Data Assimilation

William Lahoz · Boris Khattatov ·
Richard Ménard
Editors

Data Assimilation

Making Sense of Observations

Springer

Editors
Dr. William Lahoz
Norsk Institutt for Luftforskning, Norwegian Institute for Air Research
Atmospheric and Climate Research
Instituttveien 18
Kjeller 2007
Norway
wal@nilu.no

Dr. Boris Khattatov
373 Arapahoe Avenue
Boulder CO 80302
USA
boris@fusionnumerics.com

Dr. Richard Ménard
Environment Canada
Atmospheric Science & Technology Directorate
2121 Trans-Canada Highway
Dorval QC H9P 1J3
Canada
richard.menard@ec.gc.ca

ISBN 978-3-540-74702-4 e-ISBN 978-3-540-74703-1
DOI 10.1007/978-3-540-74703-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010925150

Cover design: Bauer, Thomas

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Erik Andersson European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK, erik.andersson@ecmwf.int

Michael G. Bosilovich NASA Goddard Space Flight Center, Greenbelt, MD, USA, Michael.Bosilovich@nasa.gov

Mark Buehner Meteorological Research Division, Data Assimilation and Satellite Meteorology Section, Environment Canada, Canada, mark.buehner@ec.gc.ca

Andrew Charlton-Perez Department of Meteorology, University of Reading, Reading, UK, a.j.charlton@reading.ac.uk

Stephen E. Cohn Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA, stephen.e.cohn@nasa.gov

Gabriëlle J.M. De Lannoy George Mason University, Fairfax, VA, USA; Faculty of Bioscience Engineering, Ghent University, Ghent B-9000, Belgium, gdlannoy@cola.iges.org

Hendrik Elbern Research Centre Jülich, Rhenish Institute for Environmental Research at the University of Cologne (RIU), Cologne, Germany; Helmholtz Virtual Institute for Inverse Modelling of Atmospheric Chemical Composition (IMACCO), Cologne, Germany, he@eurad.uni-koeln.de

G. David Emmitt Simpson Weather Associates (SWA), Charlottesville, VA, USA, gde@swa.com

Quentin Errera Belgian Institute for Space Aeronomy, BIRA-IASB, Brussels, Belgium, quentin@aeronomie.be

Ronald M. Errico NASA/GSFC, Greenbelt, MD, USA; Goddard Earth Science and Technology Center, University of Maryland, Baltimore, MD, USA, Ronald.M.Errico@nasa.gov

Keith Haines Environmental Systems Science Centre, University of Reading, Reading RG6 6AL, UK, kh@mail.nerc-essc.ac.uk

Paul R. Houser George Mason University, Fairfax, VA, USA, phouser@gmu.edu

Xiang-Yu Huang National Center for Atmospheric Research, Boulder, CO, USA, huangx@ucar.edu

Eugenia Kalnay University of Maryland, College Park, MD 20742-2425, USA, ekalnay@atmos.umd.edu

Boris Khatattov Fusion Numerics Inc, Boulder, CO, USA, boris@fusionnumerics.com

William Lahoz Norsk Institutt for Luftforskning, Norwegian Institute for Air Research, NILU, Kjeller, Norway, wal@nilu.no

Stephen R. Lewis Department of Physics and Astronomy, The Open University, Milton Keynes MK7 6AA, UK, S.R.Lewis@open.ac.uk

Stephen J. Lord NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA, Stephen.Lord@noaa.gov

Peter Lynch University College Dublin, Dublin, Ireland, peter.lynch@ucd.ie

Michiko Masutani NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA; Wyle Information Systems, El Segundo, CA, USA, Michiko.Masutani@noaa.gov

Richard Ménard Air Quality Research Division, Meteorological Service of Canada, Environment Canada, Dorval, Canada, richard.menard@ec.gc.ca

Nancy K. Nichols Department of Mathematics, University of Reading, Reading, UK, n.k.nichols@reading.ac.uk

Lars Nieradzik Research Centre Jülich, Rhenish Institute for Environmental Research at the University of Cologne (RIU), Cologne, Germany; Helmholtz Virtual Institute for Inverse Modelling of Atmospheric Chemical Composition (IMACCO), Cologne, Germany, e-mail: ln@eurad.unikoeln.de

Lars-Peter Riishøjgaard NASA/GSFC, Greenbelt, MD, USA; Goddard Earth Science and Technology Center, University of Maryland, Baltimore, MD, USA; Joint Center for Satellite Data Assimilation, Camp Springs, MD, USA, Lars.P.Riishojgaard@nasa.gov

Richard B. Rood University of Michigan, Ann Arbor, MI, USA, rbrood@umich.edu

Thomas W. Schlatter NOAA/Earth System Research Laboratory, Boulder, CO, USA, Tom.Schlatter@noaa.gov

Ad Stoffelen Royal Dutch Meteorological Institute (KNMI), DeBilt, The Netherlands, Ad.Stoffelen@knmi.nl

Achim Strunk Research Centre Jülich, Rhenish Institute for Environmental Research at the University of Cologne (RIU), Cologne, Germany; Helmholtz

Virtual Institute for Inverse Modelling of Atmospheric Chemical Composition (IMACCO), Cologne, Germany, as@eurad.uni-koeln.de

Richard Swinbank Met Office, Exeter, UK, richard.swinbank@metoffice.gov.uk

Olivier Talagrand Laboratoire de Météorologie Dynamique/CNRS, École Normale Supérieure, Paris, France, Talagrand@lmd.ens.fr

Jean-Noël Thépaut European Centre for Medium-Range Weather Forecasts, ECMWF, Shinfield, UK, jean-noel.thepaut@ecmwf.int

Jeffrey P. Walker Department of Civil and Environmental Engineering, The University of Melbourne, Victoria, Australia, j.walker@unimelb.edu.au

John S. Woollen NOAA/NWS/NCEP/EMC, Camp Springs, MD, USA; Science Applications International Corporation (SAIC), Mclean, VA, USA, Jack.Woollen@noaa.gov

Valery Yudin SAIC, Global Modeling Assimilation Office, Code 610.1, Goddard Space Flight Center, Greenbelt, MD 20771, USA; Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, CO, USA, vyudin@ucar.edu

Introduction

William Lahoz, Boris Khattatov, and Richard Ménard

This book came from a request from Springer to the editors to update knowledge on the science of data assimilation and incorporate developments during the last 5 years. It is designed to update the science of data assimilation since the NATO (North Atlantic Treaty Organization) Science Series Book "Data Assimilation for the Earth System" (R. Swinbank, V. Shutyaev, W.A. Lahoz, eds.) came out in 2003, and fill in some of the gaps in that book. The NATO Science Series Book was based on a set of lectures presented at the NATO Advanced Study Institute (ASI) on *Data Assimilation for the Earth System*, which was held at Maratea, Italy during May–June 2002. That ASI grew out of a concern that there was little teaching available in data assimilation, even though it had become central to modern weather forecasting, and was becoming increasingly important in a range of other Earth disciplines such as the ocean, land and chemistry.

Many changes have happened in the science of data assimilation over the last 5 years. They include the development of chemical data assimilation systems at several centres world-wide, both research and operational; the increased interaction between the research and operational communities; the use of data assimilation to evaluate research satellite data; the use of data assimilation ideas, long applied to weather forecast models, to evaluate climate models; the combination of theoretical notions from variational methods and ensemble Kalman filter methods to improve data assimilation performance; and the increased extension of data assimilation to areas beyond the atmosphere and dynamics: chemistry, ionosphere, and other planets, e.g., Mars and Venus. There has also been a consolidation in the use of data assimilation to evaluate future observations, and in the use of data assimilation in areas such as the ocean and the land.

Parallel to these changes in the science of data assimilation, another remarkable change over the last 5 years has been the increased presence of data assimilation in teaching initiatives such as Summer Schools. These include the now biennial ESA (European Space Agency) Earth Observation Summer School

W. Lahoz (✉)
Norsk Institutt for Luftforskning, Norwegian Institute for Air Research, NILU, Kjeller, Norway
e-mail: wal@nilu.no

(http://envisat.esa.int/envschool_2008/) and several others. It can now be said that data assimilation has become a mainstream topic in the teaching of Earth Observation.

The NATO Science Series book, although useful and a feature in many university lecture courses, has some gaps. These include, for example, an overview of data assimilation and its relationship to information, either in observations or models; a discussion of ensemble Kalman filter methods; a discussion of Observing System Simulation Experiments (OSSEs); a discussion of tropospheric chemical data assimilation; and a discussion of meteorology and dynamics.

This book is intended to build on the material from the NATO Science Series book, address the above changes, and fill in the above gaps. Although there will be inevitable gaps in this book, we think it will provide a useful addition to the literature on data assimilation. To achieve this, we have asked world-leading data assimilation scientists to contribute to the chapters. We hope we succeed, at least until the next data assimilation book along these lines comes out in 5 years! Finally, we dedicate this book to Andrew Crook (1958–2006) who was one of the original chapter authors.

November 2009

Part I
Theory

Data Assimilation and Information

William Lahoz, Boris Khattatov, and Richard Ménard

1 Introduction

In this introductory chapter we provide an overview of the connection between the *data assimilation* methodology and the concept of *information*, whether embodied in *observations* or *models*. In this context, we provide a step by step introduction to the need for data assimilation, culminating in an easy to understand description of the data assimilation methodology. Schematic diagrams and simple examples form a key part of this chapter.

The plan is to first discuss the need for *information*; then discuss sources of information; discuss the characteristics of this information, in particular the presence of "information gaps"; provide an objective underpinning to methods to fill in these information gaps; and discuss the benefits of combining different sources of information, in this case from *observations* that sample in space and time the system of interest (e.g. the atmosphere, the ocean, the land surface, the ionosphere, other planets), and *models* that embody our understanding of the system observed. Finally, we bring together these ideas under the heading of "data assimilation", provide a schematic of the methodology, and provide three simple examples highlighting how data assimilation *adds value*, the impact of *spatial resolution* on information, and the impact of *temporal sampling* on information.

At the end of this chapter we identify the foci of this book and the order in which they are presented in the book.

2 Need for Information

The main challenges to society, for example, *climate change*, *impact of extreme weather*, *environmental degradation* and *ozone loss*, require information for an intelligent response, including making choices on future action. Regardless of its

W. Lahoz (✉)
Norsk Institutt for Luftforskning, Norwegian Institute for Air Research, NILU, Kjeller, Norway
e-mail: wal@nilu.no

W. Lahoz et al. (eds.), *Data Assimilation*, DOI 10.1007/978-3-540-74703-1_1,

source, we wish to be able to use this information to make predictions for the future, test hypotheses, and attribute cause and effect. In this way, we are able to take action according to information provided on the future behaviour of the system of interest, and in particular future events (*prediction*); test our understanding of the system, and adjust this understanding according to new information (*hypothesis testing*); and understand the cause of events, and obtain information on possible ways of changing, mitigating or adjusting to the course of events (*attribute cause and effect*).

We can identify a generic *chain of information processing*:

- Gather information;
- Test hypotheses based on this information;
- Build methods to use this information to attribute cause and effect;
- Use these methods to make predictions.

However, we still need two ingredients: a means of gathering information, and methods to build on this information gathered. Roughly speaking, *observations* (measurements) provide the first ingredient, and *models* (conceptual, numerical or otherwise) provide the second ingredient. Note, however, that from the point of view of information, observations and models are not distinct; it is the mechanism of obtaining this information that is distinct: observations have a roughly *direct link* with the system of interest via the measurement process; models have a roughly *indirect link* with the system of interest, being an embodiment of information received from measurements, experience and theory.

3 Sources of Information

We have two broad sources of information: *measurements* of the system of interest ("observations"); and *understanding* of the temporal and spatial evolution of the system of interest ("models"). Further details about observations and models can be found in Part II, *Observations*, and Part III, *Meteorology and Atmospheric Dynamics*, respectively.

Observations (or measurements) sample the system of interest in space and time, with spatial and temporal scales dependent on the technique used to make the measurements. These measurements provide information on the system of interest and contribute to building an understanding of how the system evolves in space and time.

Understanding can be *qualitative*, e.g., how variables roughly "connect" or are related, or *quantitative*, commonly expressed in equations. A rough, qualitative connection can indicate that if the velocity of a particle increases, its kinetic energy also increases. A quantitative connection based on equations assigns a numerical relationship between the velocity and the kinetic energy, so that we can make precise

(subject to the accuracy of the calculation) the increase in kinetic energy given an increase in velocity of the particle. Equations can come from *general laws* (e.g. Newton's laws of motion), or *relations between parameters* (e.g. empirical or statistical). In general, quantification on the basis of laws tends to be more rigorous than quantification on the basis of empirical or statistical relations, mainly because laws have broad (if not universal) application, whereas empirical or statistical relations tend to apply only to specific cases.

4 Characteristics of Information

To make use of the information embodied in observations and models it is necessary to understand the characteristics of this information. In particular, we must recognize that both observations and models have *errors*. We now discuss briefly the nature of these errors.

Observations have errors which are characterized as *random* (also known as precision), *systematic* (also known as bias) and of *representativeness* (or representativity). The sum of these errors is sometimes known as the *accuracy*. Random errors have the property that they are reduced by averaging. Systematic errors, by contrast, are not reduced by averaging; if known, they can be subtracted from an observation. The representativeness error is associated with differences in the resolution of observational information and the resolution of the model interpreting this information.

Models also have errors. These errors arise through the construction of models, as models can be incomplete due to a lack of understanding or due to processes being omitted to make the problem tractable; and through their imperfect simulation of the "real world", itself sampled by observations or measurements. Thus, information, whether in the form of observations or models has errors, and these have to be taken into account. Further details about the nature of observational and model errors can be found in the following chapters in Part I, *Theory*.

Another key feature of observations (or measurements) is that they are discrete in space and time, with the result that the information provided by observations has *gaps* (Fig. 1).

It is desirable to fill gaps in the information provided by observations: first, to make this information more *complete*, and hence more useful; second, to provide information at a *regular scale* to quantify the characteristics of this information. Information at an *irregular scale* can be quantified, but this procedure is more tractable when done with a regular scale.

Assuming a need to fill in the gaps in the observational information, the question is how to do so. Conceptually, it is desirable to use information on the behaviour of the system to extend the observations and fill in the gaps. This information is provided by a model of how the system behaves; this model then allows one to organize, summarize and propagate the information from observations. Note that

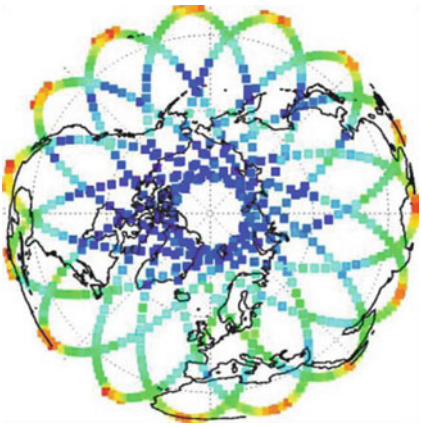

Fig. 1 Plot representing ozone data at 10 hPa (approximately 30 km in altitude) for 1 February 1997 based on the observational geometry of ozone measurements from the MLS (Microwave Limb Sounder) instrument onboard the National Aeronautics and Space Administration (NASA) UARS (Upper Atmosphere Research Satellite) satellite. For information on UARS, see http://mls.jpl.nasa.gov/uars/science.php. *Blue* denotes relatively low ozone values; *red* denotes relatively high ozone values. Note the gaps between the satellite orbits. Thanks to Finn Bjørklid (NILU) for improving this figure

there can be differences in the resolution of the observations, and the resolution of the models used to propagate the information in observations. This will introduce errors when filling in the information gaps.

We now discuss algorithms to fill in the information gaps. The idea is that the algorithm, embedded in a model, provides a set of *consistent* (i.e., mathematically, physically or otherwise) and *objective* (i.e., based on impartial principles) rules which when followed fill in the information gaps associated with observations.

5 Objective Ways of Filling in Information Gaps

What algorithm should one use to fill in the information gaps associated with observations? There are a number of features that such an algorithm should have. The most important ones are that it be *feasible* and that it be *objective* (and *consistent*). From the point of view of feasibility, one could build a hierarchy of algorithms of increasing complexity, starting, for example, with linear interpolation between observations. A simple approach such as linear interpolation is feasible (because simple) and, in cases where observations are dense enough, could be expected to be reasonably accurate. However, although in principle consistent, it is not objective (because not general) and, for example, in general it will not reflect how it is understood systems such as the atmosphere behave. A more realistic approach would be to fill in the gap using a model of how the system behaved. For example, for the

atmosphere, we could use a model that embodies the *equations of motion*; *radiative transfer*; *physical processes* such as convection; and *chemistry*. Such a model would be more expensive to apply than a simple linear interpolation, but in principle would provide a more accurate (and more objective) approach to filling in the information gaps in the observations. In practice, one strikes a balance between using a model that is feasible and using a model that is objective and consistent. Practically, one seeks a model that is *tractable* and *realistic*.

We would like to find methods that allow the interpolation, i.e., filling in of the observational information gaps using a model, to be done in an "intelligent" way. By intelligent, we mean an "objective" way which makes use of concepts for combining information that can be quantified. For example, by finding the *minimum* or *maximum* value of a quantity that can be calculated from the information available. In this way, we can think of the model as an intelligent interpolator of the observation information: intelligent because it embodies our understanding of the system; intelligent because the combination of the observational and model information is done in an objective way. Note that in practice, the model (like the observations) provides information that is discrete in space and time.

Mathematics provides rules for combining information objectively, based on principles which aim to maximize (or minimize) a quantity (e.g. a "penalty function"), or on established *statistical concepts* which relate *prior information* (understanding, which comes from prior combination of observations and models) with *posterior information* (which comes from making an extra observation).

In particular, mathematics provides a foundation to address questions such as: "What combination of the observation and model information is optimal?", and provides an estimate of the errors of the "optimum" or "best" estimate. This is known as "data assimilation" (also as Earth Observation data/model fusion), and has strong links to several mathematical disciplines, including *control theory* and *Bayesian statistics*. The data assimilation methodology adds value to the observations by filling in the observational gaps and to the model by constraining it with observations (Fig. 2 below). In this way, the data assimilation allows one to "make sense" of the observations. Further details about the theory of data assimilation can be found in the following chapters in Part I, *Theory*.

Mathematics also provides an algorithmic basis for applying data assimilation to real problems, including, for example, *weather forecasting*, where data assimilation has been very successful. In particular, over the last 25 years, the skill of weather forecasts has increased – the skill of today's 5-day forecast is comparable to the skill of the 3-day forecast 25 years ago. Furthermore, the skill of forecasts for the Southern Hemisphere is now comparable to that of the Northern Hemisphere (Simmons and Hollingsworth 2002).

Mathematics also provides a *theoretical* and *algorithmic* basis for studying the problem of data assimilation, notably by using simpler models to test ideas. The results using these simpler models can then be used to inform data assimilation developments with complex systems, such as those used for weather forecasting.

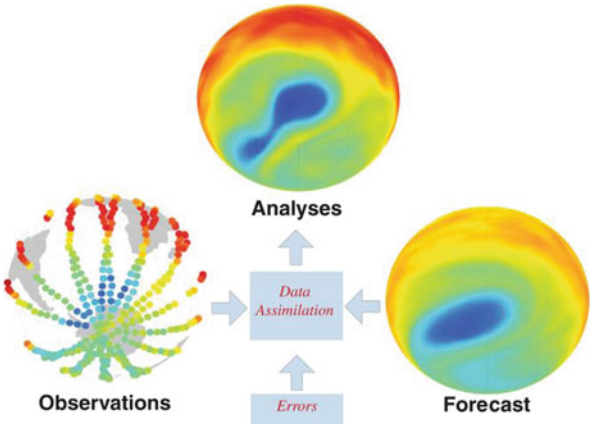

Fig. 2 Schematic of how data assimilation (DA) works and adds value to observational and model information. The data shown are various representations of ozone data at 10 hPa (about 30 km in height) on 23 September 2002. *Lower left panel*, "observations": plot representing the day's ozone data based on the observational geometry of ozone measurements from the MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) instrument onboard the European Space Agency (ESA) Envisat satellite; for information on MIPAS, see http://envisat.esa.int/instruments/mipas/. *Lower right panel*, "forecast": plot representing a 6-day ozone forecast (1200 UTC) based on output from a DA system. *Top panel*, "analyses": plot representing an ozone analysis (1200 UTC) based on output from a DA system. The DA system associated with the lower right plot and the top plot is based on that at the Met Office, and is described in Geer et al. (2006). *Blue* denotes relatively low ozone values; *red* denotes relatively high ozone values. The DA method combines the observations with a model forecast (commonly short-term, e.g., 6 or 12 h), including their errors to produce an ozone analysis. Note how the analysis (*top panel*) fill in the gaps in the observations (*lower left panel*), and the analysis captures the Antarctic ozone hole split (verified using independent data not used in the assimilation) whereas the 6-day forecast (*lower right panel*) does not. In this sense, the DA method adds value to both the observations and the model. Thanks to Alan Geer for providing the basis of this figure and for Finn Bjørklid for improving the figure

6 Simple Examples of Data Assimilation

We now provide three simple examples highlighting how data assimilation *adds value* (Example 1); the impact of *spatial resolution* on information (Example 2); and the impact of *temporal sampling* on information (Example 3).

Example 1 Combining observations with understanding of a system, where both pieces of information have finite errors, should, intuitively, increase the information about the system. There are several ways of quantifying this increase in information, one of them being the error embodied in the information, quantified by the *standard deviation*. We discuss this using a simple example where information from two scalar quantities with *Gaussian* (i.e., normally distributed) errors is combined.

Consider two observations (x_1, x_2) of variable x, with associated variances $({\sigma_1}^2, {\sigma_2}^2)$. Now assume that the observation errors are random, unbiased and normally distributed. It can be shown that the *optimum estimate* ("most probable" value) is given by:

$$x = \frac{\left(\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}\right)}{\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)},$$

with variance:

$$\sigma^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}.$$

We can also see from this example that:

$$\sigma_1 \to \infty, x \to x_2;$$
$$\sigma^2 \leq \min\{\sigma_1^2, \sigma_2^2\}.$$

We can see from this simple example that the error (variance) associated with the combined information is generally lower than the error associated with any of the two pieces of information being combined and that, at worse, it is equal to the minimum of the errors of the individual pieces of information, but never larger. We can also see obvious limiting cases, when the error of one of the pieces of information being combined becomes infinitely large, i.e., the information from this piece becomes vanishingly small. The result in this example can be generalized to two observations $(\mathbf{x}_1, \mathbf{x}_2)$ of a vector variable $\mathbf{x}$, with associated matrix error covariances $(\mathbf{S}_1, \mathbf{S}_2)$.

Although this simple example encapsulates how information is increased, this result concerning variances only holds for Gaussian errors. For errors that are not Gaussian, the variance of the combined information can be larger than that of one of the pieces of information being combined. This apparently counter-intuitive result indicates that variance is not the best way of measuring increases in information. In fact, one must use the concept of *entropy* to consider errors with general probability distributions.

Example 2 Consider a large square room, where temperature measurements are made at each corner. What is the temperature at the centre of the room? What is the temperature representative for the room? These questions concern the *spatial resolution* of information, and how the latter changes as the former changes.

To estimate the temperature at the centre of the room we could average the four corner temperatures, giving each measurement equal weight. This gives the same result assuming the temperature varies linearly between opposite corners and taking an average of the two resulting measurements. Regardless of how the final value is

computed, a *model* of how the temperature varies in the room is needed to compute the temperature at the centre of the room.

To estimate the temperature representative for the room we could proceed as above. In this case we would be averaging the "point" temperature information from each corner to provide "area" temperature information for the whole room. When we use this estimate of the "area" temperature (or any other estimate) as representative of the room temperature, we incur an error of *representativeness*. This was introduced in Sect. 4 above.

The impact of spatial resolution on the estimate for the temperature at the centre of the room can be seen as follows. If we increase the number of measurements in the room, for example along the walls or toward the centre, we tend to get a better estimate of the temperature at the centre of the room, either because we are sampling closer to the room centre, and/or we are obtaining more information of how the temperature varies in the room. Higher spatial observational sampling generally provides (at least initially) better information on the system by reducing the observational gaps. However, there comes a point where we do not get further information, e.g., sampling the temperature at close enough locations in the room gives essentially an unchanged temperature within the error of the measuring device. This illustrates the concept of observational information *saturation* with respect to other observations, where the measurement is no longer *independent* and provides no new information.

The impact of spatial resolution on the estimate for the "area" temperature of the room can be seen as follows. Assume the spatial resolution of the algorithm (i.e., model) used to estimate the "area" temperature remains fixed. As we *reduce* the spatial dimensions of the room the observational gaps become smaller, and the estimate of the "area" temperature as calculated above (or generally using any algorithm or model) initially tends to become *more accurate*. However, there comes a point where, within the error of the algorithm, we do not get further information if we continue reducing the spatial dimension of the observational gaps. We have observational information *saturation* with respect to the model.

Through representation of errors, data assimilation takes account of the spatial resolutions in the model and the observations, and the information saturation between observations, and between the observations and the model.

Example 3 Consider a person walking along a path in the forest, gathering information about their surroundings through their eyes, and keeping their eyes closed for regular intervals. How does this person keep on the path when their eyes are closed? How does the time the person keeps their eyes closed affect their progress along the path? These questions concern the rate at which information is sampled in time, i.e., *temporal sampling*.

The person gathers *observational* information about their surroundings through their eyes: "the path is straight"; "the path curves to the left". This provides information of the path to the person, who then incorporates it into a model of their surroundings. This allows the person to keep along the path when their eyes are closed: "keep straight ahead"; "turn left". When the person next opens their eyes

they can adjust (*correct*) their model of their surroundings depending on the new observational information: "turn right"; "bend down to avoid a low tree branch". The combination of observational and model information allows the person to walk along the path.

However, the amount of time the person keeps their eyes closed affects the quality of observational information they get about their surroundings. If the amount of time is relatively short, say 1 s, the quality of observational information will be relatively high and the person should be able to walk along the path without mishap. By contrast, if the amount of time is relatively long, say 1 min, the quality of observational information will be relatively low and the person would be expected to have problems walking along the path (note, however, that this depends on the nature of the path, see later). This shows how temporal sampling can affect the quality of observational information received, which in turn allows the correction of model information.

If the path is straight, the amount of time the person keeps their eyes closed can be relatively long and still allow them to be able to keep along the path without mishap. This is because the model of the path (built from observational information) is relatively simple: "keep on a straight line", and does not need relatively high temporal sampling to adjust it. Conversely, if the path has many bends without pattern in their handedness, the model of the path (again, built from observational information) is relatively complex: "keep turning in the direction of the path", and needs relatively high temporal sampling to adjust it. This shows how the complexity of the system affects the temporal sampling of observational information needed to adjust (i.e., keep "on track") a model describing the system. The appropriate complexity of a model describing the system depends on the character of the observational information gathered (observation types, errors, spatial resolution, temporal sampling).

Data assimilation, by confronting the model with observations in time and space, keeps the model on track.

7 Benefits of Combining Information

As seen in Fig. 2 above, and the examples in Sect. 6, combining information from observations and a model adds value to both the observations and the model: the information gaps in the observations are filled in; the model is constrained by the observations. Other benefits accrue from "confronting" observations and models, as is done in the data assimilation method. These benefits include the *evaluation* of both the observations and the model. This evaluation of information is crucial in *Earth Observation* (observational information); *Earth System Modelling* (model information, i.e., information which embodies our understanding); and in *melding* observations with a model, which we call "data assimilation" (merging of information). By evaluating information, shortcomings can be identified and remedied, with a consequent improvement in the collection, propagation and use of information.

8 What This Book Is About

This book develops the theme introduced in this chapter, namely, the use of data assimilation to make sense of observations. It has six foci:

- *Theory* (the eight chapters in Part I following this chapter);
- *Observations* (the three chapters in Part II);
- *Meteorology and Atmospheric Dynamics* (the three chapters in Part III);
- *Atmospheric Chemistry* (the four chapters in Part IV);
- *Wider Applications* (the three chapters in Part V);
- *The Longer View* (the three chapters in Part VI).

These foci span several cross-cutting axes: (i) the mathematics of data assimilation; (ii) observations and models; (iii) the activities of the weather centres and the activities of the research community; (iv) the different elements of the Earth System: atmosphere, ocean, land and chemistry; (v) evaluation and production of added-value analyses; and (vi) the success of the data assimilation method and future developments. These are exciting times for data assimilation and we hope this book conveys this excitement.

References

Geer, A.J., W.A. Lahoz, S. Bekki, et al., 2006. The ASSET intercomparison of ozone analyses: Method and first results. *Atmos. Chem. Phys.*, **6**, 5445–5474.

Simmons, A.J. and A. Hollingsworth, 2002. Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc*., **128**, 647–677.

Mathematical Concepts of Data Assimilation

N.K. Nichols

1 Introduction

Environmental systems can be realistically described by mathematical and numerical models of the system dynamics. These models can be used to predict the future behaviour of the system, provided that the initial states of the system are known. Complete data defining all of the states of a system at a specific time are, however, rarely available. Moreover, both the models and the available initial data contain inaccuracies and random noise that can lead to significant differences between the predicted states and the actual states of the system. In this case, observations of the system over time can be incorporated into the model equations to derive "improved" estimates of the states and also to provide information about the "uncertainty" in the estimates.

The problem of state-estimation is an inverse problem and can be treated using observers and/or filters derived by feedback design techniques (see, for example, Barnett and Cameron 1985). For the very large non-linear systems arising in the environmental sciences, however, many traditional state-estimation techniques are not practicable and new "data assimilation" schemes have been developed to generate accurate state-estimates (see, for example, Daley 1993; Bennett 1992). The aim of such schemes can be stated as follows.

> The aim of a data assimilation scheme is to use measured observations in combination with a dynamical system model in order to derive accurate estimates of the current and future states of the system, together with estimates of the uncertainty in the estimated states.

The most significant properties of the data assimilation problem are that the models are very large and non-linear, with order $O(10^7–10^8)$ state variables. The dynamics are multi-scale and often unstable and/or chaotic. The number of observations is also large, of order $O(10^5–10^6)$ for a period of 6 h, but the data are not evenly distributed in time or space and generally have "holes" where there are no

N.K. Nichols (✉)
Department of Mathematics, University of Reading, Whiteknights, Reading, RG6 6AX UK
e-mail: n.k.nichols@reading.ac.uk

W. Lahoz et al. (eds.), *Data Assimilation*, DOI 10.1007/978-3-540-74703-1_2,

observations (see chapter *Data Assimilation and Information*, Lahoz et al.). In practice the assimilation problem is generally ill-posed and the state estimates may be sensitive to errors.

There are two basic approaches to this problem. The first uses a "dynamic observer," which gives a *sequential data assimilation scheme*, and the second uses a "direct observer," which gives a *four-dimensional data assimilation scheme*. In the first case, the observations are "fed-back" into the model at each time these are available and a best estimate is produced and used to predict future states. In the second case a feasible state trajectory is found that best fits the observed data over a time window, and the estimated states at the end of the window are used to produce the next forecast. Under certain mathematical assumptions these processes solve the same "optimal" state-estimation problem. In operational systems, solving the "optimal" problem in "real-time" is not always possible, and many different approximations to the basic assimilation schemes are employed.

In the next section the data assimilation problem is formulated mathematically. In subsequent sections various techniques for solving the assimilation problem are discussed.

2 Data Assimilation for Non-linear Dynamical Systems

A variety of models is used to describe systems arising in environmental applications, as well as in other physical, biological and economic fields. These range from simple linear, deterministic, continuous ordinary differential equation models to sophisticated non-linear stochastic partial-differential continuous or discrete models. The data assimilation schemes, with minor modifications, can be applied to any general model.

We begin by assuming that for any given initial states and given inputs, the equations modelling the dynamical system uniquely determine the states of the system at all future times. This is known as the "perfect" model assumption. In the following subsections we define the data assimilation problem for this case and examine its properties. Next we determine a best linear estimate of the solution to the non-linear assimilation problem. The data assimilation scheme is then interpreted in a stochastic framework and the "optimal" state-estimate is derived using statistical arguments. We consider the case where the model includes errors in the system equations in a later section of this chapter.

2.1 Basic Least-Squares Formulation for Perfect Models

Data assimilation schemes are described here for a system modelled by the discrete non-linear equations

$$\mathbf{x}_{k+1} = \mathcal{M}_{k,k+1}(\mathbf{x}_k), \quad k = 0, \ldots, N-1, \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ denotes the vector of n model states at time t_k and $\mathcal{M}_{k,k+1} : \mathbb{R}^n \to \mathbb{R}^n$ is a non-linear operator describing the evolution of the states from time t_k to time t_{k+1}. The operator contains known inputs to the system including known external forcing functions that drive the system and known parameters describing the system.

Prior estimates, or "background estimates," $\mathbf{x}_0^b$, of the initial states $\mathbf{x}_0$ at time t_0 are assumed to be known, usually provided by a previous forecast.

The observations are assumed to be related to the system states by the equations

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k^{\mathrm{o}}, \quad k = 0, \ldots, N, \tag{2}$$

where $\mathbf{y}_k \in \mathbb{R}^{p_k}$ is a vector of p_k observations at time t_k and $\mathcal{H}_k : \mathbb{R}^n \to \mathbb{R}^{p_k}$ is a non-linear operator that includes transformations and grid interpolations. The observational errors $\boldsymbol{\varepsilon}_k^{\mathrm{o}} \in \mathbb{R}^{p_k}$ consist of instrumentation errors and representativity (or representativeness) errors (see chapter *Data Assimilation and Information*, Lahoz et al.).

For the "optimal" analysis, we aim to find the best estimates $\mathbf{x}_k^a$ for the system states $\mathbf{x}_k$, $k = 0, \ldots, N$, to fit the observations $\mathbf{y}_k$, $k = 0, \ldots, N$, and the background state $\mathbf{x}_0^b$, subject to the model equations (1). We write the problem as a weighted non-linear least-squares problem constrained by the model equations.

Problem 1 Minimize, with respect to $\mathbf{x}_0$, the objective function

$$\begin{aligned} J = &\frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^T \mathbf{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \\ &+ \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k), \end{aligned} \tag{3}$$

subject to $\mathbf{x}_k$, $k = 1, \ldots, N$, satisfying the system equations (1) with initial states $\mathbf{x}_0$.

The model is assumed here to be "perfect" and the system equations are treated as *strong constraints* on the minimization problem. The states $\mathbf{x}_k$ that satisfy the model equations (1) are uniquely determined by the initial states and therefore can be written explicitly in terms of $\mathbf{x}_0$. Substituting into the objective function (3) then allows the optimization problem to be expressed in terms of the initial states alone. The assimilation problem, Problem 1, thus becomes an unconstrained weighted least-squares problem where the initial states are the required control variables in the optimization.

The weighting matrices $\mathbf{B}_0 \in \mathbb{R}^{n \times n}$ and $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$, $k = 0, 1 \ldots, N$, are taken to be symmetric and positive definite and are chosen to give the problem a "smooth" solution. They represent, respectively, the uncertainty in the background states (prior estimates) and the observations. The objective function (3) can then be written in the compact form:

$$J(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{f}(\mathbf{x}_0)\|_2^2 \equiv \frac{1}{2}\mathbf{f}(\mathbf{x}_0)^T \mathbf{f}(\mathbf{x}_0), \tag{4}$$

where

$$\mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{B}_0^{-1/2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) \\ \mathbf{R}_0^{-1/2}(\mathcal{H}_0(\mathbf{x}_0) - \mathbf{y}_0) \\ \vdots \\ \mathbf{R}_N^{-1/2}(\mathcal{H}_N(\mathbf{x}_N) - \mathbf{y}_N) \end{pmatrix}, \tag{5}$$

and $\mathbf{x}_k = \mathcal{M}_{0,k}(\mathbf{x}_0)$, $k = 1, \ldots, N$, satisfy the system equations (1) with initial states $\mathbf{x}_0$ at time t_0 (see Lawless et al. 2005). The matrices $\mathbf{B}_0^{-1/2}$ and $\mathbf{R}_k^{-1/2}$ denote the inverses of the symmetric square roots of $\mathbf{B}_0$ and $\mathbf{R}_k$, respectively.

In this approach the initial states are treated as parameters that must be selected to minimize the weighted mean square errors between the observations predicted by the model and the measured observations over the time window and between the initial and background states. The initial state is adjusted to different positions in order to achieve the best fit, using an efficient iterative minimization algorithm.

2.2 Properties of the Basic Least-Squares Formulation

The solution $\mathbf{x}_0^a$ to the least-squares problem (4) is known as the *analysis*. The analysis may not be well-defined if $\mathbf{B}_0^{-1} = 0$, that is, if no background state is specified. In that case the number and locations of the observations may not be sufficient to determine all the degrees of freedom in the optimization problem; in other words, the system may not be "observable." If the weighting matrix $\mathbf{B}_0$ is non-singular, however, then, provided the operators $\mathcal{M}_{0,k}$ and $\mathcal{H}_k$ are continuously differentiable, the stationary points of the least-squares problem are well-defined. The weighted background term acts as a "regularization" term, ensuring the existence of a solution and also damping the sensitivity of the solution to the observational errors (Johnson et al. 2005a, b).

Under these conditions, the stationary points of the objective function (4) satisfy the gradient equation, given by

$$\nabla_{\mathbf{x}_0} J = \mathbf{J}^T \mathbf{f}(\mathbf{x}_0) = 0, \tag{6}$$

where $\mathbf{J}$ is the Jacobian of the vector function $\mathbf{f}$ defined in (5). The Jacobian can be written in the compact form

$$\mathbf{J} = \begin{pmatrix} \mathbf{B}_0^{-1/2} \\ \hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}} \end{pmatrix}, \; \hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0,1} \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{0,N} \end{pmatrix}, \tag{7}$$

where $\hat{\mathbf{R}} = \text{diag}\{\mathbf{R}_k\}$ is a block diagonal matrix containing the weighting matrices $\mathbf{R}_k$ on the diagonal. The matrices $\mathbf{M}_{0,k}$ and $\mathbf{H}_k$ denote the Jacobians of the model and observation operators $\mathcal{M}_{0,k}$ and $\mathcal{H}_k$, respectively; that is,

$$\mathbf{M}_{0,k} = \left.\frac{\partial \mathcal{M}_{0,k}}{\partial \mathbf{x}}\right|_{\mathbf{x}_0}, \quad \mathbf{H}_k = \left.\frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}\right|_{\mathcal{M}_{0,k}(\mathbf{x}_0)}.$$

If $\mathbf{B}_0$ is non-singular, then the Jacobian $\mathbf{J}$, given by (7), is of full rank and the stationary points satisfying the gradient equation (6) are well-defined. Stationary points are not unique, however, and may not yield a minimum of the non-linear assimilation problem. If a stationary point is such that the Hessian $\nabla^2_{\mathbf{x}_0} J$, of the objective function (3) (or equivalently (4)) is positive-definite at that point, then the stationary point is a local minimum of the assimilation problem (see Gratton et al. 2007). It should be noted that multiple local minima of the assimilation problem may exist.

We remark that the sensitivity of the analysis to small perturbations in the data depends on the "conditioning" of the Hessian, $\nabla^2_{\mathbf{x}_0} J$, that is, on the sensitivity of the inverse of the Hessian to small perturbations. If small errors in the Hessian lead to large errors in its inverse, then the computed solution to the data assimilation problem may be very inaccurate. In designing data assimilation schemes, it is important, therefore, to ensure that the conditioning of the Hessian is as small as feasible, or to use "preconditioning" techniques to improve the conditioning.

2.3 Best Linear Least-Squares Estimate

In general, explicit solutions to the non-linear data assimilation problem, Problem 1, cannot be found. A "best" *linear* estimate of the solution to the non-linear problem can, however, be derived explicitly. We assume that the departure of the estimated analysis $\mathbf{x}_0^a$ from the background $\mathbf{x}_0^b$ is a *linear* combination of the innovations $\mathbf{d}_k = \mathbf{y}_k - \mathcal{H}_k\left(\mathbf{x}_k^b\right)$, $k = 0, 1, \ldots, N$, and find the estimate for $\mathbf{x}_0^a$ that solves the least-squares data assimilation problem as accurately as possible.

To determine the estimate, we linearize the assimilation problem about the non-linear background trajectory $\mathbf{x}_k^b = \mathcal{M}_{0,k}\left(\mathbf{x}_0^b\right)$, $k = 1, \ldots, N$. We denote by the matrices $\mathbf{H}_k$ and $\mathbf{M}_{0,k}$ the linearizations of the observation and model operators $\mathcal{H}_k$ and $\mathcal{M}_{0,k}$, respectively, about the background trajectory; that is,

$$\mathbf{H}_k = \left.\frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}\right|_{\mathbf{x}_k^b}, \quad \mathbf{M}_{0,k} = \left.\frac{\partial \mathcal{M}_{0,k}}{\partial \mathbf{x}}\right|_{\mathbf{x}_0^b}.$$

The linearized least-squares objective function is then given by

$$\tilde{J} = \frac{1}{2}\boldsymbol{\delta}\mathbf{x}_0^T \mathbf{B}_0^{-1} \boldsymbol{\delta}\mathbf{x}_0 + \frac{1}{2}\sum_{k=0}^{N} (\mathbf{H}_k \mathbf{M}_{0,k} \boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}_k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{M}_{0,k} \boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}_k), \qquad (8)$$

where $\boldsymbol{\delta}\mathbf{x}_0 = \left(\mathbf{x}_0 - \mathbf{x}_0^b\right)$. Using the compact form of the Jacobian (7), the gradient equation of the linearized problem may be written

$$\begin{aligned}\nabla_{\mathbf{x}_0}\tilde{J} &= \mathbf{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \\ &\quad + \sum_{k=0}^{N}(\mathbf{H}_k\mathbf{M}_{0,k})^T\mathbf{R}_k^{-1}\left(\mathbf{H}_k\mathbf{M}_{0,k}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) - \left(\mathbf{y}_k - \mathcal{H}_k\left(\mathbf{x}_k^b\right)\right)\right) \\ &= \left(\mathbf{B}_0^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}}\right)\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{d}} \\ &= 0,\end{aligned} \tag{9}$$

where $\hat{\mathbf{d}} = \left(\mathbf{d}_0^T, \mathbf{d}_1^T, \ldots, \mathbf{d}_N^T\right)^T$ is the vector of innovations.

The optimal *linear* state-estimate for $\mathbf{x}_0^a$ is then the solution to the gradient equation (9) and is given by

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \hat{\mathbf{K}}\hat{\mathbf{d}}, \tag{10}$$

where

$$\hat{\mathbf{K}} = \left(\mathbf{B}_0^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1} \equiv \mathbf{B}_0\hat{\mathbf{H}}^T\left(\hat{\mathbf{H}}\mathbf{B}_0\hat{\mathbf{H}}^T + \hat{\mathbf{R}}\right)^{-1}. \tag{11}$$

The matrix $\hat{\mathbf{K}}$ is known as the *gain* matrix.

For systems where the model and observation operators are linear, the analysis (10) and (11) is an exact, unique, stationary point of the data assimilation problem, Problem 1. For non-linear systems multiple stationary points of the objective function (3) may exist and the analysis (10) and (11) is only a first order approximation to an optimal solution, due to the linearization of the non-linear model and observation operators.

The Hessian of the linearized objective function (8) at the analysis (10) and (11) is given by

$$\nabla^2_{\mathbf{x}_0}\tilde{J} = \left(\mathbf{B}_0^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}}\right). \tag{12}$$

If $\mathbf{B}_0$ is non-singular, then the matrix (12) is symmetric and positive-definite and (10) and (11) provides the "best" linear estimate of the minimum of the data assimilation problem, Problem 1, in a region of the state space near to the background.

2.4 Statistical Interpretation

The data assimilation problem, as formulated in Problem 1, determines a least-squares fit of the model predictions to the observations, subject to constraints. An estimate of the "uncertainty" in this analysis would be valuable. If additional assumptions about the stochastic nature of the errors in the initial state estimates and

the observations are made, then the solution to the data assimilation problem can be interpreted in statistical terms and the uncertainty in the analysis can be derived.

To obtain a statistical formulation of the data assimilation problem, we assume that the errors $\left(\mathbf{x}_0–\mathbf{x}_0^b\right)$ between the true initial states $\mathbf{x}_0$ and the prior background estimates $\mathbf{x}_0^b$ are randomly distributed with mean zero and covariance matrix $\mathbf{B}_0 \in \mathbb{R}^{n\times n}$. The observational errors $\boldsymbol{\varepsilon}_k^{\mathrm{o}} \in \mathbb{R}^{p_k}$, $k = 0, \ldots, N$, defined in (2), are assumed to be unbiased, serially uncorrelated, randomly distributed vectors with zero means and covariance matrices $\mathbf{R}_k \in \mathbb{R}^{p_k\times p_k}$. The observational errors and the errors in the prior estimates are assumed to be uncorrelated.

Under these basic statistical assumptions, given the prior estimates $\mathbf{x}_0^b$, and the observations $\mathbf{y}_k$, $k = 0, \ldots, N$, the "best linear unbiased estimate," or BLUE, of the true state $\mathbf{x}_0$ at time t_0 equals the best least-squares estimate (10) and (11) for the analysis $\mathbf{x}_0^a$. The uncertainty in this estimate is described by the analysis error covariance, which is given by

$$\mathbf{A} = (\mathbf{I}_n - \hat{\mathbf{K}}\hat{\mathbf{H}})\mathbf{B}_0. \tag{13}$$

Over all linear combinations of the innovations of form (10), the BLUE minimizes the analysis error covariance and is thus the solution to the assimilation problem with *minimum variance*. The analysis given by (10) and (11) is therefore the "optimal" linear estimate in this sense.

In addition to the basic statistical assumptions, the errors in the prior estimates and in the observations are commonly assumed to have Gaussian probability distributions, which are fully defined by the means and covariances specified. In this case, the solution to the data assimilation problem, Problem 1, is equal to the *maximum a posteriori Bayesian estimate* of the system states at the initial time. From Bayes Theorem we have that the posterior probability of $(\mathbf{x}_0 - \mathbf{x}_0^b)$, given the departures from the observations $(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))$, $k = 0, \ldots, N$, satisfies

$$\begin{aligned}\rho\Big(\mathbf{x}_0 - \mathbf{x}_0^b|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k),\ k = 0, \ldots, N\Big) = \\ = \alpha\rho\Big(\mathbf{x}_0 - \mathbf{x}_0^b\Big)\,\rho\Big(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k),\ k = 0, \ldots, N|\mathbf{x}_0 - \mathbf{x}_0^b\Big),\end{aligned} \tag{14}$$

where $\rho\big(\mathbf{x}_0 - \mathbf{x}_0^b\big)$ is the prior probability of $\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)$ and $\rho(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)$, $k = 0, \ldots, N|\mathbf{x}_0 - \mathbf{x}_0^b\big)$ is the conditional joint probability of $(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))$, $k = 0, \ldots, N$, given $\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)$. The scalar α is a normalizing constant that ensures that the value of the posterior probability is not greater than unity. The "optimal" analysis is then the initial state that maximizes the posterior probability.

From the assumption that the probability distributions are Gaussian, we have that

$$\rho\left(\mathbf{x}_0 - \mathbf{x}_0^b\right) \propto \exp\left[-\frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^b\right)\right]$$

and

$$\rho(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)) \propto \exp\left[-\frac{1}{2}(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))^T \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))\right],$$

for $k = 0, 1, \ldots, N$. Taking the log of the posterior probability and using the assumptions that the observational errors are uncorrelated in time and uncorrelated with the background errors, we find that

$$\begin{aligned} J(\mathbf{x}_0) &\equiv -\ln\left[\rho\big(\mathbf{x}_0 - \mathbf{x}_0^b | \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k), k = 0, \ldots, N\big)\right] \\ &= -\ln\left[\rho\big(\mathbf{x}_0 - \mathbf{x}_0^b\big)\right] - \sum_{k=0}^{N} \ln\left[\rho\big(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\big)\right]. \end{aligned} \tag{15}$$

(See Lorenc 1986, 1988.) The solution $\mathbf{x}_0$ to the data assimilation problem, Problem 1, that minimizes $J(\mathbf{x}_0)$ is therefore equivalent to the maximum Bayesian a posteriori likelihood estimate.

If the model and observation operators are linear and the errors are normally distributed (i.e., Gaussian), then the *maximum a posteriori Bayesian estimate* and the *minimum variance estimate* are equivalent. The BLUE, given explicitly by (10) and (11), with zero mean and covariance (13), is thus the unique optimal in both senses.

In practice the error distributions may not be Gaussian and the assumptions underlying the estimates derived here may not hold. Ideally, we would like to be able to determine the full probability distributions for the true states of the system given the prior estimates and the observations. This is a major topic of research and new approaches based on sampling methods and particle filters are currently being developed.

Techniques used in practice to solve the data assimilation problem, Problem 1, include sequential assimilation schemes and variational assimilation schemes. These methods are described in the next two sections.

3 Sequential Data Assimilation Schemes

We describe sequential assimilation schemes for discrete models of the form (1), where the observations are related to the states by the Eq. (2). We make the *perfect model assumption* here. We assume that at some time t_k, *prior* background estimates $\mathbf{x}_k^b$ for the states are known. The differences between the observations of the true states and the observations predicted by the background states at this time, $\left(\mathbf{y}_k - \mathcal{H}\left(\mathbf{x}_k^b\right)\right)$, known as the innovations, are then used to make a correction to the background state vector in order to obtain improved estimates $\mathbf{x}_k^a$, known as the analysis states. The model is then evolved forward from the analysis states to the next time t_{k+1} where observations are available. The evolved states of the system at the time t_{k+1} become the background (or forecast) states and are denoted by $\mathbf{x}_{k+1}^b$. The background is then corrected to obtain an analysis at this time and the process is repeated.