmatrix}\boldsymbol{q} \\ \boldsymbol{r}\end{bmatrix} = \begin{bmatrix}0 \\ \boldsymbol{y}\end{bmatrix}\,, \end{aligned} \tag{B.15}$$

which is in the form of problem (B.5). Therefore, eqn. (B.7) can be inserted, yielding

$$\boldsymbol{x}_\star = \left(\begin{bmatrix}\mathbf{I}_m & \mathbf{H}^\mathsf{T}\end{bmatrix}\begin{bmatrix}\mathbf{I}_m \\ \mathbf{H}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{I}_m & \mathbf{H}^\mathsf{T}\end{bmatrix}\begin{bmatrix}0 \\ \boldsymbol{y}\end{bmatrix} \tag{B.16}$$

$$= \left(\mathbf{I}_m + \mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T}\boldsymbol{y}\,. \tag{B.17}$$

Note that $\left(\mathbf{I}_m + \mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T}$ always exists since $\left(\mathbf{I}_m + \mathbf{H}^\mathsf{T}\mathbf{H}\right)$ is SPD.

A modification of the cost function by a weighting, $\varepsilon^2$,

$$J(\boldsymbol{x},\boldsymbol{q},\boldsymbol{r}) = \varepsilon^2\|\boldsymbol{q}\|_2^2 + \|\boldsymbol{r}\|_2^2\,, \tag{B.18}$$

does not pose a new challenge, because it can be transferred to the constraints by a scaling change of variables, $\boldsymbol{q} = \varepsilon\boldsymbol{q}'$. Thus, the solution of eqn. (B.17) can be inserted:

$$\boldsymbol{x}_\star = \left(\varepsilon^2\mathbf{I}_m + \mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T}\boldsymbol{y}\,. \tag{B.19}$$

Now, if $(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}$ or $(\mathbf{H}\mathbf{H}^\mathsf{T})^{-1}$ exists, then, respectively,

$$\lim_{\varepsilon\to 0}\quad \left(\varepsilon^2\mathbf{I}_m + \mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T} = \left(\mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T} \tag{B.20}$$

$$\text{or}\quad \lim_{\varepsilon\to 0}\quad \mathbf{H}^\mathsf{T}\left(\varepsilon^2\mathbf{I}_m + \mathbf{H}\mathbf{H}^\mathsf{T}\right)^{-1} = \mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{H}^\mathsf{T}\right)^{-1}\,. \tag{B.21}$$

But, as can be shown using Corollary A.2, or an SVD of $\mathbf{H}$,

$$\left(\varepsilon\mathbf{I}_m + \mathbf{H}^\mathsf{T}\mathbf{H}\right)^{-1}\mathbf{H}^\mathsf{T} = \mathbf{H}^\mathsf{T}\left(\varepsilon\mathbf{I}_p + \mathbf{H}\mathbf{H}^\mathsf{T}\right)^{-1} \tag{B.22}$$

for any $\varepsilon > 0$. Thus, for any $\varepsilon > 0$, the left hand sides of eqns. (B.20) and (B.21) are the same. On the other hand, for $\varepsilon = 0$, one of the right hand sides will not exist, since $P \neq M$.

Remarkably, however, both *limits* exist, even if $(\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}$ or $(\mathbf{H}\mathbf{H}^\mathsf{T})^{-1}$ does not. Moreover, the unique limit value is the pseudoinverse, $\mathbf{H}^+$, as defined in section B.3.1.

A proof of the above statements is provided by Ben-Israel and Greville [11, §3]. Heuristically, it can be argued that (i) the cost function formulation, eqn. (B.18), indicates that the limiting solution (B.19) should minimize the 2-norm of the residual, as is the first priority of $\mathbf{H}^+$, according to Theorem B.1. But, (ii) for any $\varepsilon > 0$, the solution is also trying to minimize $\|\boldsymbol{q}\|_2^2$. Thus, (iii) providing the limit exists, even if the minimum-residual solution is non-unique, one would expect the limit to pick out the solution that also minimizes $\|\boldsymbol{q}\|_2^2$.

### B.3.3 The Moore-Penrose pseudoinverse

Lastly, the Moore-Penrose characterization of the pseudoinverse:

$$\mathbf{H}\mathbf{H}^+\mathbf{H} = \mathbf{H}\,, \tag{B.23}$$

$$\mathbf{H}^+\mathbf{H}\mathbf{H}^+ = \mathbf{H}^+\,, \tag{B.24}$$

$$(\mathbf{H}\mathbf{H}^+)^\mathsf{T} = \mathbf{H}\mathbf{H}^+\,, \tag{B.25}$$

$$(\mathbf{H}^+\mathbf{H})^\mathsf{T} = \mathbf{H}^+\mathbf{H}\,, \tag{B.26}$$

can also be shown to be equivalent to the SVD and the limit definitions [13, §1].

# Appendix C

# Notation

## Abbreviations

| Abbrev. | Meaning | Reference |
|---|---|---|
| DA | data assimilation | section 1.1 |
| EnKF | ensemble Kalman filter | chapter 2 |
| HMM | hidden Markov model | section 1.2.1 |
| iid | independent and identically distributed | |
| KF | Kalman filter | section 2.1 |
| pdf | probability density function | |
| RMSE | root mean square error | section 3.1 |
| SPD | symmetric, positive-definite | page 23 |
| SVD | singular value decomposition | section B.1 |
| Th. | theorem | |

## Symbols in frequent use

| Symbol | Definition | Reference |
|---|---|---|
| $\boldsymbol{x}_t$ | The state variable at time index $t$. | section 1.2 |
| $\boldsymbol{y}_t$ | The observation at time index $t$. | section 1.2 |
| $f(\boldsymbol{x}_t)$ | The dynamical forecast model. | section 1.2 |
| $h(\boldsymbol{x}_t)$ | The observation model. | section 1.2 |
| $m$ | State vector size. | section 1.2 |
| $p$ | Observation vector size. | section 1.2 |
| $k_1{:}k_2$ | The sequence of integers from $k_1$ to $k_2$. | |
| $x_{k_1:k_2}$ | The sequence $\{x_k \ ; \ k = k_1, \ldots, k_2\}$. | |
| $p()$ | Probability density function of the random variable indicated by its argument. | |
| $\mathbf{I}_k$ | Identity matrix of size $k \times k$. | |
| $\boldsymbol{e}_i$ | Unit coordinate vector in the $i$-th dimension. | |
| $\mathbf{M}^\mathsf{T}$ | Transpose of the arbitrary matrix $\mathbf{M}$. | |
| $\mathbf{M}^+$ | Pseudoinverse of $\mathbf{M}$. | section B.3 |
| $\mathbf{M}^{1/2}, \mathbf{M}^{\mathsf{T}/2}$ | A matrix square root of $\mathbf{M}$, and its transpose. $\mathbf{M}^{1/2}\mathbf{M}^{\mathsf{T}/2} = \mathbf{M}$. | section 4.2 |
| $\mathbf{M}^{1/2}_{\text{sym}}$ | The symmetric square root of the SPD matrix $\mathbf{M}$. | section 4.2 |
| $[\mathbf{M}]_{i,j}$ | Element $(i,j)$ of $\mathbf{M}$. | |
| tr | Trace of a matrix. | |
| diag | Diagonal of a matrix. | |
| $\|\boldsymbol{v}\|^2_{\mathbf{M}}$ | $= \boldsymbol{v}^\mathsf{T}\mathbf{M}^{-1}\boldsymbol{v}$. 2-norm of the generic vector $\boldsymbol{v}$ weighted by the inverse invertible matrix $\mathbf{M}$. | |
| $\|\boldsymbol{v}\|^2_{\mathbf{M}}$ | $= \boldsymbol{v}^\mathsf{T}\mathbf{M}^{-1}\boldsymbol{v}$. 2-norm of the generic vector $\boldsymbol{v}$ weighted by the inverse invertible matrix $\mathbf{M}$. | |
| $\mathbb{E}(\boldsymbol{x})$ | $= \int \boldsymbol{x}\, p(\boldsymbol{x})\, d\boldsymbol{x}$. Expectation of the random variable $\boldsymbol{x}$. | |
| $\mathbb{V}\text{ar}(\boldsymbol{x})$ | Variance of the random variable $\boldsymbol{x}$. If $\boldsymbol{x}$ is a vector, then the output is a covariance *matrix*. $\mathbb{V}\text{ar}(\boldsymbol{x}) = \int (\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))(\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))^\mathsf{T}\, p(\boldsymbol{x})\, d\boldsymbol{x}$. | |
| $\mathbb{C}\text{ov}(\boldsymbol{x}, \boldsymbol{y})$ | Cross-covariance of $\boldsymbol{x}$ and $\boldsymbol{y}$. $\mathbb{C}\text{ov}(\boldsymbol{x}, \boldsymbol{y}) = \int (\boldsymbol{x} - \mathbb{E}(\boldsymbol{x}))(\boldsymbol{y} - \mathbb{E}(\boldsymbol{y}))^\mathsf{T}\, p(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{x}\, d\boldsymbol{y}$. | |
| $\mathcal{N}(\boldsymbol{x}\|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | Gaussian (Normal) pdf with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathcal{N}(\boldsymbol{x}\|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \|2\pi\boldsymbol{\Sigma}\|^{-1/2} \exp\big(-\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{\mu}\|^2_{\boldsymbol{\Sigma}}\big)$. | |
| $\boldsymbol{x}^a, \boldsymbol{x}^f$ | The analysis/forecast *exact* mean. $\boldsymbol{x}^a = \mathbb{E}(\boldsymbol{x}_t\|\boldsymbol{y}_{1:t})$, $\boldsymbol{x}^f = \mathbb{E}(\boldsymbol{x}_t\|\boldsymbol{y}_{1:t-1})$. | section 2.1 |
| $\mathbf{P}^a, \mathbf{P}^f$ | The analysis/forecast *exact* covariance matrix. $\mathbf{P}^a = \mathbb{V}\text{ar}(\boldsymbol{x}_t\|\boldsymbol{y}_{1:t})$, $\mathbf{P}^f = \mathbb{V}\text{ar}(\boldsymbol{x}_t\|\boldsymbol{y}_{1:t-1})$. | section 2.1 |
| $\boldsymbol{x}_n$ | An ensemble member/realisation. Typically associated with a given state (and thus time index), $\boldsymbol{x}_t$, as well as a given conditioning, e.g. $\|\boldsymbol{y}_{1:t}$. If this is of relevance, it is indicated by superscripts $a$ and $f$, as described below. | section 2.2 |
| $N$ | Ensemble size. | chapter 2 |
| $\mathbb{1}$ | Vector of ones. Length: $N$. | |
| $\mathbf{E}, \mathbf{E}^a, \mathbf{E}^f$ | The ensemble matrix, whose columns are the individual members. $\mathbf{E} = \big[\boldsymbol{x}_1, \ \ldots \ \boldsymbol{x}_n, \ \ldots \ \boldsymbol{x}_N\big] \in \mathbb{R}^{m \times N}$. The superscript $a$ (resp. $f$) indicates that it is an "analysis" (resp. "forecast") entity: it estimates $\boldsymbol{x}_t$ based on $\boldsymbol{y}_{1:t}$ (resp. $\boldsymbol{y}_{1:t-1}$). | section 2.2 |

| Symbol | Definition | Reference |
|---|---|---|
| $\mathbf{A}, \mathbf{A}^a, \mathbf{A}^f$ | The ensemble anomalies matrix. $\mathbf{E} = \boldsymbol{x}\mathbb{1}^\mathsf{T} + \mathbf{A}$. | section 2.2 |
| $\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}^a, \bar{\boldsymbol{x}}^f$ | The *ensemble* mean. $\bar{\boldsymbol{x}} = \frac{1}{N}\mathbf{E}\mathbb{1}$. | section 2.2 |
| $\bar{\mathbf{P}}, \bar{\mathbf{P}}^a, \bar{\mathbf{P}}^f$ | The *ensemble* covariance. $\bar{\mathbf{P}} = \frac{1}{N-1}\mathbf{A}\mathbf{A}^\mathsf{T}$. | section 2.2 |
| $\bar{\mathbf{K}}, \mathbf{K}$ | The ensemble and exact Kalman gain matrices. | section 2.3 |
| $\bar{\boldsymbol{\delta}}$ | Mean innovation. | section 4.1 |
| $\mathbf{Y}$ | Anomalies of the observed ensemble. | section 4.1 |
| $\boldsymbol{\Pi}$ | Orthogonal projection matrix. | page 8 |

# Bibliography


[1] Sigurd Aanonsen, Geir Nævdal, Dean Oliver, Albert Reynolds, and Brice Vallès. The ensemble Kalman filter in reservoir engineering–a review. *SPE Journal*, 14(3):393–412, 2009.

[2] M. U. Altaf, T. Butler, T. Mayo, X. Luo, C. Dawson, A. W. Heemink, and I. Hoteit. A comparison of ensemble Kalman filters for storm surge assimilation. *Monthly Weather Review*, 142(8):2899–2914, 2014.

[3] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.

[4] Jeffrey L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001.

[5] Jeffrey L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131(4):634–642, 2003.

[6] Jeffrey L. Anderson. A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138(11):4186–4198, 2010.

[7] Jeffrey L. Anderson and Stephen L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, 1999.

[8] Jose Angulo, Hwa-Lung Yu, Andrea Langousis, Alexander Kolovos, Jinfeng Wang, A. Madrid, and George Christakos. Spatiotemporal infectious disease modeling: A BME-SIR approach. *PloS One*, 8(9), 2013.

[9] J. D. Annan, D. J. Lunt, J. C. Hargreaves, and P. J. Valdes. Parameter estimation in an atmospheric GCM using the ensemble Kalman filter. *Nonlinear processes in geophysics*, 12(3):363–371, 2005.

[10] Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

[11] Adi Ben-Israel and Thomas N. E. Greville. *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 15. Springer-Verlag, New York, second edition, 2003.

[12] Riccardo Benedetti. Scoring rules for forecast verification. *Monthly Weather Review*, 138 (1):203–211, 2010.

[13] Andrew F. Bennett. *Inverse Methods in Physical Oceanography*. Cambridge University Press, 1992.

[14] Laurent Bertino, Geir Evensen, and Hans Wackernagel. Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71(2):223–241, 2003.

[15] J. Eric Bickel. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, 4(2):49–65, 2007.

[16] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[17] Craig H. Bishop, Brian J. Etherton, and Sharanya J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129(3):420–436, 2001.

[18] Marc Bocquet and Pavel Sakov. An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1521–1535, 2014.

[19] Marc Bocquet, Carlos A. Pires, and Lin Wu. Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, 138(8):2997–3023, 2010.

[20] Marc Bocquet, Patrick N. Raanes, and Alexis Hannart. Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, 22(6):645–662, 2015.

[21] Jean-Michel Brankart, Clément Ubelmann, Charles-Emmanuel Testut, Emmanuel Cosme, Pierre Brasseur, and Jacques Verron. Efficient parameterization of the observation error covariance matrix for square root or ensemble Kalman filters: application to ocean altimetry. *Monthly Weather Review*, 137(6):1908–1927, 2009.

[22] Yoram Bresler. Two-filter formulae for discrete-time non-linear Bayesian smoothing. *International Journal of Control*, 43(2):629–641, 1986.

[23] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[24] Jochen Bröcker. Reliability, sufficiency, and the decomposition of proper scores. *Quarterly Journal of the Royal Meteorological Society*, 135(643):1512–1519, 2009.

[25] G. Burgers, P. Jan van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6):1719–1724, 1998.

[26] Emmanuel J. Candès and Michael B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[27] Alberto Carrassi and Stéphane Vannitsem. Accounting for model error in variational data assimilation: A deterministic formulation. *Monthly Weather Review*, 138(9):3369–3386, 2010.

[28] Alberto Carrassi, Michael Ghil, Anna Trevisan, and Francesco Uboldi. Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(2): 023112, 2008.

[29] Yan Chen and Dean S. Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.

[30] Yan Chen and Dean S. Oliver. History matching of the Norne full field model using an iterative ensemble smoother-(SPE-164902). In *75th EAGE Conference & Exhibition incorporating SPE EUROPEC*, 2013.

[31] Yan Chen and Dean S. Oliver. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, 17(4):689–703, 2013.

[32] Remi Chou, Yvo Boers, Martin Podt, and Matthieu Geist. Performance evaluation for particle filters. In *Information Fusion (FUSION), Proceedings of the 14th International Conference on*, pages 1–7. IEEE, 2011.

[33] H.M. Christensen. Decomposition of a new proper score for verification of ensemble forecasts. *Monthly Weather Review*, 143(5):1517–1532, 2015.

[34] Alexandre A. Emerick and Albert C. Reynolds. Investigation of the sampling performance of ensemble-based methods with a simple reservoir model. *Computational Geosciences*, 17 (2):325–350, 2013.

[35] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99: 10–10, 1994.

[36] G. Evensen. The ensemble Kalman filter for combined state and parameter estimation. *Control Systems, IEEE*, 29(3):83–104, 2009.

[37] Geir Evensen. Advanced data assimilation for strongly nonlinear dynamics. *Monthly Weather Review*, 125(6):1342–1354, 1997.

[38] Geir Evensen. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6):539–560, 2004.

[39] Geir Evensen. *Data Assimilation*. Springer, 2 edition, 2009.

[40] Geir Evensen, Joakim Hove, Hilde Meisingset, Edel Reiso, Knut Sponheim Seim, Øystein Espelid, et al. Using the EnKF for assisted history matching of a North Sea reservoir model. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2007.

[41] C. Farmer. Bayesian field theory applied to scattered data interpolation and inverse problems. *Algorithms for Approximation*, pages 147–166, 2007.

[42] M. Fisher, M. Leutbecher, and G. A. Kelly. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3235–3246, 2005.

[43] Marco Luca Flavio Frei. *Ensemble Kalman Filtering and Generalizations*. PhD thesis, ETH Zurich, Dept. of Mathematics, 2013.

[44] Reinhard Furrer and Thomas Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98(2): 227–255, 2007.

[45] C. W. Gardiner. *Handbook of Stochastic Methods*. Springer, 2004.

[46] Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[47] Gene H. Golub and Charles F. Van Loan. *Matrix Computations. 1996*. Johns Hopkins University, Press, Baltimore, MD, USA, third edition, 1996.

[48] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

[49] Mohinder S. Grewal and Angus P. Andrews. Applications of Kalman filtering in aerospace 1960 to the present. *Control Systems, IEEE*, 30(3):69–78, 2010.

[50] A. K. Gupta and D. K. Nagar. *Matrix variate distributions*, volume 104 of *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics*. Chapman & Hall/CRC, Boca Raton, FL, 2000.

[51] K. S. Gurumoorthy, C. Grudzien, A. Apte, A. Carrassi, and C. K. R. T. Jones. Rank deficiency of Kalman error covariance matrices in linear perfect model. *arXiv preprint arXiv:1108.0158*, mar 2015.

[52] William W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31(2):221–239, 1989.

[53] R. G. Hanea, G. J. M. Velders, A. J. Segers, M. Verlaan, and A. W. Heemink. A hybrid Kalman filter algorithm for large-scale atmospheric chemistry data assimilation. *Monthly Weather Review*, 135(1):140–151, 2007.

[54] Alexis Hannart, Alberto Carrassi, Marc Bocquet, Michael Ghil, Philippe Naveau, Manuel Pulido, Juan Ruiz, and Pierre Tandeo. DADA: data assimilation for the detection and attribution of weather and climate-related events. *Climatic Change*, 136(2):155–174, 2016.

[55] P. C. Hansen. Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms*, 6(1):1–35, 1994.

[56] Arnold W. Heemink, Martin Verlaan, and Arjo J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, 129(7):1718–1728, 2001.

[57] R. C. Hilborn and J. C. Sprott. *Chaos And Nonlinear Dynamics: An Introduction For Scientists And Engineers*. Oxford University Press, New York, 1994.

[58] Matthew J. Hoffman, Steven J. Greybush, R. John Wilson, Gyorgyi Gyarmati, Ross N. Hoffman, Eugenia Kalnay, Kayo Ide, Eric J. Kostelich, Takemasa Miyoshi, and Istvan Szunyogh. An ensemble Kalman filter data assimilation system for the Martian atmosphere: Implementation and simulation experiments. *Icarus*, 209(2):470–481, 2010.

[59] Mevin B. Hooten, Jessica Anderson, and Lance A. Waller. Assessing North American influenza dynamics with a statistical SIRS model. *Spatial and spatio-temporal epidemiology*, 1(2):177–185, 2010.

[60] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[61] Peter L. Houtekamer and Herschel L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.

[62] Peter L. Houtekamer, Herschel L. Mitchell, Gerard Pellerin, Mark Buehner, Martin Charron, Lubos Spacek, and Bjarne Hansen. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, 133(3):604–620, 2005.

[63] Brian R. Hunt, Eric J. Kostelich, and Istvan Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1):112–126, 2007.

[64] Randall J. Hunt, John Doherty, and Matthew J. Tonkin. Are models too simple? arguments for increased parameterization. *Groundwater*, 45(3):254–262, 2007.

[65] K. Ide, P. Courtier, M. Ghil, and A. Lorenc. Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan*, 75 (1B):181–189, 1997.

[66] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, volume 63. Academic Press, 1970.

[67] Ian T. Jolliffe and David B. Stephenson. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, 2012.

[68] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *Automatic Control, IEEE Transactions on*, 45(3):477–482, 2000.

[69] R. E. Kalman et al. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[70] James L. Kaplan and James A. Yorke. Chaotic behavior of multidimensional difference equations. In *Functional Differential equations and approximation of fixed points*, pages 204–227. Springer, 1979.

[71] Duc H. Le, Alexandre A. Emerick, Albert C. Reynolds, et al. An adaptive ensemble smoother with multiple data assimilation for assisted history matching. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2015.

[72] François Le Gland, Valerie Monbet, and Vu-Duc Tran. Large sample asymptotics for the ensemble Kalman filter. Research Report RR-7014, INRIA, 2009.

[73] Erich Leo Lehmann and George Casella. *Theory of Point Estimation*, volume 31. Springer, 1998.

[74] Jing Lei and Peter Bickel. A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Monthly Weather Review*, 139(12):3964–3973, 2011.

[75] John M. Lewis, Sivaramakrishnan Lakshmivarahan, and Sudarshan Dhall. *Dynamic Data Assimilation: A Least Squares Approach*, volume 13. Cambridge University Press, 2006.

[76] A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

[77] Andrew C. Lorenc. The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, 2003.

[78] Andrew C. Lorenc, Neill E. Bowler, Adam M. Clayton, Stephen R. Pring, and David Fairbairn. Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Monthly Weather Review*, 143(1):212–229, 2014.

[79] Rolf J. Lorentzen and Geir Nævdal. An iterative ensemble Kalman filter. *Automatic Control, IEEE Transactions on*, 56(8):1990–1995, 2011.

[80] E. N. Lorenz. Atmospheric predictability experiments with a large numerical model. *Tellus*, 34(6):505–513, 1982.

[81] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[82] Edward N. Lorenz. Predictability: A problem partly solved. In *Proc. ECMWF Seminar on Predictability*, volume 1, pages 1–18, Reading, UK, 1996.

[83] Edward. N. Lorenz. Designing chaotic models. *Journal of the Atmospheric Sciences*, 62 (5):1574–1587, 2005.

[84] Edward N. Lorenz and Kerry A. Emanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55(3): 399–414, 1998.

[85] Dana Mackenzie. Ensemble Kalman filters bring weather models up to date. *Siam news*, 36(8):10–03, 2003.

[86] Jan Mandel, Lynn S. Bennethum, Jonathan D. Beezley, Janice L. Coen, Craig C. Douglas, Minjeong Kim, and Anthony Vodacek. A wildland fire model with data assimilation. *Mathematics and Computers in Simulation*, 79(3):584–606, 2008.

[87] Jan Mandel, Loren Cobb, and Jonathan D. Beezley. On the convergence of the ensemble Kalman filter. *Applications of Mathematics*, 56(6):533–541, 2011.

[88] Sammy Metref, E. Cosme, C. Snyder, and P. Brasseur. A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. *Nonlinear Processes in Geophysics*, 21(4):869–885, 2014.

[89] Herschel L. Mitchell, P. L. Houtekamer, and Gérard Pellerin. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Monthly Weather Review*, 130 (11):2791–2808, 2002.

[90] Robb J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[91] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[92] Geir Nævdal, Kristian Thulin, Hans Julius Skaug, Sigurd Ivar Aanonsen, et al. Quantifying Monte Carlo uncertainty in the ensemble Kalman filter. *SPE Journal*, 16(01):172–182, 2011.

[93] Shin'ya Nakano. Hybrid algorithm of ensemble transform and importance sampling for assimilation of non-Gaussian observations. *Tellus A*, 66, 2014.

[94] Shuli Niu, Yiqi Luo, Michael C. Dietze, Trevor F. Keenan, Zheng Shi, Jianwei Li, and F. Stuart Chapin III. The role of data assimilation in predictive ecology. *Ecosphere*, 5(5): 65, 2014.

[95] Met Office. Met office numerical weather prediction models, 2015. Online. Retreived from http://www.metoffice.gov.uk/research/modelling-systems/unified-model/weather-forecasting. December 2015.

[96] Dean S. Oliver. Minimization for conditional simulation: Relationship to optimal transport. *Journal of Computational Physics*, 265:1–15, 2014.

[97] Dean S. Oliver and Yan Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, 2011.

[98] Edward Ott, Brian R. Hunt, Istvan Szunyogh, Aleksey V. Zimin, Eric J. Kostelich, Matteo Corazza, Eugenia Kalnay, D. J. Patil, and James A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004.

[99] Luigi Palatella, Alberto Carrassi, and Anna Trevisan. Lyapunov vectors and assimilation in the unstable subspace: theory and applications. *Journal of Physics A: Mathematical and Theoretical*, 46(25):254020, 2013.

[100] Lies Peters, R. J. Arts, G. K. Brouwer, C. R. Geel, Stan Cullick, Rolf J. Lorentzen, Yan Chen, K. N. B. Dunlop, Femke C. Vossepoel, Rong Xu, et al. Results of the Brugge benchmark study for flooding optimization and history matching. *SPE Reservoir Evaluation & Engineering*, 13(3):391–405, 2010.

[101] Dinh Tuan Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207, 2001.

[102] James E. Potter and Robert Gottlieb Stern. Statistical filtering of space navigation measurements. *Massachusetts Institute of Technology, Experimental Astronomy Laboratory*, 1963.

[103] Patrick N. Raanes. *Improvements to Ensemble Methods for Data Assimilation in the Geosciences*. PhD thesis, University of Oxford, January 2016. https://ora.ox.ac.uk/objects/uuid:9f9961f0-6906-4147-a8a9-ca9f2d0e4a12.

[104] Sebastian Reich and Colin J. Cotter. Large scale inverse problems. computational methods and applications in the earth sciences: Ensemble filter techniques for intermittent data assimilation. *Radon Series on Computational and Applied Mathematics*, 13:91–134, 2013.

[105] Rolf H. Reichle, Dennis B. McLaughlin, and Dara Entekhabi. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114, 2002.

[106] Albert Coburn Reynolds, Alexandre Anoze Emerick, et al. History-matching production and seismic data in a real field case using the ensemble smoother with multiple data assimilation. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2013.

[107] William Sacher and Peter Bartello. Sampling errors in ensemble Kalman filtering. Part II: Application to a barotropic model. *Monthly Weather Review*, 137(5):1640–1654, 2009.

[108] Pavel Sakov and Laurent Bertino. Relation between two common localisation methods for the EnKF. *Computational Geosciences*, 15(2):225–237, 2011.

[109] Pavel Sakov and Peter R. Oke. A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2):361–371, 2008.

[110] Pavel Sakov and Peter R. Oke. Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, 136(3):1042–1053, 2008.

[111] Pavel Sakov, F. Counillon, L. Bertino, K. A. Lisæter, P. R. Oke, and A. Korablev. TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, 8(4):633, 2012.

[112] Pavel Sakov, Dean S. Oliver, and Laurent Bertino. An iterative EnKF for strongly nonlinear systems. *Monthly Weather Review*, 140(6):1988–2004, 2012.

[113] Barry Saltzman. Finite amplitude free convection as an initial value problem – I. *Journal of the Atmospheric Sciences*, 19(4):329–341, 1962.

[114] Simo Särkkä. Bayesian estimation of time-varying systems: Discrete-time systems, 2010. Unpublished. Retrieved from http://www.lce.hut.fi/~ssarkka/course_k2010/full_course_booklet.pdf. December 2015.

[115] Jeffrey Shaman and Alicia Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.

[116] S. Sherman. Non-mean-square error criteria. *Information Theory, IRE transactions on*, 4 (3):125–126, 1958.

[117] Emir H. Shuford Jr., Arthur Albert, and H. Edward Massengill. Admissible probability measurement procedures. *Psychometrika*, 31(2):125–145, 1966.

[118] Chris Snyder. Particle filters, the "optimal" proposal and high-dimensional systems. In *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, 2011.

[119] A. S. Stordal, H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. Vallès. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, 15(2):293–305, 2011.

[120] Michael K. Tippett, Jeffrey L. Anderson, Craig H. Bishop, Thomas M. Hamill, and Jeffrey S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131(7):1485–1490, 2003.

[121] Ricardo Todling and Stephen E. Cohn. Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Monthly Weather Review*, 122(11):2530–2557, 1994.

[122] Lloyd N. Trefethen. *Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations*. Unpublished, 1996. Retrieved from http://people.maths.ox.ac.uk/trefethen/pdetext.html. December 2015.

[123] Lloyd N. Trefethen and David Bau, III. *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[124] Peter Jan van Leeuwen. Comment on "Data assimilation using an ensemble Kalman filter technique". *Monthly Weather Review*, 127(6):1374–1377, 1999.

[125] Peter Jan van Leeuwen. Particle filtering in geophysical systems. *Monthly Weather Review*, 137(12):4089–4114, 2009.

[126] M. Verlaan and A. W. Heemink. Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, 11(5):349–368, 1997.

[127] Qing Wang, Sanjeev R. Kulkarni, and Sergio Verdú. Divergence estimation for multidimensional densities via-nearest-neighbor distances. *Information Theory, IEEE Transactions on*, 55(5):2392–2405, 2009.

[128] Xuguang Wang and Craig H. Bishop. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, 60(9): 1140–1158, 2003.

[129] Xuguang Wang and Ting Lei. GSI-based four dimensional ensemble-variational (4DEnsVar) data assimilation: formulation and single resolution experiments with real data for NCEP Global Forecast System. *Monthly Weather Review*, 142(9):3303–3325, 2014.

[130] Steven V. Weijs, Ronald Van Nooijen, and Nick Van De Giesen. Kullback-Leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition. *Monthly Weather Review*, 138(9):3387–3399, 2010.

[131] Jeffrey S. Whitaker and Thomas M. Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, 2002.

[132] Jeffrey S. Whitaker, Thomas M. Hamill, Xue Wei, Yucheng Song, and Zoltan Toth. Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, 136(2):463–482, 2008.

[133] C. K. Wikle and L. M. Berliner. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, 230(1-2):1–16, 2007.

[134] Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 100. Academic Press, 2011.

[135] C. Wunsch. *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications*. Cambridge University Press, 2006.

[136] Xiaozhen Xiong, Ionel Michael Navon, and Bahri Uzunoglu. A note on the particle filter with posterior Gaussian resampling. *Tellus A*, 58(4):456–460, 2006.