CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

Part I

General Background

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

1
Data Assimilation: General Background

1.1 Introduction

Data assimilation includes two main components: simulation model and data. The simulation model is defined as a mathematical/numerical system that can simulate an event or a process. In most typical settings the simulation model is a prediction model based on partial differential equations (PDEs) that often includes empirical parameters. Data are generally associated with observations made by a measuring instrument, although data could also imply a product obtained by processing observations. Using an example from meteorology, data include observations such as atmospheric temperature and satellite radiances. The goal of data assimilation is to combine the information from a simulation model and data in order to improve the knowledge of the system, described by the simulation model. Apparently, the formulation of data assimilation will depend on interpretation of the *knowledge of the system*. Before we attempt to clarify a possible interpretation, it is useful to further understand the simulation model and data.

In agreement with common applications in geosciences and engineering, we narrow our discussion to a dynamic-stochastic PDE-based prediction model. Prediction models are developed with the general idea of improving the prediction of various phenomena of interest. From the theory of PDEs it is known that various parameters can impact the result of PDE integration, such as initial conditions (ICs), model errors (MEs), and empirical model parameters (EMPs). It is widely recognized that our knowledge of these parameters is never perfect, implying uncertainty of these parameters and uncertainty of the prediction calculated using such uncertain parameters.

Since the ultimate goal of using prediction models is to produce an improved prediction, it is natural to prefer a prediction that is in some way optimal. Such a prediction should be reliable, implying a desire to have a very small uncertainty associated with prediction. Then, the question is: How can the prediction be improved? First, it is anticipated that by improving the mentioned parameters (ICs, MEs, EMPs) and reducing their uncertainty would result in a desirable prediction. One could also try to improve model equations by including missing physical processes, coupling relevant components, and/or improving spatiotemporal resolution (if the prediction model is discretized). However, the only way to improve prediction is to introduce new information about the model parameters or model equations. The new information could come from another model with superior performance, but the most common source of new information about the real world comes

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

from observations. An additional source of information could be introduced from past model performances if it is believed that the prediction model has some skill. If the prediction model has no skill, then observations are the only source of information, and one has to rely on using purely statistical methods. If the prediction model has some skill, however, then it is possible to combine the information from observations and from past model performances and then rely on using data assimilation.

Note that all sources of information, from observations and from prediction models, are uncertain. We already suggested that imperfect knowledge of model parameters (ICs, MEs, and/or EMPs), as well as model equations, implies an imperfect prediction. Information from observation is also not perfect. There are instrument errors, transmission errors, local errors, as well as the so-called representativeness errors. The instrument error is associated with every measuring instrument and can vary depending on the accuracy of the instrument. The errors created during a transmission from observation site to central location may not be detected in some instances and will contribute to observation error. Local errors refer to unforeseen errors of the local observation site, such as artificial heat sources and the impact of local vegetation. The representativeness error is the error caused by model prediction that is not representative of the actual observation. This can refer to inadequate model resolution, volume-averaged model variable versus point observation, etc. Therefore, observations also have errors, i.e., uncertainties.

Given that the two main components of data assimilation, prediction model and data, are inherently uncertain, then the output of data assimilation, the knowledge of the system, is expected to be uncertain as well. Uncertainty can be measured in many different ways. One can think of uncertainty as a measure of the difference between an estimate and the truth, if the truth is known. Unfortunately, the true value of the field is rarely known, except in a controlled experiments such as an observation system simulation experiment (OSSE). The theory of probability offers a mathematically consistent, formal way of dealing with uncertainties, and is used in our approach to data assimilation. A comprehensive object that describes the probabilistic system is the probability density function (PDF). Therefore, one can think of the PDF as the actual knowledge of the system, implying that the ultimate goal of data assimilation is to estimate the PDF. As will be shown in Chapters 3, 7, 8, 9, and 12, estimating the PDF is quite a challenging problem in realistic high-dimensional applications of data assimilation, mostly limiting practical data assimilation to estimating the first PDF moment (e.g., mean) and eventually the second PDF moment (e.g., covariance), with only an occasional capability of estimating the higher-order PDF moments.

Another critical aspect of data assimilation is the processing of information. Both prior model realizations and data contain information that can potentially contribute to improving the state of knowledge. Shannon's information theory (Shannon and Weaver, 1949), also based on using the probabilistic approach, offers the mathematical formalism for quantifying and processing information. Although still not used to its maximum, this information theory is a very handy tool for data assimilation. Implied from the above discussion of the impact of model parameters, such as ICs, MEs, and EMPs, on the prediction made by the model, and the aspiration of data assimilation to improve prediction by modifying model parameters ICs, MEs, and/or EMPs, the control theory is also an important tool of

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

data assimilation. The implied dynamic-stochastic characterization of a prediction model also implies the important role of statistics and possibly chaotic nonlinear dynamics in data assimilation. Given that data assimilation is typically multivariate and applied to vectors and matrices, it relies heavily on using linear algebra and functional analysis.

There are several other considerations that are important for data assimilation. Realistic physical phenomena and processes, and their relation to observed variables, are all inherently nonlinear. As such, the treatment of nonlinearity in data assimilation plays an important role in choosing the adequate control theory methods and limiting the utility of linear algebra. The dynamical aspect of prediction models, generally characterized by time-dependent phenomena, implies that prediction uncertainties have to be dynamical and time-dependent as well. Given the sensitivity of PDEs to the initial (and boundary) conditions, data assimilation has to provide dynamically balanced ICs that would not cause spurious perturbations in prediction. In the case of chaotic nonlinear dynamics, as most realistic dynamical systems are, data assimilation needs to capture and eventually remove the errors of growing and neutral modes from the ICs.

With all these components, probability theory, statistics, information theory, control theory, linear algebra, and functional analysis, make data assimilation very complex and challenging.

1.2 Historical Background

First attempts to address what we now call data assimilation could be traced to data fitting and regression analysis applied in astronomy, most notably by Legendre (1805) and Gauss (1809). In solving the problem Gauss assumed normally distributed errors and introduced the normal probability distribution. Around that time Laplace (1814) introduced the Bayesian approach by developing a mathematical system on inductive reasoning based on probability. Starting with these discoveries, and after a considerable development of mathematical tools and theories, the modern-age data assimilation was made possible.

Early methods for data assimilation were deterministic and essentially represented a function fitting to measurements. This included the interpolation methods with distance-based interpolation weights in order to determine the relative importance of observations, such as the objective analysis schemes of Bergthórsson and Döös (1955), Gilchrist and Cressman (1954), Cressman (1959), and Barnes (1964). While useful for operational numerical weather prediction (NWP) of that time, these methods did not explicitly include probabilistic considerations. Other deterministic methods include nudging data assimilation (Hoke and Anthes, 1976; Davies and Turner, 1977), sometimes also referred to as four-dimensional data assimilation (4DDA) or a dynamic relaxation method. Later developments of the method include a generalization to accept uncertainties (e.g., Zou et al., 1992). Nudging implies a change of the original dynamical equations of a prediction model to include a forcing term. The coefficients associated with the forcing are generally determined by fitting the model state closer to the observations. Although nudging has been improved to implicitly accept uncertainties, it does not rely on using the Bayesian approach and does not attempt to estimate PDF moments as probabilistic data assimilation does.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

Probably the first data assimilation method that is critically relevant for understanding modern-age data assimilation is the Kalman filter (KF) (Kalman and Bucy, 1961), initially developed for signal processing. It provides a mathematically consistent methodology based on probability and Bayesian principles that produces a minimum variance solution. The KF is also helpful in describing the role of dynamics in forecast error covariance, as well in model error covariance. Since the KF is defined for linear systems, it fully resolves the Gaussian PDF and in that sense represents a satisfactory solution to general probabilistic data assimilation problems. There are, however, major obstacles in making the KF a practical data assimilation method. For one, it is a linear filtering method and as such it cannot satisfactorily address nonlinearities in the prediction model and observations. Another major obstacle is the required matrix inversion, which becomes practically impossible to calculate in realistic high-dimensional applications. Strictly relying on the Gaussian PDF assumption is also a disadvantage of the KF, given that prediction model variables and observations could have non-Gaussian errors.

The first practical method that incorporates the basic data assimilation setup with Bayesian and probabilistic assumptions is the optimal interpolation (OI) method of Gandin (1963), sometimes referred to as statistical interpolation. This is a minimum variance estimator and as such it can be related to the KF and other probabilistic data assimilation methods. A more detailed overview of OI can be found in Daley (1991) (see chapters 3, 4, and 5 therein). The OI method is very much a simplified version of the KF. The OI employs a linear observation operator, in early versions only the identity matrix. For nonlinear observations, such as satellite radiances, an inversion algorithm (i.e., retrieval) that produces a model variable from observations is required. The forecast error covariance is modeled and includes separate vertical and horizontal correlations. By construction the forecast error, covariance is homogeneous (i.e., all grid points are treated equally) and isotropic (all directions are treated equally). In addition, the covariance is stationary, being approximated by a correlation function with statistically estimated correlation parameters. Since it is related to the KF, OI can also produce an estimate of the posterior error covariance. However, such an estimate is not reliable since the input covariances and parameters are not accurate. The OI is also local, in the sense that only observations within a certain distance from the model point impact the analysis at that point. Although theoretically and practically an important step in probabilistic data assimilation development, when measured against our motivation to produce a reliable estimate of PDFs, OI leaves much to be desired. At best it can produce a meaningful estimate of the first PDF moment only, however with serious limitations related to preferred capabilities such as the nonlinearity of observation operators and dynamical structure of forecast error covariance.

Another fundamental development that led to current variational data assimilation (VAR) methods was the introduction of variational principles in data assimilation by Sasaki (1958). While at the time it was understood as a method for objective analysis based on least squares, the new method for the first time introduced variational formalism and minimization under the geostrophic constraint and also under the more general balance constraint between winds and geopotential. Then, in a trilogy of papers (Sasaki, 1970a, 1970b, 1970c) expanded

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

the previous approach to include the time dependency of observations and established a basis for future development of four-dimensional variational data assimilation (4DVAR) methodology.

While the use of variational principles in data assimilation have been known since the early work of Sasaki (1958), it took almost 25 years before variational methodology had another push into the field of data assimilation, mostly because of the advancements of computers in NWP. Addressing the deficiencies of OI, most importantly the local character of the analysis, nonlinearity of observations, and to some extent the specification of forecast error covariance, variational methods for data assimilation were revived in the mid 1980s (e.g., Lewis and Derber, 1985; Le Dimet and Talagrand, 1986). The subtypes of variational methods include three-dimensional variational (3DVAR) (e.g., Parrish and Derber, 1992) and 4DVAR data assimilation (e.g., Navon et al., 1992; Županski, 1993; Courtier et al., 1994). They include a global minimization (i.e., over all model points) of the cost function that can incorporate nonlinear observations and solves the inversion problem using adjoint equations. The forecast error covariance is improved over OI as it includes complex cross-correlations with additional dynamical balance constraints, but the correlations are still modeled. On the positive side, the modeling of error covariance allows the covariance to be of full rank, meaning that all degrees of freedom (DOF) required for solving the analysis problem are included. The covariance is stationary, although in 4DVAR there is limited capability to introduce time dependence during the assimilation window. Also, variational methods primarily estimate the first moment of PDFs. Although it is possible to estimate the second PDF moment, especially in 4DVAR, there is no feedback of uncertainties from one data assimilation cycle to the next implying a limited use of Bayesian inference. The main advantage of variational methods is their capability to assimilate nonlinear observations, in particular the satellite radiances that now represent the major source of information in meteorology (e.g., Derber and Wu, 1998). By introducing 4DVAR the prediction model itself could be used as a constraint in optimization. The cost of applying VAR has increased compared to previous methods, but it can still be considered efficient since potentially costly matrix inversions are avoided. The variational methods are still used in practice.

Immediately following this development of variational methods, ensemble Kalman filtering (EnKF) methods have been introduced to data assimilation (Evensen, 1994; Houtekamer and Mitchell, 1998). The EnKF successfully addressed the problem of the nonlinear prediction model in the KF by introducing the Monte Carlo approach to the KF forecast step. At the same time the forecast error covariance is dynamic, and is therefore an improvement on the stationary and modeled error covariance used in variational methods. The most important impact of the EnKF was that a realistic data assimilation could be used to produce the first two moments of the PDF, the mean and the covariance, although still under a Gaussian PDF assumption. One of the issues of the EnKF is not being able to account for nonlinearity of observations, since the same linear KF analysis equation is used. More recently (e.g., Sakov et al., 2012), an iterative EnKF was introduced in a manner similar to the iterated KF to address the nonlinearity of observation operators. Implementing the EnKF requires the assimilation of perturbed observations, which results in the calculation of numerous analyses for each ensemble member, and therefore an increase in the cost. Square-root

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

EnKF methods were introduced to reduce the computational cost, by directly calculating the mean of the analysis. Including a large number of ensembles required to resolve a realistic data assimilation problem proved to be practically impossible due to the computational cost and storage requirements. This prompted a need for covariance localization to increase the number of DOF of the low-rank ensemble covariance that could be feasible. This localization greatly helped the EnKF and related ensemble methods to remain of practical significance, although a modification of the dynamically based ensemble forecast error covariance via convolution with a prespecified localizing covariance matrix is required. Covariance localization implies that practical EnKF methods can be interpreted, in terms of forecast error covariance, as an intermediate approach between the full-rank EnKF (with all DOF) and the local OI method. The analysis solution in the EnKF with localization is essentially local since only observations within a certain distance can impact the analysis point.

Both EnKF and variational methods have practical and theoretical limitations. Variational methods have the capability of addressing nonlinearity through applying global numerical optimization. The EnKF is inherently designed to use the linear analysis equation of the KF. An alternative way of bridging this issue was introduced by the maximum likelihood ensemble filter (MLEF) (Županski, 2005), in which it was shown how the calculation of adjoint operators could be avoided by using nonlinear ensemble perturbations and applied in variational-like minimization of the cost function. As with other ensemble methods, MLEF includes the flow-dependent ensemble covariance and estimates the posterior uncertainty.

The implied limitation of error covariance representation in the practical EnKF due to an insufficient number of DOF and to some extent the nondynamical impact of covariance localization, even though the covariance is flow dependent, can result in an analysis that is not of the desired quality. The same could be said for variational methods, where the use of stationary and modeled error covariance is not sufficiently realistic and can produce unsatisfactory analysis. As a result, hybrid ensemble-variational methods that allow a combination of the flow-dependent ensemble, but low-rank covariance, and the stationary variational, but full-rank, error covariance were introduced (Lorenc, 2003; Buehner, 2005; Wang et al., 2007; Bonavita et al., 2012; Clayton et al., 2013).

Data assimilation can also be viewed as an application of Pontryagin's minimum principle (PMP) (e.g., Pontryagin et al., 1961; Lakshmivarahan et al., 2013) where a least squares fit of an idealized path to dynamics law follows from Hamiltonian mechanics. In this application of optimal control theory, the problem is posed as finding the best possible forcing for taking a dynamical system from one state to another, in the presence of dynamical constraints. This forcing is also related to accounting for ME in data assimilation. While the use of forcing reminds us of nudging, the PMP method is more general since it includes an optimization subject to dynamical constraints as well as uncertainties (Lakshmivarahan and Lewis, 2013). Similar to previous methods, it searches for optimal analysis that could be interpreted as the first PDF moment, but estimation of the posterior uncertainties is not an essential part of the method. It is possible to view 4DVAR as a special case of PMP.

The above historical overview also indicates the current status of practical data assimilation development. Other methods with stronger theoretical foundations have been introduced to data assimilation, such as particle filters (PFs) (e.g., van Leeuwen, 2009; Chorin et al., 2010), but they still have limitations for realistic high-dimensional applications.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

However, by directly calculating arbitrary PDFs through the Bayesian framework they have a theoretical advantage in accounting for nonlinearity and non-Gaussianity and therefore offer numerous possibilities for the future development of data assimilation.

1.3 Terminologies and Notation

Data assimilation consists of two major elements – a model of the dynamical system and a set of data (i.e., observations), and aims to procure optimal estimates of model states by combining model forecasts and observations. We represent the model states and the observations in terms of vectors, $\mathbf{x}$ and $\mathbf{y}^o$, respectively. The *true* state, $\mathbf{x}^t$, can never be obtained but can be estimated through an adequate estimation procedure. Such an estimate, made at a given time, is called the *analysis*, $\mathbf{x}^a$. The *estimate* is also denoted by $\hat{\mathbf{x}}$ and is interchangeably used with $\mathbf{x}^a$. The *background*, $\mathbf{x}^b$, is an a priori estimate of $\mathbf{x}^t$ before the analysis is conducted. For the notations in data assimilation, we generally follow Ide et al. (1997).

Data assimilation represents a process to obtain $\mathbf{x}^a$, as close to $\mathbf{x}^t$ as possible, by correcting $\mathbf{x}^b$ using a correction, $\Delta\mathbf{x}$. Mathematically, it is formulated, in its simplest form, as:

$$\mathbf{x}^a = \mathbf{x}^b + \Delta\mathbf{x}. \tag{1.1}$$

Note that $\Delta\mathbf{x}$ is a function of both $\mathbf{y}^o$ and $\mathbf{x}^b$, and it is called an *analysis increment*.

1.3.1 Observation Equation

A variety of observations, assembled in $\mathbf{y}^o$, are used for data assimilation (see Figure 1.1). As observations are much fewer than model states and are irregularly distributed, direct comparison between observations and model states is unfeasible. Thus, we define a nonlinear function, H, called an *observation operator*, that transforms the state vector from the state space, $\mathcal{R}^m$, to the observation vector in the observation space, $\mathcal{R}^n$. The observation is described in terms of the true state as:

$$\mathbf{y}^o = H\mathbf{x}^t + \boldsymbol{\varepsilon}^o, \tag{1.2}$$

where $\boldsymbol{\varepsilon}^o$ is the observation (measurement) error. Equation (1.2) is called the *observation equation* or the *observation model*.

1.3.2 Observation Error Statistics

We assume that the measurement error $\boldsymbol{\varepsilon}^o$ in (1.2) is random and independent, and hence have zero mean, i.e.,

$$mean(\varepsilon_i^o) = E(\varepsilon_i^o) = 0 \ \text{ for } \ i = 1, \ldots, n. \tag{1.3}$$

This implies that $\mathbf{y}^o$ in (1.2) depends only on $\mathbf{x}^t$ and all other variation in $\mathbf{y}^o$ is random. For the random errors, the variance and the covariance of the errors are

$$var(\varepsilon_i^o) = E(\varepsilon_i^o \varepsilon_i^o) = \sigma_i^2 \ \text{ for } \ i = 1, \ldots, n \tag{1.4}$$

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-83176-5 — Principles of Data Assimilation
Seon Ki Park , Milija Zupanski
Excerpt
More Information

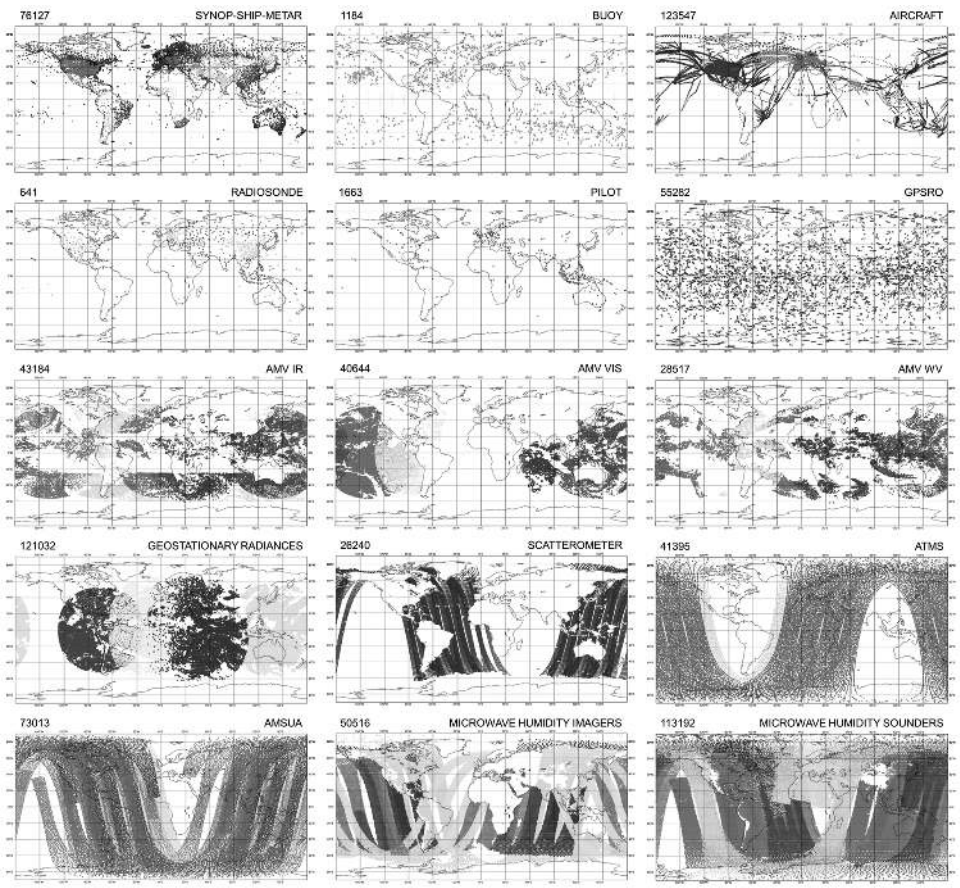

Figure 1.1 Various observation data at the global scale, available on 0000 UTC 18 August 2020, with the observation platforms (top-right corners) and the number of data used for data assimilation (top-left corners). The details of legends in the subfigures refer to the data coverage from the European Centre for Medium-Range Weather Forecasts (ECMWF, 2020). CC BY-NC-ND 4.0 License.

and

$$cov(\varepsilon_i^o, \varepsilon_j^o) = E(\varepsilon_i^o \varepsilon_j^o) = 0 \ \text{ for } \ i, j = 1, \ldots, n \ \text{ and } \ i \neq j, \tag{1.5}$$

respectively. Here, σ_i^2 is the squared standard deviation of ε_i^o and (1.4) assumes that the variance of $\boldsymbol{\varepsilon}^o$ is constant; thus, not dependent on $\mathbf{x}^t$. With zero covariances in (1.5), the variables in $\boldsymbol{\varepsilon}^o$ are uncorrelated with each other. By combining the three assumptions in (1.3)–(1.5), we have

$$\begin{aligned} mean(\boldsymbol{\varepsilon}^o) &= E(\boldsymbol{\varepsilon}^o) = \mathbf{0}, \\ cov(\boldsymbol{\varepsilon}^o) &= E(\boldsymbol{\varepsilon}^o (\boldsymbol{\varepsilon}^o)^T) = \sigma^2 \mathbf{I} = \mathbf{R}. \end{aligned} \tag{1.6}$$

© in this web service Cambridge University Press & Assessment www.cambridge.org