

# Chapter 9

## An Invitation to Meteorological Data Assimilation

Ágnes Bodó and Petra Csomós

**Abstract** The chapter introduces the basic data assimilation methods used in meteorological modelling. After briefly recalling the mathematical notions and tools needed, we present the optimal interpolation, the variational methods, and the Kálmán Filter techniques in one and more dimensions. In order to illustrate the use of the methods introduced, we present the results of numerical experiments done for simple models.

**Keywords** Data assimilation • Ensemble transform Kálmán filter • Kálmán filter • Optimal interpolation • Variational method

### 9.1 Introduction

The present chapter serves as an introduction to data assimilation methods used in meteorological modelling. Our aim is to present the mathematical derivation of the various methods and their applications to simple test models. Data assimilation literally means that one aims at combining information from several sources leading to a result which is better in some sense than the original data. Since we are after the best weather forecast possible, meteorological data assimilation aims at combining all the information being gathered about the present state of the atmosphere: observations, numerical prediction, climatological data, etc. The mathematical question is then how to combine all these data in order to get a result being nearest to the true state of atmosphere. For a detailed introduction in this field we refer to Kalnay [13] and Evensen [5] and the references therein.

---

Á. Bodó

Department of Applied Analysis and Computational Mathematics, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary  
e-mail: [bodoagi@cs.elte.hu](mailto:bodoagi@cs.elte.hu)

P. Csomós (✉)

Department of Applied Analysis and Computational Mathematics & MTA–ELTE Numerical Analysis and Large Networks Research Group, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary  
e-mail: [csomos@cs.elte.hu](mailto:csomos@cs.elte.hu)

For the sake of simplicity, we consider only two information sources being typical in meteorology: observations and numerical forecast obtained by a numerical prediction model. Both of them can be considered as vectors containing the values of the seven meteorological variables, that is, temperature, wind velocity in three directions, pressure, density, and relative humidity, at each point of a certain three-dimensional spatial mesh covering the whole atmosphere or its smaller region. Let  $x \in \mathbb{R}^n$  denote the vector of numerical forecast and  $y \in \mathbb{R}^m$  the vector of observations. In practice we usually have  $m \ll n$  (nowadays  $n \approx 10^7$ ,  $m \approx 10^5$ ). Hence, we are looking for that combination of  $x$  and  $y$ , called *analysis* in meteorology, which approximates best the true state of the atmosphere. Since the analysis at time  $t$  is the best approximation of the true state of the atmosphere at that time, it plays two roles. On one hand, it serves as the weather forecast for time  $t$ , being presented to the public. On the other hand, it is the best candidate for the initial value of a numerical weather prediction model computing the numerical forecast for the next time level, that is, for time  $t + \Delta t$  with some time step  $\Delta t > 0$ . Due to its latter role, it should be compatible with the model's variables, that is, it should be a vector of size  $n$ . Hence, we denote the analysis by  $x_a \in \mathbb{R}^n$ .

In the present chapter we introduce the basic data assimilation methods used in numerical weather prediction models, such as optimal interpolation, variational methods, and Kálmán Filter techniques. In Sect. 9.2 we summarise the mathematical tools needed later on. In Sects. 9.3 and 9.4 the optimal interpolation and the variational method are introduced in one and more dimensions, respectively. Section 9.5 serves as an introduction to the various Kálmán Filter techniques, and in Sect. 9.6 we present two test models and with the help of numerical experiments we compare the data assimilation methods. Section 9.7 serves as an outlook on various procedures used in nonlinear data assimilation.

## 9.2 Mathematical Background

In what follows we introduce the notions from mathematical statistics needed later on.

**Definition 9.1** Let  $\Omega \neq \emptyset$  denote the sample space being the space of all possible outcomes and let a  $\sigma$ -algebra  $\mathcal{E}$  denote the set of all events where each event is a set containing zero or more outcomes, that is, an event is a subset of the sample space  $\Omega$ . Then the function  $P : \mathcal{E} \rightarrow [0, 1]$  is called a probability function if it possesses the following properties:  $P(\Omega) = 1$  and it is countably additive, that is, for all  $A_n \in \mathcal{E}$ ,  $n = 1, \dots, N$  with  $A_i \cap A_j, i \neq j$  one has

$$P\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N P(A_n).$$

The triple  $(\Omega, \mathcal{E}, P)$  is called a probability space. The measurable function  $x : \Omega \rightarrow \mathbb{R}$  is called a real-valued random variable if  $\{\omega \in \Omega : x(\omega) \leq r\} \in \mathcal{E}$  for all  $r \in \mathbb{R}$  meaning that the set of events  $\omega$ , for which  $x(\omega) \leq r$  holds, is again an event, that is, we can talk about its probability. We will use the expression vector-valued random variable  $x \in \mathbb{R}^n$ , if the coordinate functions of  $x : \Omega \rightarrow \mathbb{R}^n$  are real-valued random variables.

By having a random variable at hand, one can define its statistical quantities which play an important role in data assimilation. From now on we suppose that  $\Omega$  is the finite union of the intervals  $I_j \subset \mathbb{R}$  for  $j = 1, \dots, n$  with  $n \in \mathbb{N}$ .

**Definition 9.2** Let  $(\Omega, \mathcal{E}, P)$  be a probability space and  $x = (x^{(1)}, \dots, x^{(n)}) : \Omega \rightarrow \mathbb{R}^n$  be a vector-valued random variable. We define the following quantities.

1. The cumulative distribution function  $F_x : \mathbb{R}^n \rightarrow \mathbb{R}$  of the vector-valued random variable  $x$  is defined as

$$F_x(\xi^{(1)}, \dots, \xi^{(n)}) = P(x^{(1)} < \xi^{(1)}, \dots, x^{(n)} < \xi^{(n)})$$

for all  $\xi = (\xi^{(1)}, \dots, \xi^{(n)}) \in \mathbb{R}^n$ . Two random variables are called identically distributed if they possess the same distribution function.

2. The probability density function  $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$  (if exists) of the vector-valued random variable  $x$  is the function which fulfills

$$F_x(\xi) = \int_{-\infty}^{\xi^{(1)}} \dots \int_{-\infty}^{\xi^{(n)}} f_x(t^{(1)}, \dots, t^{(n)}) dt^{(n)} \dots dt^{(1)}$$

for any  $\xi = (\xi^{(1)}, \dots, \xi^{(n)}) \in \mathbb{R}^n$ .

In what follows we define the most important notions characterising a random variable. Its expectation is intuitively the long-run average value of repetitions of the experiment it represents. The variance measures how far a set of numbers is spread out, and the covariance measures how much two random variables depend on each other.

**Definition 9.3**

1. The expectation  $\mathbb{E}$  of the vector-valued random variable  $x \in \mathbb{R}^n$  is defined as

$$\mathbb{E}(x) := (\mathbb{E}(x^{(1)}), \dots, \mathbb{E}(x^{(n)})) \quad \text{with}$$

$$\mathbb{E}(x^{(i)}) := \int_{I_j} t f_{x^{(i)}}(t) dt \quad \text{for all } i = 1, \dots, n$$

(if exists). Let  $X = (x_1, \dots, x_k) \in \mathbb{R}^{n \times k}$  be the matrix containing the  $k$  pieces of vector-valued random variables  $x_1, \dots, x_k \in \mathbb{R}^n$  in its columns. Then the notation  $\mathbb{E}(X)$  means  $(\mathbb{E}(X))_{ij} := \mathbb{E}(x_j^{(i)})$  for all  $i = 1, \dots, n$  and  $j = 1, \dots, k$ , i.e., we take the expectation elementwise.

2. Let  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  be vector-valued random variables. Their covariance is defined as

$$\text{cov}(x, y) := \mathbb{E}((x - \mathbb{E}(x))(y - \mathbb{E}(y))^\top) \in \mathbb{R}^{n \times m},$$

where  $\top$  denotes the transposition, that is,  $xy^\top \in \mathbb{R}^{n \times m}$  is the dyadic product of the vectors  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$ . We note that  $\mathbb{V}(x) := \text{cov}(x, x) \in \mathbb{R}^{n \times n}$  is called the variation of the random variable  $x$ . Since we have

$$(\mathbb{V}(x))_{ij} = \text{cov}(x^{(i)}, x^{(j)}) \quad \text{for all } i, j = 1, \dots, n,$$

that is, the entries of  $\mathbb{V}(x)$  are the covariances of the elements of  $x$ ,  $\mathbb{V}(x)$  is also called the *covariance matrix* of the random variable  $x$ .

The following properties will be used frequently.

1. The expectation  $\mathbb{E}$  is a linear function.
2. The matrix  $\mathbb{V}(x)$  is symmetric and positive semidefinite for all random variables  $x$  (whenever it exists).

One often investigates the jointly behaviour of two random variables but the knowledge of their distribution functions is usually not sufficient. Therefore, we need to define the joint distribution function of two random variables.

#### Definition 9.4

1. The joint distribution function  $F_{x,y} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  of the vector-valued random variables  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  is defined as

$$F_{x,y}(\xi, \eta) := P(x^{(1)} < \xi^{(1)}, \dots, x^{(n)} < \xi^{(n)}, y^{(1)} < \eta^{(1)}, \dots, y^{(m)} < \eta^{(m)})$$

for all  $\xi \in \mathbb{R}^n$ ,  $\eta \in \mathbb{R}^m$ .

2. The vector-valued random variables  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  are called independent if

$$F_{x,y}(\xi, \eta) = F_x(\xi)F_y(\eta) \quad \text{for all } \xi \in \mathbb{R}^n, \eta \in \mathbb{R}^m.$$

3. The vector-valued random variables  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  are called uncorrelated if

$$\text{cov}(x, y) = 0 \in \mathbb{R}^{n \times m}.$$

We note that if two random variables are independent, then they are uncorrelated as well.

In some cases the random variable  $x$  is unknown and it is approximated by another random variable  $\tilde{x}$  called an *estimator* of  $x$  with the following properties.

**Definition 9.5** Let  $x$  be a vector-valued random variable and  $\tilde{x}$  one of its estimators.

1. The estimator  $\tilde{x}$  is called unbiased if  $\mathbb{E}(\tilde{x}) = \mathbb{E}(x)$ .
2. The estimator  $\tilde{x}$  is called optimal if the trace  $\text{tr} \mathbb{E}((\tilde{x} - x)(\tilde{x} - x)^\top)$  is minimal among all possible estimators.

We note that for a real-valued random variable  $x \in \mathbb{R}$ , the optimal estimator  $\tilde{x}$  possesses the minimal variance  $\mathbb{V}(\tilde{x})$ .

The sample (or empirical) mean and the sample covariance are statistics computed from one or more random variables. These will be important later on when the data assimilation methods are introduced.

**Definition 9.6**

1. The sample mean  $\mathbb{E}_{x_1 \dots x_k}$  of the vector-valued random variables  $x_1, \dots, x_k \in \mathbb{R}^n$  is defined as

$$\mathbb{E}_{x_1 \dots x_k} := \frac{1}{k} \sum_{j=1}^k x_j \in \mathbb{R}^n.$$

2. The sample covariance matrix  $\mathbb{V}$  of the vector-valued random variables  $x_1, \dots, x_k \in \mathbb{R}^n$  is defined as

$$\mathbb{V} := \frac{1}{k-1} \sum_{j=1}^k (x_j - \mathbb{E}_{x_1 \dots x_k})(x_j - \mathbb{E}_{x_1 \dots x_k})^\top \in \mathbb{R}^{n \times n}.$$

When introducing the basic data assimilation methods used in meteorology, we will need the following result presented e.g. in Johnson and Wichern [11]

**Proposition 9.1** *Let  $x$  be a vector-valued random variable and  $x_1, \dots, x_k$  its mutually independent and identically distributed estimators. Then the following assertions hold.*

1. *The sample mean  $\mathbb{E}_{x_1 \dots x_k}$  is an unbiased estimator of the expectation  $\mathbb{E}(x)$ .*
2. *The sample covariance matrix  $\mathbb{V}$  is an unbiased estimator of the covariance matrix  $\mathbb{V}(x)$ .*

### 9.3 Optimal Interpolation and Variational Method in One Dimension

This section is devoted to the introduction of the basic data assimilation methods when applied to one-dimensional problems, for example estimating the unknown true temperature  $x_t \in \mathbb{R}$  at a point. To do so we make two measurements, that is, we take the real-valued random variables  $x, y$  and look for their (in some sense best) combination, that is, the real-valued estimator  $x_a$ . In meteorological data assimilation, we always suppose the following.

*Assumptions 9.1* Let  $x$  and  $y$  be real-valued random variables and let  $x_a$  be an estimator of the constant true state  $x_t \in \mathbb{R}$ . We suppose the following.

1. The estimator  $x_a$  is the linear combination of  $x$  and  $y$ , that is,  $x_a = \alpha_1 x + \alpha_2 y$  for some constants  $\alpha_1, \alpha_2 \in \mathbb{R}$ .
2. The estimator  $x_a$  is unbiased, that is,  $\mathbb{E}(x_a) = \mathbb{E}(x_t) = x_t$ .
3. The estimator  $x_a$  is optimal, that is, the  $\mathbb{E}((x_a - x_t)^2)$  is minimal.
4. The measurements  $x$  and  $y$  are unbiased, that is,  $\mathbb{E}(x) = \mathbb{E}(y) = x_t$ .
5. The measurements  $x$  and  $y$  are uncorrelated, that is,  $\text{cov}(x, y) = 0$ .
6. The values of the variances  $\mathbb{V}(x)$  and  $\mathbb{V}(y)$  are given.

We present first the result of the *optimal interpolation* being a least mean square estimate.

**Theorem 9.1** *Under Assumptions 9.1, the estimator  $x_a$  has the form*

$$x_a = x + \frac{\mathbb{V}(x)}{\mathbb{V}(x) + \mathbb{V}(y)}(y - x). \quad (9.1)$$

*Proof* Instead of just checking Assumptions 9.1, we present a constructive proof. From the linearity of the expectation  $\mathbb{E}$  and the estimator  $x_a$  in  $x$  and  $y$ , it follows that for some  $\alpha_1, \alpha_2 \in \mathbb{R}$  the following identity holds

$$\mathbb{E}(x_a) = \mathbb{E}(\alpha_1 x + \alpha_2 y) = \alpha_1 \mathbb{E}(x) + \alpha_2 \mathbb{E}(y) = (\alpha_1 + \alpha_2)x_t.$$

Since the estimator  $x_a$  is unbiased, we have that  $\alpha_1 + \alpha_2 = 1$ , hence, we obtain the form

$$x_a = (1 - \alpha)x + \alpha y = x + \alpha(y - x)$$

for some constant  $\alpha \in \mathbb{R}$ . In order to minimize the variance  $\mathbb{V}(x_a)$ , we note first that Definition 9.3 implies

$$\mathbb{V}(x_a) = \mathbb{E}((x_a - \mathbb{E}(x_a))^2) = \mathbb{E}((x_a - x_t)^2),$$

and similarly for  $\mathbb{V}(x) = \mathbb{E}(\varepsilon_x^2)$  and  $\mathbb{V}(y) = \mathbb{E}(\varepsilon_y^2)$ , where  $\varepsilon_x := x - x_t$  and  $\varepsilon_y := y - x_t$  denote the errors of the measurements  $x$  and  $y$ , respectively, being real-valued random variables as well. Hence, we have the identity

$$\begin{aligned} \mathbb{V}(x_a) &= \mathbb{E}((x_a - x_t)^2) = \mathbb{E}((x + \alpha(y - x) - x_t)^2) \\ &= \mathbb{E}((x_t + \varepsilon_x + \alpha(x_t + \varepsilon_y - x_t - \varepsilon_x) - x_t)^2) = \mathbb{E}(\varepsilon_x + \alpha(\varepsilon_y - \varepsilon_x))^2 \\ &= \mathbb{E}((1 - \alpha)^2 \varepsilon_x^2 + \alpha^2 \varepsilon_y^2 - 2\alpha(1 - \alpha)\varepsilon_x \varepsilon_y) \\ &= (1 - \alpha)^2 \mathbb{E}(\varepsilon_x^2) + \alpha^2 \mathbb{E}(\varepsilon_y^2) - 2\alpha(1 - \alpha)\mathbb{E}(\varepsilon_x \varepsilon_y). \end{aligned}$$

Since the measurements  $x$  and  $y$  are unbiased and uncorrelated, we have

$$0 = \text{cov}(x, y) = \mathbb{E}((x - \mathbb{E}(x))(y - \mathbb{E}(y))) = \mathbb{E}((x - x_t)(y - x_t)) = \mathbb{E}(\varepsilon_x \varepsilon_y)$$

by Definitions 9.4 and 9.5. This implies the result

$$\mathbb{V}(x_a) = (1 - \alpha)^2 \mathbb{V}(x) + \alpha^2 \mathbb{V}(y),$$

which is minimal if its derivative with respect to the parameter  $\alpha$  vanishes:

$$0 = \frac{d}{d\alpha} \mathbb{V}(x_a) = \frac{d}{d\alpha} (1 - \alpha)^2 \mathbb{V}(x) + \alpha^2 \mathbb{V}(y) = -2(1 - \alpha) \mathbb{V}(x) + 2\alpha \mathbb{V}(y)$$

which implies

$$\alpha = \frac{\mathbb{V}(x)}{\mathbb{V}(x) + \mathbb{V}(y)}$$

completing the proof.

We note that formula (9.1) contains all the information given: the measurements  $x$ ,  $y$  and their variances  $\mathbb{V}(x)$ ,  $\mathbb{V}(y)$ . In cases when the formula above is not feasible to compute (e.g. in more dimensions presented later on), usually a statistical cost function is minimised. As before, let  $x$  and  $y$  be estimators for the true state  $x_t$  with probability density functions  $f_x$  and  $f_y$ , respectively. The analysis  $x_a$  is then derived by maximising the maximum likelihood function  $L : z \mapsto f_x(z)f_y(z)$  for the real-valued random variable  $z$ . Such methods are called *variational methods*.

*Assumptions 9.2* We suppose that the real-valued random variables  $x$  and  $y$  are equally distributed and are of normal distribution with given variances  $\mathbb{V}(x)$  and  $\mathbb{V}(y)$ .

**Theorem 9.2** *Under Assumptions 9.1 and 9.2, the solution of the maximum likelihood method leads to the same solution (9.3) as the optimal interpolation.*

*Proof* Since  $x$  and  $y$  are of normal distribution, the maximum likelihood function has the following form for any  $\xi \in \mathbb{R}$ :

$$\begin{aligned} L(\xi) &= f_x(\xi)f_y(\xi) \\ &= \frac{1}{\sqrt{2\pi\mathbb{V}(x)}} e^{-\frac{1}{2} \frac{(x-\xi)^2}{\mathbb{V}(x)}} \frac{1}{\sqrt{2\pi\mathbb{V}(y)}} e^{-\frac{1}{2} \frac{(y-\xi)^2}{\mathbb{V}(y)}} \\ &= \frac{1}{2\pi\sqrt{\mathbb{V}(x)\mathbb{V}(y)}} e^{-\frac{1}{2} \frac{(x-\xi)^2}{\mathbb{V}(x)} - \frac{1}{2} \frac{(y-\xi)^2}{\mathbb{V}(y)}}. \end{aligned}$$

The function  $L$  is maximal if the absolute value of the exponent

$$J(\xi) := \frac{1}{2} \frac{(x - \xi)^2}{\mathbb{V}(x)} + \frac{1}{2} \frac{(y - \xi)^2}{\mathbb{V}(y)} \quad (9.2)$$

is minimal, that is, its derivative with respect to  $\xi$  vanishes. Hence, we obtain

$$x_a = \frac{\mathbb{V}(y)}{\mathbb{V}(x) + \mathbb{V}(y)}x + \frac{\mathbb{V}(x)}{\mathbb{V}(x) + \mathbb{V}(y)}y = x + \frac{\mathbb{V}(x)}{\mathbb{V}(x) + \mathbb{V}(y)}(y - x)$$

which completes the proof.

The function  $J$  defined by formula (9.2) is called *cost function* in meteorological data assimilation. We note that it is a quadratic function.

## 9.4 Optimal Interpolation and Variational Method in More Dimensions

In the previous section we have seen how the optimal interpolation and the variational method work in one dimension. Since in meteorology one aims at estimating the true state of the whole atmosphere, or at least the true values of the meteorological variables in the spatial grid points, the measurements  $x$  and  $y$  are (quite long) vectors. Hence, in this section we seek the best combination of the model's forecast  $x \in \mathbb{R}^n$  and the observations  $y \in \mathbb{R}^m$  by supposing the same as in Assumption 9.1 in the appropriate form. To do so, we introduce first the operator  $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , called *observation operator*, which maps the forecast vector  $x$  onto the grid of the observations' vector  $y$ .

*Assumptions 9.3* Let  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  be vector-valued random variables with  $m \leq n$  and  $x_a \in \mathbb{R}^n$  be an estimator of the constant true state  $x_t \in \mathbb{R}^n$ .

1. The estimator  $x_a$  is a linear function of  $x$  and  $y$ , that is, there exists a matrix  $K \in \mathbb{R}^{n \times m}$  such that

$$x_a = x + K(y - \mathcal{H}(x)). \quad (9.3)$$

2. The estimator  $x_a$  and the data  $x, y$  are unbiased, that is,  $\mathbb{E}(x_a) = \mathbb{E}(x) = x_t$  and  $\mathbb{E}(y) = \mathcal{H}(x_t)$ , and are of normal distribution.
3. The estimator  $x_a$  is optimal in the sense of Definition 9.5.
4. The data are uncorrelated, that is,  $\text{cov}(x, y) = 0$ .
5. The values of the variances  $\mathbb{V}(x)$  and  $\mathbb{V}(y)$  are given.
6. The observation operator  $\mathcal{H} = H \in \mathbb{R}^{m \times n}$  is linear.

*Remark 9.1* Let  $\varepsilon_x := x - x_t \in \mathbb{R}^n$  and  $\varepsilon_y := y - \mathcal{H}(x_t) \in \mathbb{R}^m$  denote the errors of  $x$  and  $y$ , respectively, being vector-valued random variables as well. Since the data  $x$



and  $y$  are unbiased, the variances have the form

$$\mathbb{V}(x) = \mathbb{E}((x - \mathbb{E}(x))(x - \mathbb{E}(x))^{\top}) = \mathbb{E}((x - x_t)(x - x_t)^{\top}) = \mathbb{E}(\varepsilon_x \varepsilon_x^{\top})$$

and similarly for  $\mathbb{V}(y) = \mathbb{E}(\varepsilon_y \varepsilon_y^{\top})$ . Hence, they are usually called *error covariance matrices*. Moreover, since the data  $x$  and  $y$  are uncorrelated, we have

$$0 = \text{cov}(x, y) = \mathbb{E}((x - \mathbb{E}(x))(y - \mathbb{E}(y))^{\top}) = \mathbb{E}((x - x_t)(y - x_t)^{\top}) = \mathbb{E}(\varepsilon_x \varepsilon_y^{\top})$$

and similarly  $\mathbb{E}(\varepsilon_y \varepsilon_x^{\top}) = 0$ .

The next question is how to choose the matrix  $K$ , called Kálmán gain matrix, in order to obtain an optimal estimator  $x_a$ .

**Theorem 9.3** *Under Assumptions 9.3, the Kálmán gain matrix  $K$  in formula (9.3) has the form*

$$K = \mathbb{V}(x)H^{\top}(\mathbb{V}(y) + H\mathbb{V}(x)H^{\top})^{-1}. \quad (9.4)$$

*Proof* Since the vector-valued random variables  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  are of normal distribution, their probability density functions have the following form for any  $\xi \in \mathbb{R}^n$ :

$$f_x(\xi) = \frac{1}{\sqrt{2\pi|\mathbb{V}(x)|}} e^{-\frac{1}{2}(x-\xi)^{\top}\mathbb{V}(x)^{-1}(x-\xi)},$$

$$f_y(\xi) = \frac{1}{\sqrt{2\pi|\mathbb{V}(y)|}} e^{-\frac{1}{2}(y-H\xi)^{\top}\mathbb{V}(y)^{-1}(y-H\xi)},$$

where  $|\cdot|$  denotes the determinant of the corresponding matrix. Thus, the maximum likelihood function reads as

$$L(\xi) := f_x(\xi)f_y(\xi)$$

$$= \frac{1}{\sqrt{2\pi|\mathbb{V}(x)||\mathbb{V}(y)|}} e^{-\frac{1}{2}(x-\xi)^{\top}\mathbb{V}(x)^{-1}(x-\xi) - \frac{1}{2}(y-H\xi)^{\top}\mathbb{V}(y)^{-1}(y-H\xi)}.$$

The function  $L$  is maximal if the absolute value of the exponent

$$J(\xi) := \frac{1}{2}(x - \xi)^{\top}\mathbb{V}(x)^{-1}(x - \xi) + \frac{1}{2}(y - H\xi)^{\top}\mathbb{V}(y)^{-1}(y - H\xi) \quad (9.5)$$

is minimal, that is, if its derivative vanishes:

$$\frac{d}{d\xi}J(\xi) = \mathbb{V}(x)^{-1}(x - \xi) + H^{\top}\mathbb{V}(y)^{-1}(y - H\xi) = 0.$$

Hence, we obtain

$$x_a = x + (\mathbb{V}(x)^{-1} + H^{\top}\mathbb{V}(y)^{-1}H)^{-1}H^{\top}\mathbb{V}(y)^{-1}(y - Hx).$$

From the identities

$$(\mathbb{V}(x)^{-1} + H^\top \mathbb{V}(y)^{-1} H)^{-1} H^\top \mathbb{V}(y)^{-1} = \mathbb{V}(x) H^\top (\mathbb{V}(y) + H \mathbb{V}(x) H^\top)^{-1}$$

and

$$H^\top \mathbb{V}(y)^{-1} (H \mathbb{V}(x) H^\top + \mathbb{V}(y)) = (\mathbb{V}(x)^{-1} + H^\top \mathbb{V}(y)^{-1} H) \mathbb{V}(x) H^\top,$$

we have that

$$(\mathbb{V}(x)^{-1} + H^\top \mathbb{V}(y)^{-1} H)^{-1} H^\top \mathbb{V}(y)^{-1} = \mathbb{V}(x) H^\top (\mathbb{V}(y) + H \mathbb{V}(x) H^\top)^{-1},$$

which completes the proof.

**Theorem 9.4** *Under Assumptions 9.3, for any matrix  $K \in \mathbb{R}^{n \times m}$ , the analysis error covariance matrix  $\mathbb{V}(x_a)$  is given by*

$$\mathbb{V}(x_a) = (\mathbf{I} - KH) \mathbb{V}(x) (\mathbf{I} - KH)^\top + K \mathbb{V}(y) K^\top. \quad (9.6)$$

*If the Kálmán gain matrix  $K$  has the special form defined in (9.4), the expression becomes*

$$\mathbb{V}(x_a) = (\mathbf{I} - KH) \mathbb{V}(x). \quad (9.7)$$

*Proof* From formula (9.3), we obtain for the errors that

$$\begin{aligned} \varepsilon_a - \varepsilon_x &= x_a - x_t - x + x_t = K(y - Hx) \\ &= K(\varepsilon_y + Hx_t - Hx) = K(\varepsilon_y + H(x_t - x)) \\ &= K(\varepsilon_y - H\varepsilon_x), \end{aligned}$$

which implies

$$\varepsilon_a = \varepsilon_x + K\varepsilon_y - KH\varepsilon_x = (\mathbf{I} - KH)\varepsilon_x + K\varepsilon_y.$$

Hence, the error covariance matrix  $\mathbb{V}(x_a)$  of the analysis can be expressed as

$$\begin{aligned} \mathbb{V}(x_a) &= \text{cov}(\varepsilon_a) = \mathbb{E}((\varepsilon_a - \mathbb{E}(\varepsilon_a))(\varepsilon_a - \mathbb{E}(\varepsilon_a))^\top) \\ &= \mathbb{E}(((\mathbf{I} - KH)\varepsilon_x + K\varepsilon_y)((\mathbf{I} - KH)\varepsilon_x + K\varepsilon_y)^\top) \\ &= (\mathbf{I} - KH) \mathbb{E}(\varepsilon_x \varepsilon_x^\top) (\mathbf{I} - KH)^\top + (\mathbf{I} - KH) \mathbb{E}(\varepsilon_x \varepsilon_y^\top) K^\top \\ &\quad + K \mathbb{E}(\varepsilon_y \varepsilon_x^\top) (\mathbf{I} - KH)^\top + K \mathbb{E}(\varepsilon_y \varepsilon_y^\top) K^\top. \end{aligned}$$

Remark 9.1 further implies

$$\begin{aligned}
\mathbb{V}(x_a) &= (\mathbf{I} - KH)\mathbb{V}(x)(\mathbf{I} - KH)^\top + K\mathbb{V}(y)K^\top \\
&= \mathbb{V}(x) - \mathbb{V}(x)H^\top K^\top - KH\mathbb{V}(x) + KH\mathbb{V}(x)H^\top K^\top + K\mathbb{V}(y)K^\top \\
&= \mathbb{V}(x) - KH\mathbb{V}(x) - KH\mathbb{V}(x) + KH\mathbb{V}(x)H^\top K^\top + K\mathbb{V}(y)K^\top \quad (9.8) \\
&= (\mathbf{I} - KH)\mathbb{V}(x) + \Delta
\end{aligned}$$

with  $\Delta := -KH\mathbb{V}(x) + KH\mathbb{V}(x)H^\top K^\top + K\mathbb{V}(y)K^\top$ . We only have to prove now that  $\Delta = 0$  hold. From formula (9.4) we have

$$K = \mathbb{V}(x)H^\top (H\mathbb{V}(x)H^\top + \mathbb{V}(y))^{-1} = \mathbb{V}(x)^\top H^\top ((H\mathbb{V}(x)H^\top + \mathbb{V}(y))^{-1})^\top,$$

which implies

$$K^\top = (H\mathbb{V}(x)H^\top + \mathbb{V}(y))^{-1}H\mathbb{V}(x)$$

and

$$H\mathbb{V}(x) = (H\mathbb{V}(x)H^\top + \mathbb{V}(y))K^\top = H\mathbb{V}(x)H^\top K^\top + \mathbb{V}(y)K^\top.$$

Then from the identity

$$KH\mathbb{V}(x) = KH\mathbb{V}(x)H^\top K^\top + K\mathbb{V}(y)K^\top$$

we finally conclude the proof with

$$0 = -KH\mathbb{V}(x) + KH\mathbb{V}(x)H^\top K^\top + K\mathbb{V}(y)K^\top = \Delta.$$

By inserting the form (9.4) of  $K$  into formula (9.6), one obtains the identity (9.7) which was to prove.

Besides the specific form (9.3) of the analysis  $x_a$ , we will show its optimality as well. To do so we will need the following technical Lemma.

**Lemma 9.1** *Let  $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$  be a continuously differentiable function, and  $A \in \mathbb{R}^{m \times a}$ ,  $B \in \mathbb{R}^{m \times m}$  be arbitrary fixed matrices for some  $m, a \in \mathbb{N}$ . Then the following holds for its derivative for any  $K \in \mathbb{R}^{n \times m}$ .*

1. For  $g(K) = \text{tr} KA$  one has  $\frac{\partial g}{\partial K} = A^\top$ .
2. For  $g(K) = \text{tr} KBK^\top$  one has  $\frac{\partial g}{\partial K} = KB^\top + KB$ .

*Proof* For the whole proof we refer to Schönemann [16]. For the conviction of the reader we note that since the function  $g$  is continuously differentiable with respect to  $K = (K_{jk})_{j,k} \in \mathbb{R}^{n \times m}$  ( $j = 1, \dots, n$  and  $k = 1, \dots, m$ ), its derivative can be

expressed as

$$\frac{\partial g}{\partial K} = \begin{pmatrix} \frac{\partial g}{\partial K_{11}} & \cdots & \frac{\partial g}{\partial K_{1m}} \\ \vdots & \cdots & \vdots \\ \frac{\partial g}{\partial K_{n1}} & \cdots & \frac{\partial g}{\partial K_{nm}} \end{pmatrix}. \quad (9.9)$$

We can now state the main result of this section.

**Theorem 9.5** *Under Assumptions 9.3, the analysis  $x_a$ , given by the formula (9.3) with the Kálmán gain matrix (9.4), is optimal in the sense of Definition 9.5.*

*Proof* The analysis  $x_a$  is optimal if the trace of the matrix

$$\mathbb{E}((x_a - x_t)(x_a - x_t)^\top)$$

is minimal. Since  $x_a$  is an unbiased estimate, this is equivalent to the minimisation of  $\text{tr } \mathbb{V}(x_a)$ . Formula (9.7) in Theorem 9.4 implies that

$$\mathbb{V}(x_a) = (\mathbf{I} - KH)\mathbb{V}(x),$$

therefore, its trace is given in (9.8) as

$$\text{tr } \mathbb{V}(x_a) = \text{tr } \mathbb{V}(x) + \text{tr } KH\mathbb{V}(x)H^\top K^\top - 2 \text{tr } KH\mathbb{V}(x) + \text{tr } K\mathbb{V}(y)K^\top.$$

Since the expression above is minimal if its derivative with respect to the matrix  $K$  vanishes, we need to compute  $\frac{\partial \text{tr } \mathbb{V}(x_a)}{\partial K}$ . We use Lemma 9.1/2 first for the matrix  $B := H\mathbb{V}(x)H^\top$ , and obtain

$$\begin{aligned} \frac{\partial \text{tr } KH\mathbb{V}(x)H^\top K^\top}{\partial K} &= K(H\mathbb{V}(x)H^\top)^\top + KH\mathbb{V}(x)H^\top \\ &= (H\mathbb{V}(x)H^\top K^\top)^\top + KH\mathbb{V}(x)H^\top. \end{aligned}$$

Similarly, for the choice  $B := \mathbb{V}(y)$ , Lemma 9.1/2 implies

$$\frac{\partial \text{tr } K\mathbb{V}(y)K^\top}{\partial K} = (\mathbb{V}(y)K^\top)^\top + K\mathbb{V}(y).$$

Finally, for the matrix  $A := H\mathbb{V}(x)$ , Lemma 9.1/1 implies

$$\frac{\partial \text{tr } KH\mathbb{V}(x)}{\partial K} = \mathbb{V}(x)^\top H^\top.$$

So the derivative of  $\text{tr } \mathbb{V}(x_a)$  is given by

$$\begin{aligned} \frac{\partial \text{tr } \mathbb{V}(x_a)}{\partial K} &= (H\mathbb{V}(x)H^\top K^\top)^\top + KH\mathbb{V}(x)H^\top + K\mathbb{V}(y) \\ &\quad + (\mathbb{V}(y)K^\top)^\top - 2\mathbb{V}(x)^\top H^\top \\ &= 2KH\mathbb{V}(x)H^\top + 2K\mathbb{V}(y) - 2\mathbb{V}(x)H^\top, \end{aligned}$$

which is zero if and only if

$$K = \mathbb{V}(x)H^\top (H\mathbb{V}(x)H^\top + \mathbb{V}(y))^{-1}$$

holds, which completes the proof.

Since formula (9.3) together with formula (9.4) is the best linear unbiased estimate, this method is called BLUE from the initials. We note that if the observation operator  $\mathcal{H}$  is nonlinear but linearisable around  $x_a$  (i.e. there exists  $H \in \mathbb{R}^{m \times n}$  being the first derivative of  $\mathcal{H}$  at  $x_a$ ) then formula BLUE reads as

$$x_a = x + K(y - \mathcal{H}(x))$$

and together with (9.4) yield an approximatively optimal estimate to  $x_a$ , being however the only analysis which is possible to compute in practice in this way.

## 9.5 Kálmán Filter Techniques

In the previous section we presented the two basic data assimilation methods used in meteorology. From formulae BLUE (9.3) and (9.4) one can see how important role the error covariance matrices  $\mathbb{V}(x)$  and  $\mathbb{V}(y)$  play. Their computation, however, is a challenging task in practice. As a first attempt, they are usually supposed to be constant in time, however, in reality they may strongly depend on the weather situation. In the present study we focus on  $\mathbb{V}(x)$  and assume that  $\mathbb{V}(y)$  is constant in time. This can be supposed, because the spatial propagation of that part of  $x_a$  which causes the changes to  $x$  (called analysis increment) is based on  $\mathbb{V}(x)$  solely. We present now a procedure due to Kálmán [12] to update the value of the error covariance matrix  $\mathbb{V}(x)$  of the model's forecast in each time step. To do so, we need to introduce a model operator. Since the model operator describes time-dependent processes, we denote it by  $\mathcal{M}_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$  acting between the time levels  $i$  and  $i + 1$ . It contains the spatially and temporally discretised version of the partial differential equations describing the atmosphere's dynamics and the physical parametrisations. By applying the BLUE data assimilation method, the numerical forecast  $x^{\{i+1\}}$  at

time level  $i + 1$  is then obtained from the analysis  $x_a^{\{i\}}$  valid at the  $i$ th time level as

$$\begin{aligned} x^{\{i+1\}} &= \mathcal{M}_i(x_a^{\{i\}}), \\ x_a^{\{i+1\}} &= x^{\{i+1\}} + K_{i+1}(y^{\{i+1\}} - \mathcal{H}(x^{\{i+1\}})) \end{aligned}$$

for all  $i \in \mathbb{N}$ , where  $x_a^{(0)}$  is a given initial value (e.g. from another numerical weather prediction model) and the Kálmán gain matrix is defined by formula (9.4), that is,

$$K_i = \mathbb{V}(x^{\{i\}})H_i^\top (\mathbb{V}(y^{\{i\}}) + H_i\mathbb{V}(x^{\{i\}})H_i^\top)^{-1}$$

for all  $i \in \mathbb{N}$ , where  $H_i$  denotes the linear observation operator. We note that if one takes the derivative of the nonlinear observation operator  $\mathcal{H}$  at  $x_a^{\{i\}}$  instead of  $\mathcal{H}$  itself, the method described above only leads to an approximation to  $x_a^{\{i+1\}}$  at the  $i$ th time level. We denote the model's error at time level  $i$  by  $\varepsilon_{\mathcal{M}}^{\{i\}}$  and its error covariance matrix by  $\mathbb{V}(\mathcal{M}_i(x_t^{\{i\}})) := \mathbb{E}(\varepsilon_{\mathcal{M}}^{\{i\}}(\varepsilon_{\mathcal{M}}^{\{i\}})^\top)$ . As before, we suppose that the various errors are uncorrelated.

*Assumptions 9.4* We suppose that the model's error and the error of the other data  $x$  and  $y$  are uncorrelated. We further suppose that the model operator and the observation operator are linear for all  $i \in \mathbb{N}$ , that is,  $\mathcal{M}_i = M_i \in \mathbb{R}^{n \times n}$  and  $\mathcal{H}_i = H_i \in \mathbb{R}^{m \times n}$ .

In what follows we present the *Kálmán Filter* method for updating the error covariance matrix.

**Theorem 9.6** *Under Assumptions 9.4, the update of the forecast's error covariance matrix reads as*

$$\mathbb{V}(x^{\{i+1\}}) = M_i\mathbb{V}(x_a^{\{i\}})M_i^\top + \mathbb{V}(M_ix_t^{\{i\}}) \quad \text{for all } i \in \mathbb{N}. \quad (9.10)$$

*Proof* We consider the following two relations:

$$\begin{cases} x^{\{i+1\}} = M_ix_a^{\{i\}}, \\ x_t^{\{i+1\}} = M_ix_t^{\{i\}} - \varepsilon_{\mathcal{M}}^{\{i\}}, \end{cases}$$

where  $x_t^{\{i\}}$  denotes the (unknown) true state at the  $i$ th time level. By subtracting the second equation from the first, one obtains

$$x^{\{i+1\}} - x_t^{\{i+1\}} = M_ix_a^{\{i\}} - M_ix_t^{\{i\}} + \varepsilon_{\mathcal{M}}^{\{i\}}.$$

Due to the linearity of the observation and the model operators, we can write

$$x^{\{i+1\}} - x_t^{\{i+1\}} = M_i(x_a^{\{i\}} - x_t^{\{i\}}) + \varepsilon_{\mathcal{M}}^{\{i\}}.$$

Since  $x_a^{\{i\}} - x_t^{(i)} = \varepsilon_a^{\{i\}}$  for all  $i \in \mathbb{N}$ , we have

$$\begin{aligned}
 \mathbb{V}(x^{\{i+1\}}) &= \mathbb{E}(\varepsilon_x^{\{i+1\}}(\varepsilon_x^{\{i+1\}})^\top) \\
 &= \mathbb{E}((M_i(x_a^{\{i\}} - x_t^{(i)} + \varepsilon_{\mathcal{M}}^{\{i\}})(M_i(x_a^{\{i\}} - x_t^{(i)} + \varepsilon_{\mathcal{M}}^{\{i\}})^\top) \\
 &= \mathbb{E}((M_i\varepsilon_a^{\{i\}} + \varepsilon_{\mathcal{M}}^{\{i\}})(M_i\varepsilon_a^{\{i\}} + \varepsilon_{\mathcal{M}}^{\{i\}})^\top) \\
 &= M_i\mathbb{E}(\varepsilon_a^{\{i\}}(\varepsilon_a^{\{i\}})^\top)M_i^\top + \mathbb{E}(\varepsilon_{\mathcal{M}}\varepsilon_{\mathcal{M}}^\top) \\
 &\quad + \mathbb{E}(\varepsilon_{\mathcal{M}}(\varepsilon_a^{\{i\}})^\top)M_i^\top + M_i\mathbb{E}(\varepsilon_a^{\{i\}}(\varepsilon_{\mathcal{M}})^\top).
 \end{aligned}$$

Since the different kinds of data are uncorrelated and from Definition 9.3, we obtain for all  $i \in \mathbb{N}$  that

$$\mathbb{V}(x^{\{i+1\}}) = M_i\mathbb{V}(x_a^{\{i\}})M_i^\top + \mathbb{V}(M_ix_t^{\{i\}}),$$

which completes the proof.

We remark that if the model operator  $\mathcal{M}_i$  is nonlinear but linearisable, formula (9.10) stays valid but gives only an approximation to the update of the error covariance matrix. We note that in meteorology  $M_i$  and  $M_i^\top$  are called tangent linear and adjoint model, respectively.

Formula (9.10) seems to be promising but it is absolutely not feasible for meteorological purposes. Due to the large number of grid points,  $n \approx 10^7$ , that is, the size  $n$  times  $n$  of the matrix  $M_i$  and its transpose makes the matrix product impossible to compute in a reasonable time. Hence, some other procedures are needed to approximate its effect. All the attempts in this direction originate from the ensemble predictions so far. Instead of taking only one initial analysis  $x_a^{\{i\}}$ , let us consider  $k \in \mathbb{N}$  pieces of them, that is, we take  $x_{a,j}^{\{i\}}$  for  $j = 1, \dots, k$ . At the end of the section we list some techniques how they are generated in practice. Proposition 9.1 implies that the error covariance matrix  $\mathbb{V}(x_a^{\{i\}})$  of the analysis can be estimated by

$$\mathbb{V}_{x_{a,1} \dots x_{a,k}}^{\{i\}} := \frac{1}{k-1} \sum_{j=1}^k \left( x_{a,j}^{\{i\}} - \frac{1}{k} \sum_{j=1}^k x_{a,j}^{\{i\}} \right) \left( x_{a,j}^{\{i\}} - \frac{1}{k} \sum_{j=1}^k x_{a,j}^{\{i\}} \right)^\top \quad (9.11)$$

for all  $i \in \mathbb{N}$ . Formula (9.11) is the basic of all presented methods approximating the effect of the Kálmán Filter (9.10). In what follows we will sometimes drop the index of the time level in order to ease the notation.

*Ensemble Kálmán Filter* enables us to update the forecast's error covariance matrix by multiplying smaller matrices, that is, it desires much less computational effort than the original Kálmán Filter (9.10), see e.g. in Houtekamer and Mitchell [10] and Evensen [5]. Given the analysis ensemble members  $x_{a,j}^{\{i\}}$  for  $j = 1, \dots, k$

we compute their sample mean from Definition 9.6 as

$$\mathbb{E}_{x_{a,1}\dots x_{a,k}}^{\{i\}} := \frac{1}{k} \sum_{j=1}^k x_{a,j}^{\{i\}} \quad \text{for all } i \in \mathbb{N}.$$

We define the matrices  $Z_a^{\{i\}}, Z_x^{\{i\}} \in \mathbb{R}^{n \times k}$  of the analysis and forecast perturbations, respectively, such that they contain the vectors

$$\frac{1}{\sqrt{k-1}}(x_{a,j}^{\{i\}} - \mathbb{E}_{x_{a,1}\dots x_{a,k}}^{\{i\}}) \quad \text{and} \quad \frac{1}{\sqrt{k-1}}(x_j^{\{i\}} - \mathbb{E}_{x_1\dots x_k}^{\{i\}})$$

in their  $j$ th column, respectively, for all  $j = 1, \dots, k$  and  $i \in \mathbb{N}$ . Proposition 9.1 implies that

$$\begin{aligned} \mathbb{V}(x_{a,j}^{\{i\}}) &\approx \mathbb{V}_{x_{a,1}\dots x_{a,k}}^{\{i\}} = Z_a^{\{i\}}(Z_a^{\{i\}})^\top \in \mathbb{R}^{n \times n} \quad \text{and} \\ \mathbb{V}(x_j^{\{i\}}) &\approx \mathbb{V}_{x_1\dots x_k}^{\{i\}} = Z_x^{\{i\}}(Z_x^{\{i\}})^\top \in \mathbb{R}^{n \times n} \end{aligned}$$

for any  $j = 1, \dots, k$ , where the approximation sign means an unbiased estimate. The forecast ensemble is now generated by updating the analysis perturbations by the model, that is,

$$Z_x^{\{i+1\}} = M_i Z_x^{\{i\}} \quad \text{for all } i \in \mathbb{N}. \quad (9.12)$$

Then we automatically obtain formula (9.10) for negligible  $\mathbb{V}(M_i x_t^{\{i\}})$  as

$$\begin{aligned} \mathbb{V}(x^{\{i+1\}}) &= Z_x^{\{i+1\}}(Z_x^{\{i+1\}})^\top = M_i Z_a^{\{i\}}(Z_a^{\{i\}} M_i)^\top \\ &= M_i Z_a^{\{i\}}(Z_a^{\{i\}})^\top M_i^\top = M_i \mathbb{V}(x_a^{\{i\}}) M_i^\top. \end{aligned}$$

Ensemble Kálmán Filter's advantage is that one needs to integrate with the model only  $k$  times in formula (9.12). We note that in the original setting the ensemble members  $x_{a,j}$  stem from the application of multiply analyses, i.e., application of the BLUE estimate (9.3) multiple times with a set explicitly perturbed observations (with a perturbation size in the range of the observation error variances) and a set of implicitly perturbed forecasts. In this case the estimate is optimal. If the model  $\mathcal{M}_i$  is nonlinear but linearisable, formula (9.12) reads as  $Z_x^{\{i+1\}} = \mathcal{M}_i(Z_x^{\{i\}})$ , and formula (9.10) gives only an approximation to the error covariance matrix.

*Ensemble Transform Kálmán Filter* is a technique which not only updates the forecast's error covariance matrix but also generates ensemble members for the next assimilation step. It is based on the idea that, as in the case of Ensemble Kálmán Filter, there is a relation between the analysis's and the forecast's perturbations. From the analysis ensemble, the new forecast members  $x_j^{\{i+1\}}$  are obtained by integrating with the model. By introducing the matrix  $Z_x^{\{i\}} \in \mathbb{R}^{k \times k}$  as before, we



are after the transformation matrix  $T \in \mathbb{R}^{k \times k}$  for which  $Z_a^{\{i+1\}} = Z_x^{\{i\}} T^{\{i\}}$  holds for all  $i \in \mathbb{N}$ . Bishop et al. [1] showed that  $T = V(\Lambda + I)^{-1/2}$  with

$$Z_x^\top H^\top \mathbb{V}(y)^{-1} H Z_x = V \Lambda V^\top.$$

Thus, matrix  $V$  contains the normalised eigenvectors and  $\Lambda$  the eigenvalues of the matrix on the left-hand side. Therefore, an eigenvalue decomposition has to be solved in each time step. By choosing a control member  $x_{a,1}^{\{i\}}$ , the columns of the matrix  $Z_a^{\{i\}}$  contain the perturbations to be added to  $x_{a,1}^{\{i\}}$  in order to generate the ensemble members. Given the analysis ensemble  $x_{a,j}^{\{i\}}$  for  $j = 1, \dots, k$  and  $i \in \mathbb{N}$ , the algorithm of the Ensemble Transform Kálmán Filter together with BLUE data assimilation (9.3) and optimal Kálmán gain matrix (9.4) is the following for all  $i \in \mathbb{N}$ :

$$\begin{aligned} x_j^{\{i+1\}} &:= \mathcal{M}_i(x_{a,j}^{\{i\}}) \quad \text{for } j = 1, \dots, k \\ (Z_x^{\{i+1\}})_j &:= \frac{1}{\sqrt{k-1}} (x_j^{\{i+1\}} - \mathbb{E}_{x_1, \dots, x_k}^{\{i+1\}}) \quad \text{for } j = 1, \dots, k \\ (Z_x^{\{i+1\}})^\top H_{i+1}^\top \mathbb{V}(y^{\{i+1\}})^{-1} H_{i+1} Z_x^{\{i+1\}} &= V^{\{i+1\}} \Lambda^{\{i+1\}} (V^{\{i+1\}})^\top \\ T^{\{i+1\}} &:= V^{\{i+1\}} (\Lambda^{\{i+1\}} + I)^{-1/2} \\ Z_a^{\{i+1\}} &:= Z_x^{\{i+1\}} T^{\{i+1\}} \\ \mathbb{V}_{x_1, \dots, x_k}^{\{i+1\}} &:= Z_x^{\{i+1\}} (Z_x^{\{i+1\}})^\top \\ K_{i+1} &:= \mathbb{V}_{x_1, \dots, x_k}^{\{i+1\}} H_{i+1}^\top (\mathbb{V}(y^{\{i+1\}}) + H_{i+1} \mathbb{V}_{x_k, \dots, x_k}^{\{i+1\}} H_{i+1}^\top)^{-1} \\ x_{a,1}^{\{i+1\}} &:= x_1^{\{i+1\}} + K_{i+1} (y^{\{i+1\}} - \mathcal{H}(x_1^{\{i+1\}})) \\ x_{a,j}^{\{i+1\}} &:= x_{a,1}^{\{i+1\}} + (Z_a^{\{i+1\}})_j \end{aligned}$$

generating the new analysis ensemble members  $x_{a,j}^{\{i+1\}}$  and the updated approximate value  $\mathbb{V}_{x_1, \dots, x_k}^{\{i+1\}}$  of the forecast's error covariance matrix.

We note again that the same procedure works for nonlinear model and observation operators  $\mathcal{M}_i$  and  $\mathcal{H}_i$  as well, however, it only leads to an approximative time evolution of the error covariance matrix. For more methods in the nonlinear case we refer to Sect. 9.7.

Previously, we supposed that there existed  $k$  pieces of analysis perturbations  $x_{a,j}^{\{i\}}$ ,  $j = 1, \dots, k$  being valid at time level  $i$ . The question arises how they are generated in practice. The perturbation of the observations has already been mentioned. One can of course randomly perturb the actual analysis  $x_a^{\{i\}}$  field itself. The *time-lagged approach* uses a mixture of two analyses initiated from two different time levels but being valid at the same time level, see e.g. in Hoffman and Kalnay [9]. This latter techniques will, however, not necessarily lead to perturbations being near to

the directions along that the model stretches the most, which would be one of the most beneficial requirements.

To this end, the *breeding method* was initiated where some initial random perturbations are added to the nonlinear model, and these models integrate the same initial analysis field  $x_a^{(0)}$ . The solution to the unperturbed model is then subtracted from the other solutions at each step, and the appropriately scaled differences are added to the unperturbed solution again to generate the new analysis perturbations for the next step. After some time, breeding method yields the so-called bred vectors approximating the directions in phase space where the instabilities grow fastest. The technique is described e.g. in Tóth and Kalnay [18, 19], and Kalnay [13].

Another popular perturbation generating technique is the method of *singular vectors*. The idea behind the method is the following. One considers a spatially discretised partial differential equation leading to an ordinary differential equation of the form  $\frac{d}{dt}x(t) = \mathcal{N}(x(t))$ ,  $t \geq t_0$  for the continuously differentiable functions  $x: \mathbb{R}_0^+ \rightarrow \mathbb{R}^n$ ,  $\mathcal{N}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  for some  $n \in \mathbb{N}$ . If the initial value  $x(t_0) = x_0$  is oppressed by a certain error  $e_0$ , a first-order approximation to the time evolution of the error term  $e(t)$  can be obtained from the linearised equation  $\frac{d}{dt}e(t) = \mathcal{J}(t)e(t)$  with the initial value  $e(t_0) = e_0$ , where  $\mathcal{J}(t) = \mathcal{N}'(x(t))$  denotes the Jacobian of  $\mathcal{N}$  taken at the state  $x(t)$  for all  $t \geq t_0$ . Then there exists a matrix  $\Psi(t) \in \mathbb{R}^{n \times n}$  such that the solution to this problem has the form  $x(t) = e^{\Psi(t-t_0)}x_0$  for all  $t \geq t_0$ . Since the matrix  $\Psi(t)$  is difficult to compute exactly (it is the sum of infinitely many terms containing the integral of various commutators of  $\mathcal{J}(t)$ ), certain approximation is computed in practice (e.g., Magnus method). By choosing the initial error term such that  $\|e_0\| = \varepsilon$  for some  $\varepsilon > 0$ , that is, being on the surface on the  $n$  dimensional sphere of radius  $\varepsilon$ . Our aim is now to determine how this sphere evolves subject the nonlinear model  $\mathcal{N}$ . To this end, we denote the propagator by  $E(t, t_0) := e^{\Psi(t-t_0)}$  and the scalar product in  $\mathbb{R}^n$  by  $\langle \cdot, \cdot \rangle$ . We compute now

$$\|e(t)\|^2 = \|E(t, t_0)e_0\|^2 = \langle E(t, t_0)e_0, E(t, t_0)e_0 \rangle = \langle E(t, t_0)^\top E(t, t_0)e_0, e_0 \rangle.$$

Due to the norm inequality we also have that

$$\|e(t)\|^2 = \|E(t, t_0)e_0\|^2 \leq \|E(t, t_0)\|^2 \|e_0\|^2 = \varepsilon^2 \|E(t, t_0)\|^2.$$

Altogether we have  $\langle E(t, t_0)^\top E(t, t_0)e_0, e_0 \rangle \leq \varepsilon^2 \|E(t, t_0)\|^2$ , that is,

$$\langle E(t)e_0, e_0 \rangle \leq 1 \tag{9.13}$$

with the matrix

$$E(t) = \frac{1}{\varepsilon^2 \|E(t, t_0)\|^2} E(t, t_0)^\top E(t, t_0).$$

Formula (9.13) gives the equation of an ellipsoid. The directions of its axes are given by the eigenvectors of the matrix  $E(t)$ . These directions give, namely, the directions

along which the nonlinear model  $\mathcal{N}$  stretches/compresses the error function  $e(t)$  initially lying on the sphere. When generating the analysis perturbations, one is interested in that directions where the stretching the larger is. The direction of the largest stretching is given by the eigenvector belonging to the largest eigenvalue of the matrix  $E(t)$ , and so on. Since the eigenvectors/eigenvalues of the matrix  $E(t, t_0)^\top E(t, t_0)$  are called singular vectors/values of the matrix  $E(t, t_0)$ , we call this procedure the method of singular vectors. In order to get the singular vectors belonging to the leading singular values, one needs to integrate with the tangent linear model forward in time and then with the adjoint model backward in time many times, see e.g. in Errico [4].

We note, however, that in numerical weather prediction, the method of singular vector is always combined with either ensemble analyses with perturbed observations (EDA) or with perturbations generated by the Ensemble Transform Kálmán Filter. As we have already mentioned, Ensemble Transform Kálmán Filter generates not only the forecast's error covariance matrix  $\mathbb{V}^{\{i+1\}}$  but the analysis perturbations as well. Comparisons between breeding method and Ensemble Transform Kálmán Filter, between singular vectors and EDA, and between EDA and Ensemble Transform Kálmán Filter are presented in Wang and Bishop [2, 24], and [7], respectively.

## 9.6 Numerical Experiments

In order to illustrate the use of the methods introduced above, we present the results of numerical experiments done for simple models. The reason of choosing these models for the experiments is twofold. On one hand, they are of low dimensions with  $n = 1$  and  $n = 3$ , respectively, therefore, Kálmán Filter can directly be applied and there is no need to apply one of its approximations (such as Ensemble or Ensemble Transform Kálmán Filter). On the other hand, the exact solution of the first system is known, therefore, the behaviour of the data assimilation methods can easily be explained. Although the second system does not admit a known exact solution, it shares certain properties with the meteorological models (such us nonlinearity, sensitivity to the initial values, etc.), making it a perfect test model to study the performance of data assimilation methods.

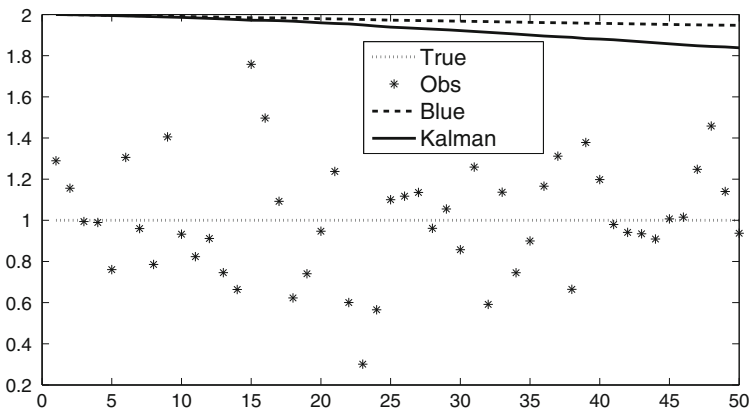
### 9.6.1 Linear Iteration

We consider the system  $x^{\{i+1\}} = x^{\{i\}}$  for  $x^{\{i\}} \in \mathbb{R}$  and  $i \in \mathbb{N}$  with  $x^{\{0\}} = 1$ . One can see that the true state equals  $x_t = x^{\{0\}} = 1$ . We suppose that the observations are unbiased and normally distributed perturbations of the true state:

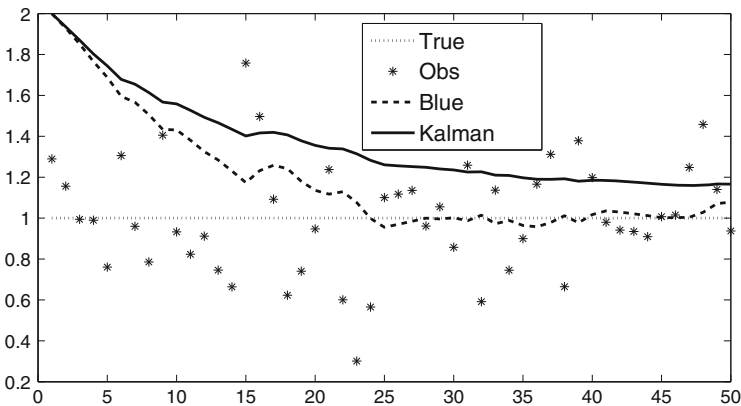
$$y^{\{i\}} = x_t + N(0, \mathbb{V}(y)) \quad \text{for all } i \in \mathbb{N} \quad (9.14)$$

where  $\mathbb{V}(y) = 0.3$  is given. The simulations aim at illustrating the effect of the forecast's error covariance matrix  $\mathbb{V}(x)$ . Since  $\mathbb{V}(x)$  represents the reliability of the forecast, we can study how the solution changes depending on how much we rely on the forecast. Another goal is to show the advantage of Kálmán Filter, therefore, we present the same numerical experiments using BLUE (9.3)(9.4) and BLUE together with Kálmán Filter (9.10). This enables us to study how the analysis, being initially  $x_a^{(0)} = 2$  far away from the true state  $x_t = 1$ , evolves in time in the two different cases.

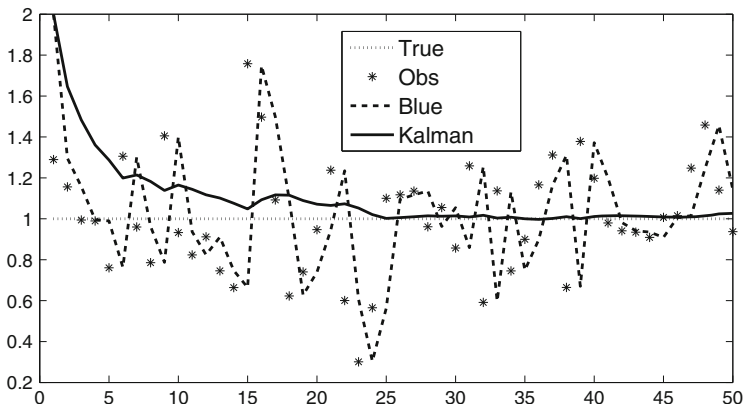
In Figs. 9.1, 9.2, and 9.3 the numerical results are shown for three values of the error covariance matrix  $\mathbb{V}(x) = 10^{-4}, 10^{-2}, 10$ , respectively. Figure 9.1 illustrates the case when we trust the forecast very much: The results of both the BLUE and the Kálmán Filter methods are far from the true value  $x_t = 1$  and follow the initial (wrong) analysis value  $x_a^{(0)} = 2$ . Figure 9.2 corresponds to the case when we treat



**Fig. 9.1** Linear iteration with  $\mathbb{V}(x) = 10^{-4}$



**Fig. 9.2** Linear iteration with  $\mathbb{V}(x) = 10^{-2}$



**Fig. 9.3** Linear iteration with  $\mathbb{V}(x) = 10$

the observations more reliable than in the latter case but still less reliable than the forecast. One can see that we obtain better results: Both methods converge to the true value  $x_t = 1$ . Figure 9.3 illustrates the case when  $\mathbb{V}(x)$  has a much greater initial value than  $\mathbb{V}(y)$ , that is, we believe the observations much more reliable than the forecast. Then the numerical results of the BLUE method completely follows the observations, while the Kálmán Filter method updates  $\mathbb{V}(x)$  in a perfect way: Its result finds the true solution very quickly.

### 9.6.2 Lorenz System

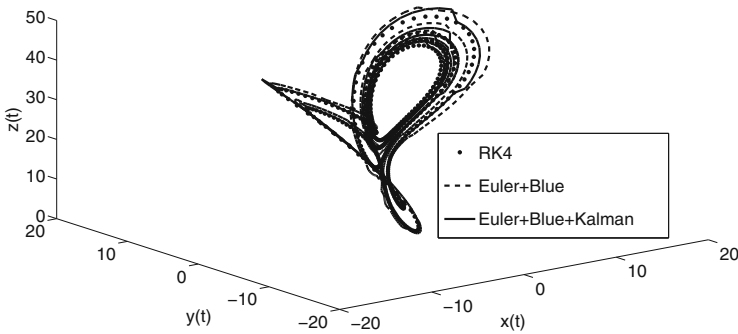
Our second example is the nonlinear three-dimensional Lorenz system. In 1963 Edward Lorenz developed a simplified mathematical model for atmospheric convection in [14]. The model is a system of three ordinary differential equations now known as the Lorenz equations:

$$\frac{d}{dt}x(t) = \sigma(y(t) - x(t)), \tag{9.15}$$

$$\frac{d}{dt}y(t) = x(t)(\rho - x(t)) - y(t), \tag{9.16}$$

$$\frac{d}{dt}z(t) = x(t)y(t) - \beta x(t), \tag{9.17}$$

where  $x, y, z : (0, \infty) \rightarrow \mathbb{R}$  are the unknown functions, and  $\sigma, \rho, \beta \in \mathbb{R}$  are parameters with specific values  $\sigma = 10, \rho = 28, \beta = \frac{8}{3}$ . Since its exact solution is not known, we solve the system numerically by using the first-order Euler method and the fourth-order Runge–Kutta method with time step  $\Delta t = 0.01$ . We consider the latter as the observations  $y$  at each time step. The solution with the Euler method is considered as the model’s forecast. In the simulations we fix the covariance



**Fig. 9.4** The different trajectories in the phase space with parameters  $\mathbb{V}(\mathcal{M}_i(x_i)) = 10^{-2}$ ,  $t = 2$ ,  $N = 10$

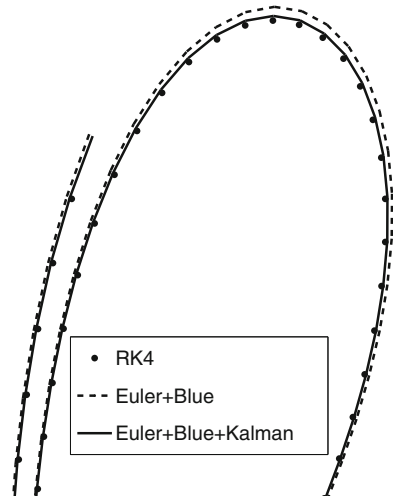
matrices as  $\mathbb{V}(y) = (\Delta t)^8 \cdot \mathbf{I} \in \mathbb{R}^{3 \times 3}$  and  $\mathbb{V}(x) = \frac{1}{2} \mathbb{V}(y)$ . Our aim is to investigate the role of the covariance matrix of the model's error  $\mathbb{V}(\mathcal{M}_i(x_i))$ , therefore, we set it to the following three value:  $\mathbb{V}(\mathcal{M}_i(x_i)) = 1, 10^{-2}, 10^{-10}$ . As before, we apply BLUE (9.3), (9.4) with and without Kálmán Filter (9.10).

Figure 9.4 shows the trajectories in the phase space  $(x, y, z)$  from the initial point  $(2, 5, 10)$  by applying the fourth-order Runge–Kutta (RK4) method and the explicit Euler method with data assimilation method BLUE with and without Kálmán Filter. One can immediately see that the three solutions differ, a more detailed study will follow. We analyse first how the solution depends on the frequency of the data assimilation, that is, on the number  $N$  of the time steps after which the analysis  $x_a$  is computed. Figure 9.5 shows the results when the data assimilations methods are performed in each time step. One can see that all the three trajectories are closed to each other at the beginning, however, the Kálmán Filter method performs better then the BLUE method alone. Figure 9.6 illustrates the case when the data assimilation methods are performed only at every 10th time step. One can see that both methods have greater distance form the observations (RK4) as before. Furthermore, they experience “jumps” after each 10 time steps when their solutions are forced to follow the reliable observations by the data assimilation procedure. The same phenomena can be observed in Fig. 9.7 when data assimilation is performed in each 50th time steps. Both Figs. 9.6 and 9.7 show that the solutions cover each other at the beginning, hence, the Kálmán Filter benefits from the update of the error covariance matrix  $\mathbb{V}(x)$  only after the first data assimilation step.

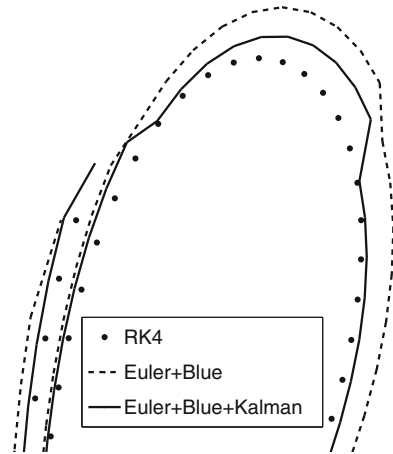
In Fig. 9.8 the relative error of BLUE

$$\text{err}^{\{i\}} := \frac{\|x_a^{\{i\}} - y^{\{i\}}\|_2}{\|y^{\{i\}}\|_2}$$

**Fig. 9.5** Lorenz system with data assimilation frequency  $N = 1$

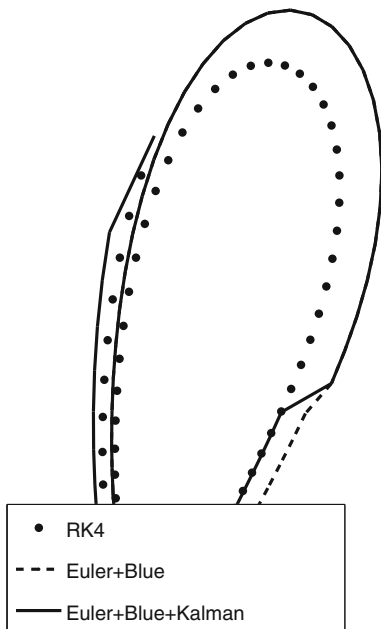


**Fig. 9.6** Lorenz system with data assimilation frequency  $N = 10$

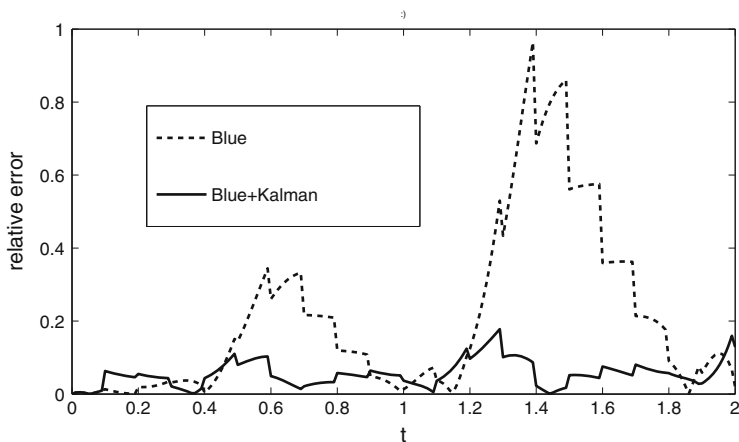


with and without Kálmán Filter is shown. One can see that the Kálmán Filter always performs better than the BLUE alone.

We investigated the effect of the model’s error covariance matrix  $\mathbb{V}(\mathcal{M}_i(x_t))$  as well. Figure 9.9 shows our results for the values  $\mathbb{V}(\mathcal{M}_i(x_t)) = 1, 10^{-8},$  and  $10^{-16}$ , respectively. One can see that in the first case the solutions follow almost the same trajectories. The explanation is that in this case  $\mathbb{V}(\mathcal{M}_i(x_t)) = 1$ , that is, the model is considered unreliable, therefore, the solutions rely on the measurements (obtained by the fourth-order Runge–Kutta method RK4). If the value of  $\mathbb{V}(\mathcal{M}_i(x_t))$  is decreased, the data assimilation methods treat the model more reliable and try to follow its trajectory. In the case  $\mathbb{V}(\mathcal{M}_i(x_t)) = 10^{-16}$  the situation is clear: The BLUE method still follows the measurements (because  $\mathbb{V}(\mathcal{M}_i(x_t))$  does not play



**Fig. 9.7** Lorenz system with data assimilation frequency  $N = 50$



**Fig. 9.8** Relative error of the BLUE method with and without Kálmán Filter technique

any role in its computation), however, the Kálmán Filter method tries to converge to the model’s trajectory.

The explanation of the expected behaviour is the following. The measurements stem from the use of the fourth-order Runge–Kutta method being more accurate



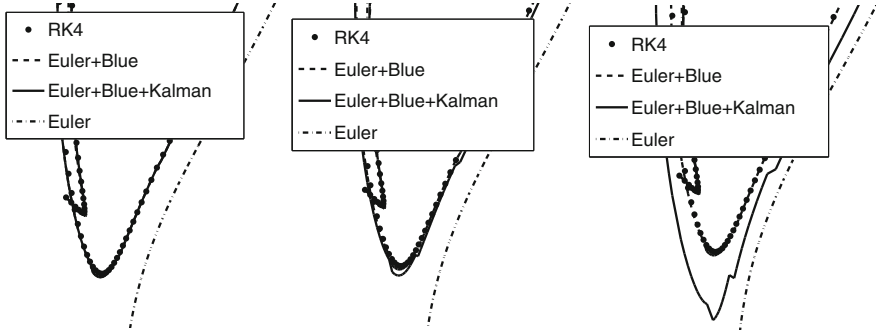


Fig. 9.9 Lorenz system with  $\mathbb{V}(\mathcal{M}_i(x_i)) = 1, 10^{-8}, 10^{-16}$ , respectively

than the first-order Euler method which provides the model’s forecast. Without applying any data assimilation methods, the model’s forecast (indicated by “Euler” in the Figures) differs very much from the (more accurate) measurements (indicated by “RK4” in the Figures). Hence, contrary to the case of the models used in numerical weather prediction where the true state of the atmosphere is somewhere “between” the measurements and the model’s forecast, in this setting it is clearly known that the (unknown) exact solution is nearer to the measurements’ trajectory. Application of a data assimilation method results in a more accurate solution which approaches therefore the trajectory of the measurements. Exactly this scenario can be observed in Fig. 9.9: Both data assimilation methods (BLUE and BLUE with Kálmán Filter) improves the model’s forecast. In the third case, when the inaccurate model is undeservedly trusted too much (i.e. its error covariance matrix is small,  $\mathbb{V}(\mathcal{M}_i(x_i)) = 10^{-16}$ ), the Kálmán Filter follows the trajectory of the Euler method causing a significant error in the analysis.

The results above illustrate that the use of a flow-dependent data assimilation method (e.g. Kálmán Filter or its approximate versions) itself is not enough for improving the weather forecast, setting the appropriate value of the model’s error covariance matrix  $\mathbb{V}(\mathcal{M}_i(x_i))$  is important as well. Since the model’s error includes not only the numerical error originated from the space and time discretisation of the corresponding partial differential equations, but also the error done by the parametrisations of various physical processes and the error of the boundary conditions, very little is known about the its covariance matrix  $\mathbb{V}(\mathcal{M}_i(x_i))$ . It is usually modelled by adding some noise with zero mean to the forecast (or to each member of the forecast ensemble). Although there are several results related, see e.g. in Raynaud et al. [15], Trémolet [20], Düben and Palmer [3] and the references therein, the further study of this issue is highly anticipated.

The studies presented above aim at giving an insight how the parameters of the data assimilation methods effect the solution’s accuracy. We showed that the analysis  $x_a$  depends very much on the frequency of the assimilation step and on the corresponding error covariance matrices. Hence, their right choice is crucial for the efficient use of data assimilation methods.

## 9.7 Outlook: Nonlinear Data Assimilation

We have seen previously that the solution to the linear data assimilation problem (9.3) is known and given in formula (9.4). The solution to the nonlinear data assimilation problem, that is, when the model and observation operators  $\mathcal{M}$  and  $\mathcal{H}$  are nonlinear, is given and studied e.g. by van Leeuwen and Evensen in [21]. Since its derivation is based on Bayes' theorem, and in practice the probability density functions can be far from being Gaussian, there is a demand for new techniques which (1) do not use linearisation and (2) lead to a nonlinear analysis.

In this section we present some ideas how to proceed when the system is not linear. For a more detailed introduction, we refer the reader to van Leeuwen [23]. The most common approaches of treating the nonlinearity are the use of incremental variational analysis and the particle filter methods.

We present the *incremental variational analysis* by applying it to the four-dimensional variational analysis (4D-Var), see e.g. in Talagrand and Courtier [17], Trémolet [20], which belongs to the class of variational data assimilation techniques presented in Sect. 9.4. Its cost function is similar to that presented in (9.5), however, it takes into account the effect of the various observations at their proper time levels. We consider the nonlinear model  $\mathcal{M}_i$ , the nonlinear observation operator  $\mathcal{H}_i$ , the observations  $y^{\{i\}}$  at the  $i$ th time level, and the previous forecast  $x$  being valid at the time level  $t_0$ . Then the cost function  $J(\xi)$  of the 4D-Var method reads as

$$J(\xi_0) := \frac{1}{2}(x - \xi)^\top \mathbb{V}(x)^{-1}(x - \xi_0) + \frac{1}{2} \sum_{i=0}^I (y^{\{i\}} - \mathcal{H}_i(\xi^{\{i\}}))^\top \mathbb{V}(y)^{-1}(y^{\{i\}} - \mathcal{H}_i(\xi^{\{i\}})) \quad (9.18)$$

with  $\xi^{\{i\}}$  subject the nonlinear model  $\xi^{\{i\}} = \mathcal{M}_{i-1}(\xi^{\{i-1\}})$ ,  $i = 1, \dots, I$ . By denoting  $\delta_0 := K(y - \mathcal{H}(x))$ , the identity (9.1) reads as  $x_a = x + \delta_0$ , hence, formula (9.18) can be rewritten as

$$J(\delta_0) := \frac{1}{2} \delta_0^\top \mathbb{V}(x)^{-1} \delta_0 + \frac{1}{2} \sum_{i=0}^I (y^{\{i\}} - \mathcal{H}_i(\xi^{\{i\}}))^\top \mathbb{V}(y)^{-1}(y^{\{i\}} - \mathcal{H}_i(\xi^{\{i\}})).$$

The terms in the sum can be approximated by using the linearisation of the nonlinear operators  $\mathcal{M}_i$  and  $\mathcal{H}_i$  around the state  $\xi^{\{i-1\}} := \xi^{\{i\}} - \delta^{\{i-1\}}$  with the vector-valued random variables  $\delta_i \in \mathbb{R}^n$ ,  $i = 0, \dots, I$ :

$$\begin{aligned} \mathcal{M}_i(\xi^{\{i\}}) &\approx \mathcal{M}_{i-1}(\xi^{\{i-1\}}) + \mathcal{M}'_{i-1}(\xi^{\{i-1\}}) \delta^{\{i-1\}}, \quad i = 1, \dots, I, \\ \mathcal{H}_i(\xi^{\{i\}}) &\approx \mathcal{H}_{i-1}(\xi^{\{i-1\}}) + \mathcal{H}'_{i-1}(\xi^{\{i-1\}}) \delta^{\{i-1\}}, \quad i = 1, \dots, I. \end{aligned} \quad (9.19)$$

Putting these formulas together, one obtains an approximative cost function whose minimisation leads to an approximation to  $\tilde{\delta}_0$ . In order to take the linear operators

$\mathcal{M}'_i(\xi^{i\beta})$  and  $\mathcal{H}'_i(\xi^{i\beta})$  at the proper time levels, one needs an outer loop to compute the linearisations (9.19) at each time level.

The inner loop contains then the minimisation of the cost function, that is, we seek that state  $x_a := \xi$  for which  $\frac{d}{d\xi}J(\xi) = 0$ . This problem can be rewritten in the form  $A\delta = b$  where matrix  $A$  contains the linear operators  $\mathcal{M}'_i(\xi^{i\beta})$  and  $\mathcal{H}'_i(\xi^{i\beta})$  as well, and the vector  $b$  contains the terms  $y^{i\beta} - \mathcal{H}_i(\xi^{i\beta})$ . Such problems are usually solved by the *conjugate gradient method* which global error depends on the condition number  $\kappa := \|A\|\|A^{-1}\|$ , that is, it converges fast if  $\kappa$  is small enough. There exist several preconditioning technics used to reduce the condition number of the problem and to obtain faster convergence, see e.g. in Faragó and Karátson [6]. A survey about the (pre)conditioning of the model operators appearing in meteorological modelling can be found e.g. in Haben et al. [8].

Another approach to treat nonlinearity is the *particle filtering*. The idea behind it is already presented in Sect. 9.5 about Ensemble Kálmán Filter, that is, the model's probability density function is approximated by using random ensemble members (also called as particles). More precisely, in this case we need to approximate the conditional density function which measures the probability density of the atmosphere's actual state given the specific observations. Then the conditional density function is represented as the weighted sum of Dirac functions positioned at the various particles (i.e., model states). Intuitively, we choose various particles and make them propagate with time subject the nonlinear model. For the next step we consider only those particles which “arrived” near to the observations, and by a resampling procedure we generate new particles from them. The weights in the sum correspond to the particles' distance from the observations. Then we repeat the cycle with the same amount of particles as in initial step. Since the derivation of particle filtering is based on the conditional probability theory (e.g. Bayesian statistics, stochastic filtering, Monte–Carlo methods), this is out of the scope of the present chapter, however, a detailed introduction can be found in van Leeuwen [22].

**Acknowledgements** The authors thank Gergely Bölöni for the careful reading, the corrections, and the useful suggestions. The financial support of the “Stiftung Aktion Österreich–Ungarn” (project nr. 89öu11) is kindly acknowledged.

## References

1. Bishop, C.H., Etherton, B.J., Majumdar, S.J.: Adaptive sampling with the ensemble transform Kalman filter. *Mon. Weather Rev.* **129**, 420–436 (2001)
2. Buizza, R., Leutbecher M., Isaksen, L.: Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Q. J. Roy. Meteorol. Soc.* **134**, 2051–2066 (2008)
3. Düben, P.D., Palmer, T.N.: Benchmark tests for numerical weather forecasts on inexact hardware. *Mon. Weather Rev.* **142**, 3809–3829 (2014)
4. Errico, R.M.: What is an adjoint model? *Bull. Am. Meteorol. Soc.* **78**, 2577–2591 (1997)
5. Evensen, G.: *Data assimilation. The Ensemble Kalman Filter*. Springer, Berlin (2009)

6. Faragó, I., Karátson, J.: *Numerical Solution of Nonlinear Elliptic Problems via Preconditioning Operators: Theory and Applications*. Nova Science Publishers, New York (2002)
7. Gustafsson, N., Bojarova, J., Vigne, O.: A hybrid variational ensemble data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Processes Geophys.* **21**, 303–323 (2014)
8. Haben, S.A., Lawless, A.S., Nichols, N.K.: Conditioning of incremental variational data assimilation, with application to the met office system. *Tellus A* **63**, 782–792 (2011)
9. Hoffman, R.N., Kalnay, E.: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A* **354**, 100–118 (1983)
10. Houtekamer, P.L., Mitchell, H.L.: Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**, 797–811 (1998)
11. Johnson, R.A., Wichern, D.W.: *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, Upper Saddle River (2007)
12. Kalman, R.E.: A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960)
13. Kalnay, E.: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge (2002)
14. Lorenz, E.N.: Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963)
15. Raynaud, L., Berre, L., Desroziers, G.: Accounting for model error in the Météo-France ensemble data assimilation system. *Q. J. R. Meteorol. Soc.* **2138**, 249–262 (2012)
16. Schönemann, P.H.: On the formal differentiation of traces and determinants. *Multivar. Behav. Res.* **20**, 113–139 (1985)
17. Talagrand, O., Courtier, P.: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory. *Q. J. R. Meteorol. Soc.* **113**, 1311–1328 (1987)
18. Toth, Z., Kalnay, E.: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330 (1993)
19. Toth, Z., Kalnay, E.: Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.* **125**, 3297–3319 (1997)
20. Trémolet, Y.: Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**, 2483–2504 (2006)
21. van Leeuwen, P.J., Evensen, G.: Data assimilation and inverse methods in terms of probabilistic formulation. *Mon. Weather Rev.* **124**, 2898–2913 (1996)
22. van Leeuwen, P.J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Q. J. R. Meteorol. Soc.* **136**, 1991–1999 (2010)
23. van Leeuwen, P.J., Cheng, Y., Reich, S.: *Nonlinear Data Assimilation*. Springer, Berlin (2015)
24. Wang, X., Bishop, C.H.: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**, 1140–1158 (2003)