Journal of the Meteorological Society of Japan, Vol. 75, No. 1B, pp. 181–189, 1997

Unified Notation for Data Assimilation: Operational, Sequential and Variational

By Kayo Ide[1]

Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1565, USA

Philippe Courtier[2]

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK

Michael Ghil

Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1565, USA

and

Andrew C. Lorenc

Meteorological Office, Bracknell RG12 2SZ, UK

(Manuscript received 7 August 1995, in revised form 16 July 1996)

Abstract

The need for unified notation in atmospheric and oceanic data assimilation arises from the field's rapid theoretical expansion and the desire to translate it into practical applications. Self-consistent notation is proposed that bridges sequential and variational methods, on the one hand, and operational usage, on the other. Over various other mottoes for this risky endeavor, the authors selected: *"When **I** use a word," Humpty Dumpty said, in rather a scornful voice tone, "it means just what I choose it to mean — neither more nor less."* Lewis Carroll, 1871.

1. Introduction

Model-based assimilation of observations, or data assimilation for short, has evolved into a major area of academic research in dynamic meteorology and physical oceanography, as well as of operational numerical prediction for atmospheric and oceanic flows (Panel, 1991). Its major sources of theoretical and practical ideas include by now the engineering (Bucy and Joseph, 1987; Gelb, 1974; Jazwinski, 1970), mathematical (Gill *et al.*, 1981; Lions, 1971; Marchuk, 1975) and geophysical (Bennett, 1992; Daley, 1991; Ghil and Malanotte-Rizzoli, 1991) literatures. The Second WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography, held in Tokyo in March 1995, has shown substantial progress in the practical implementation of advanced data-assimilation methods, based on both the control-theoretical and sequential-estimation approach. These two approaches, because of their different heritage, have sometimes used very different sets of notation, which obscure the fact that they attempt to solve, in complementary ways, similar problems. This has hindered mutual understanding amongst developers and users of data-assimilation methods, provided a stumbling block in the operational application of novel methods based on either approach or a combination of both, and hampered further theoretical

1 Corresponding author

2 Present affiliation: CNES, 2, place Maurice Quentin, 75039 Paris Cedex 01, France

advances.

Notation is the language of the sciences and differing sources of notation guarantee a Tower of Babel that impedes effective communication amongst the theoreticians and practitioners of data assimilation. It also renders the literature more difficult to master for newcomers, whether students or researchers from other fields. The authors are keenly aware that no single notation is ideal and can make every user happy. This "modest proposal" is made with the hope that it will catch on, and thereby help advance data assimilation in both meteorology and oceanography, by facilitating the mastery of past and present methods based on either approach and fostering the development of future, more advanced methods, based on either combinations of these two approaches or on entirely novel approaches.

Basic concepts and definitions appear in Section 2. Recommended, self-consistent notation following the sequential approach is outlined in Section 3 and following the variational approach in Section 4. Methods based on various combinations of these two approaches are under investigation and we hope to contribute to the study and implementation of such combined methods by suggesting unified notation. Conclusions and a summary table follow in Section 5. An Appendix includes tentative recommendations for continuous space-and-time notation, at one of the research frontiers of the field. This paper is restricted, by its nature and limitations of space, to present basic concepts and notation, which are merely illustrated by a few examples. For a more exhaustive presentation of the rapidly increasing number of assimilation methods, and their complete description, we refer to the books and review papers already cited, and to those included in this volume or cited below.

2. Basic concepts and definitions

A discrete model for the evolution of an atmospheric, oceanic, or coupled system from time t_i to time t_{i+1} is governed by an equation

$$\mathbf{x}^{\mathrm{f}}(t_{i+1}) = M_i[\mathbf{x}^{\mathrm{f}}(t_i)], \tag{1}$$

where $\mathbf{x}$ and M are the model's state vector and its corresponding dynamics operator, respectively. The state vector $\mathbf{x}$ has dimension n. The dynamics M of the model evolution (1) in a computer simulation or prediction is commonly nonlinear and deterministic (see also the Appendix), while the true geofluid may differ from (1) by random or systematic errors. The state vector $\mathbf{x}$ is obtained usually by discretization of the full partial differential equations that we assume govern the flow, using finite differences, finite elements or (pseudo-)spectral methods. The error covariance matrix associated with $\mathbf{x}$ is given by $\mathbf{P}$. The theoretically interesting case of continuous, rather than discrete, space and time is treated in the Appendix.

Observations $\mathbf{y}^{\mathrm{o}}$ at time t_i are defined by

$$\mathbf{y}_i^{\mathrm{o}} = H_i[\mathbf{x}^{\mathrm{t}}(t_i)] + \boldsymbol{\varepsilon}_i, \tag{2}$$

where H is an observation operator, and $\boldsymbol{\varepsilon}$ is a noise process. The observation vector $\mathbf{y}_i^{\mathrm{o}} \equiv \mathbf{y}^{\mathrm{o}}(t_i)$ has dimension p_i. A major problem of data assimilation is that, typically, $p_i \ll n$. The operator H can also be nonlinear, like M, and both can contain explicit time dependence — denoted here by the subscript i — in addition to the implicit dependence via the state vector $\mathbf{x}_i^{\mathrm{f}} \equiv \mathbf{x}^{\mathrm{f}}(t_i)$. The noise process $\boldsymbol{\varepsilon}$ is commonly assumed to have zero mean and we denote its covariance matrix by $\mathbf{R}$; it consists of instrumental and representativeness errors whose covariance matrices are $\mathbf{E}$ and $\mathbf{F}$, respectively, with $\mathbf{R} = \mathbf{E} + \mathbf{F}$. More generally, $\boldsymbol{\varepsilon}_i$ may have a nonzero, time- and state-dependent mean, and the covariance matrices $\mathbf{E}$, $\mathbf{F}$ and $\mathbf{R}$ may all depend on t_i or $\mathbf{x}_i$.

In general, subscripts are used to: i) denote the discrete-time index, as in Eqs. (1, 2) above; ii) refer to the corresponding space-related indices, on a regular or irregular grid or for spectral coefficients; and iii) indicate (in parentheses) the vector (*e.g.*, state or observation) with respect to which an error covariance matrix is defined (see Section 3 for an example). Superscripts, on the other hand, are used to: i) describe mathematical operations, such as $(\cdot)^{\mathrm{T}}$, $(\cdot)^{-1}$, $(\cdot)^{-\mathrm{I}}$, and $(\cdot)'$ for transpose, inverse, generalized inverse, and linearization, respectively; and ii) refer to the nature of vectors or matrices in the data-assimilation process, such as $(\cdot)^{\mathrm{a}}$, $(\cdot)^{\mathrm{b}}$, $(\cdot)^{\mathrm{f}}$, and $(\cdot)^{\mathrm{o}}$ for analysis, background (often called "first guess," but see Remarks 3.1.4, 4.1.3 and 4.1.4 below), forecast, and observation. Upper- and lower-case characters in superscripts indicate (i) (some) mathematical operations and (ii) the type of the variable, respectively. Furthermore, $(\cdot)'$ may be dropped when the original operator is linear.

3. Sequential methods

As mentioned already in Section 1, we cannot cover here all the methods inspired by the sequential-estimation approach to data assimilation. Two methods, the extended Kalman filter (EKF: Gelb, 1974; Ghil and Malanotte-Rizzoli, 1991) and so-called optimal interpolation (OI: Daley, 1991) have been selected, on the theoretical and operational side, to illustrate the suggested notation. As more advanced methods develop, we hope that this notation is flexible enough to grow with the methods.

3.1 Extended Kalman filter (EKF)

The discretized dynamics of the true geofluid $\mathbf{x}^{\mathrm{t}}$ is assumed to differ from that of the numerical model (1) by stochastic perturbations

$$\mathbf{x}^{\mathrm{t}}(t_{i+1}) = M_i[\mathbf{x}^{\mathrm{t}}(t_i)] + \boldsymbol{\eta}(t_i), \tag{3}$$

where $\boldsymbol{\eta}$ is a noise process with zero mean and covariance matrix $\mathbf{Q}$; the perturbations can be thought of as representing subgrid-scale processes not resolved by the model (1). The same comments about the possible state- and time-dependence of $\mathbf{Q}_i$ and of a nonzero mean for $\boldsymbol{\eta}_i$ apply as for $\mathbf{R}_i$ and $\boldsymbol{\varepsilon}_i$.

The EKF consists of a forecast step

$$\mathbf{x}^{\mathrm{f}}(t_i) = M_{i-1}[\mathbf{x}^{\mathrm{a}}(t_{i-1})], \tag{4a}$$
$$\mathbf{P}^{\mathrm{f}}(t_i) = \mathbf{M}_{i-1}\mathbf{P}^{\mathrm{a}}(t_{i-1})\mathbf{M}_{i-1}^{\mathrm{T}} + \mathbf{Q}(t_{i-1}), \tag{4b}$$

followed by an update step — traditionally called "analysis" step in numerical weather forecasting — in which the observation available at time t_i is blended with the previous information, carried forward by the forecast step (4a,b).

$$\mathbf{x}^{\mathrm{a}}(t_i) = \mathbf{x}^{\mathrm{f}}(t_i) + \mathbf{K}_i\mathbf{d}_i, \tag{5a}$$
$$\mathbf{P}^{\mathrm{a}}(t_i) = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}^{\mathrm{f}}(t_i). \tag{5b}$$

Here we introduced the linearizations $\mathbf{M} \equiv M'$ and $\mathbf{H} \equiv H'$, while

$$\mathbf{d}_i = \mathbf{y}_i^{\mathrm{o}} - H_i[\mathbf{x}^{\mathrm{f}}(t_i)] \tag{5c}$$

defines the observational residual (in the older meteorological literature) or innovation vector (in the engineering and some of the more recent data assimilation literature). The Kalman gain $\mathbf{K}_i$ is provided by

$$\mathbf{K}_i = \mathbf{P}^{\mathrm{f}}(t_i)\mathbf{H}_i^{\mathrm{T}}[\mathbf{H}_i\mathbf{P}^{\mathrm{f}}(t_i)\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i]^{-1}. \tag{6}$$

Remarks:

1. Observations are processed whenever available, and then discarded. If no observations are available at time t_i, $\mathbf{H}_i = \mathbf{0} = \mathbf{K}_i$ and the analysis step is omitted.

2. When the observation operator H or the model M are linear, $H = \mathbf{H}$ or $M = \mathbf{M}$, one may use operator (italics) or matrix (bold) notation, as the context dictates. Otherwise, the linearization may be about the instantaneous trajectory $\mathbf{x}^{\mathrm{a,f}}(t_i)$ (in the true EKF) or about other states $\tilde{\mathbf{x}}_i$ approximating it, such as a steady state or a state updated less frequently than every time step (in various simplified versions of the EKF).

3. In practice, various simplifications are introduced to describe $\mathbf{P}$ (*e.g.*, Section 3.2 here), in order to overcome — at least in part — the computational burden involved in the matrix calculations of Eqs. (4, 5).

4. In operational practice one may imagine the use of an innovation vector $\mathbf{d}_i$ in which $\mathbf{x}^{\mathrm{f}}(t_i)$ of Eq. (5c) is replaced by another "background state" $\mathbf{x}^{\mathrm{b}}$, obtained for instance by averaging over an ensemble of forecasts, which also produces an approximation of $\mathbf{P}^{\mathrm{f}}$ by $\mathbf{P}^{\mathrm{b}}$ (Evensen, 1994). In the work of Gandin (1963), who introduced the OI methodology to dynamic meteorology, $\mathbf{x}^{\mathrm{b}}$ was a climatological state. Rutherford (1972) proposed $\mathbf{x}^{\mathrm{b}} = \mathbf{x}^{\mathrm{f}}$, as numerical forecasts had become by then better background fields for objective analysis than climatology. This led, however, to the currently common confusion between forecast, first-guess and background fields.

3.2 Optimal interpolation (OI)

OI was, until recently, the most widespread scheme in operational use for weather prediction (Lorenc, 1986; Thiébaux and Pedder, 1987) and is coming into growing use in oceanographic data assimilation (Carton and Hackert, 1989; Derber and Rosati, 1989). It is a particular suboptimal filter, in which the EKF's error covariance matrix $\mathbf{P}^{\mathrm{f}}$ is replaced by an approximation, $\mathbf{B}$; this can be expressed as a product of variances, placed in a diagonal matrix $\mathbf{D}$, and of correlations, placed in a matrix $\mathbf{C}$ with unit diagonal (see Ghil and Malanotte-Rizzoli, 1991, and references there):

$$\mathbf{B} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}. \tag{7}$$

In order to get some of the benefits of the EKF, a simplified prediction scheme for the error variances only is introduced, while $\mathbf{C}$ is kept constant in time:

$$\mathbf{x}^{\mathrm{f}}(t_{i+1}) = M[\mathbf{x}^{\mathrm{a}}(t_i)], \tag{8a}$$
$$\mathbf{D}^{\mathrm{f}}(t_{i+1}) = N[\mathbf{D}^{\mathrm{a}}(t_i)]; \tag{8b}$$

various simple schemes N have been implemented to compute $\mathbf{D}^{\mathrm{f}}$. The analysis step is then

$$\mathbf{x}^{\mathrm{a}}(t_i) = \mathbf{x}^{\mathrm{f}}(t_i) + \mathbf{K}_i^{\mathrm{OI}}\mathbf{d}_i, \tag{9a}$$
$$\mathbf{B}^{\mathrm{a}}(t_i) = (\mathbf{I} - \mathbf{K}_i^{\mathrm{OI}}\mathbf{H}_i)\mathbf{B}^{\mathrm{f}}(t_i) \times (\mathbf{I} - \mathbf{K}_i^{\mathrm{OI}}\mathbf{H}_i)^{\mathrm{T}} + \mathbf{K}_i^{\mathrm{OI}}\mathbf{R}_i(\mathbf{K}_i^{\mathrm{OI}})^{\mathrm{T}}, \tag{9b}$$
$$\mathbf{D}^{\mathrm{a}}(t_i) = \mathrm{diag}[\mathbf{B}^{\mathrm{a}}(t_i)]. \tag{9c}$$

The suboptimal gain $\mathbf{K}_i^{\mathrm{OI}}$ is provided by replacing $\mathbf{P}^{\mathrm{f}}$ with $\mathbf{B}^{\mathrm{f}}$ in Eq. (6), *i.e.*,

$$\mathbf{K}_i^{\mathrm{OI}} = \mathbf{B}^{\mathrm{f}}(t_i)\mathbf{H}_i^{\mathrm{T}}[\mathbf{H}_i\mathbf{B}^{\mathrm{f}}(t_i)\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i]^{-1}. \tag{10}$$

The particular form of statistical interpolation (Daley, 1991) commonly referred to as OI is usually reserved for practical implementations which further approximate Eqs. (7)–(10). Two salient features of OI, as implemented at major numerical weather prediction centers in the early 1980s, are: i) A local approximation in Eq. (9a), by either selecting only a few observations near each grid point that is updated (*e.g.*, McPherson *et al.*, 1979) or by decomposing the domain into regular, small subdomains and neglecting the "influence" of observations outside

each subdomain (*e.g.*, Lorenc, 1981). ii) A resulting approximation in Eqs. (9b, 10) by using direct evaluation of a correlation model $\mathbf{C}$ at the selected observation locations to yield $\mathbf{H}_i\mathbf{B}^{\mathrm{f}}(t_i)\mathbf{H}_i^{\mathrm{T}}$.

Remarks:

1. Many suboptimal filters are currently being developed, attempting to improve upon OI at a lesser computational cost than the EKF (*e.g.*, Todling and Cohn, 1994). We recommend to follow the usage of Eq. (10) in designating the particular suboptimal gain matrix involved by a capital Roman superscript.

2. OI may easily be extended to interpolation of fields over space and time ("four-dimensional OI"). The resulting analysis vector may be interpreted as the minimizer of an expected error covariance, or as the minimizer of a quadratic objective function (see below). Thus it must also satisfy Euler-Lagrange equations. Bennett *et al.* (1997) solve such equations using representers, which are shown to be related to the forecast error covariances $\mathbf{B}^{\mathrm{f}}$; in fact the representer matrix is just $\mathbf{H}\mathbf{B}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}$.

4. Variational methods

The two methods selected here for illustrating the notation are the adjoint method for the time-dependent strong-constraint problem (known widely as "4D-Var") and a recent, "incremental" version thereof. 4D-Var is, in some sense, the theoretical counterpart of the EKF within the family of control-theoretically inspired methods. Incremental 4D-Var represents a particular attempt at its operational implementation.

4.1 Four-dimensional variational assimilation (4D-Var)

4D-Var minimizes the objective function J that measures the weighted sum of squares of distances J^{b} to the background state $\mathbf{x}^{\mathrm{b}}$ and J^{o} to the observations $\mathbf{y}^{\mathrm{o}}$ distributed over a time interval $[t_0, t_n]$,

$$J[\mathbf{x}(t_0)] = \frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)]^{\mathrm{T}}\mathbf{B}_0^{-1}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)] + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{y}_i - \mathbf{y}_i^{\mathrm{o}})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{y}_i^{\mathrm{o}}), \quad (11)$$

where we are using the notation $\mathbf{y}_i \equiv H_i[\mathbf{x}(t_i)]$. Here $\mathbf{B}^{-1}$ is an *a priori* weight matrix, with $\mathbf{B}$ meant to approximate the error-covariance matrix of $\mathbf{x}^{\mathrm{b}}$, and the minimization of (11) is done with respect to the initial state $\mathbf{x}(t_0)$. The formulation (11) reflects the imposition of a *strong constraint* (Sasaki, 1970). Alternative formulations that only impose a *weak constraint* and their connection to various smoothers derived from sequential-estimation results are given by Bennett (1992) and by Ménard and Daley (1996).

Efficient methods for performing the minimization of J require its partial derivatives with respect to the elements of $\mathbf{x}(t_0)$, given by

$$\left[\frac{\partial J}{\partial \mathbf{x}(t_0)}\right]^{\mathrm{T}} = \mathbf{B}_0^{-1}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)] + \sum_{i=0}^{n}\mathbf{M}(t_{i+1}, t_0)^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{y}_i^{\mathrm{o}}), \quad (12a)$$

where

$$\mathbf{M}(t_i, t_0)^{\mathrm{T}} = \prod_{j=0}^{i-1}\mathbf{M}(t_{j+1}, t_j)^{\mathrm{T}} \quad (12b)$$

and $\mathbf{M}(t_{i+1}, t_i) \equiv \mathbf{M}_i$. This follows from

$$\left[\frac{\partial J}{\partial \mathbf{x}_{i+1}^{\mathrm{f}}}\right]^{\mathrm{T}} = \mathbf{M}(t_{i+1}, t_i)^{\mathrm{T}}\left[\frac{\partial J}{\partial \mathbf{x}_i^{\mathrm{f}}}\right]^{\mathrm{T}}, \quad (13a)$$

$$\left[\frac{\partial J}{\partial \mathbf{x}_i}\right]^{\mathrm{T}} = \mathbf{H}_i^{\mathrm{T}}\left[\frac{\partial J}{\partial \mathbf{y}_i}\right]^{\mathrm{T}}. \quad (13b)$$

$\mathbf{M}(t_{i+1}, t_i)^{\mathrm{T}}$ is usually called the adjoint model, and $\mathbf{H}_i^{\mathrm{T}}$ the adjoint observation operator; to be precise, both are adjoints of the suitably linearized dynamics and observation operator, respectively. When using partial derivatives with respect to elements of a discretized representation, as above, the adjoint is identical to the transpose. We recommend using the $(\cdot)^{\mathrm{T}}$ notation in such cases; the $(\cdot)^*$ notation is recommended for adjoints of (partial differential) operators acting on continuous fields (see Appendix) or when the gradient is defined with respect to an inner product using a positive-definite weight matrix distinct from the identity.

Remarks:

1. All vectors are considered to be column vectors unless transposed, while the derivative of a scalar J with respect to a (column) vector $\mathbf{x}$ is a row vector, thus requiring the transpose on the left-hand side of Eq. (12a). For function — rather than vector — spaces (see Appendix), adjoint operators act on the dual space of that which contains the state functions.

2. 4D-Var reduces to three-dimensional variational assimilation (3D-Var) if the time dimension is taken out. A number of operational weather forecast centers are considering a replacement of OI by 3D-Var.

3. The notation $\mathbf{x}^{\mathrm{b}}$, as mentioned already in Remark 4 of Section 3.1, is different from $\mathbf{x}^{\mathrm{a}}$ or $\mathbf{x}^{\mathrm{f}}$ since it is meant, in general, to be the best estimate of the current state $\mathbf{x}(t)$ prior to using the observations at time t. In 3D-Var, $\mathbf{x}^{\mathrm{b}}$ will commonly be the result of a short-range forecast — hence the similarity to "sequential" methods — while in 4D-Var it may be the result of a previous 4D-Var assimilation.

4. The variational algorithm (11)–(13) requires an initial vector for the minimization of (11), *i.e.*, a first guess $\mathbf{x}^{\mathrm{g}}$ of the result. If the minimization problem (11) has a unique minimum then its solution does not depend on the first guess $\mathbf{x}^{\mathrm{g}}$. A very general sufficient condition for the objective function $J(\mathbf{x})$ to have a unique minimum is that it be strictly convex. The only case for which this condition is easy to verify, when the dimension n of $\mathbf{x}$ is large, is that of J being quadratic, which implies that the dynamics and observations are linear. Otherwise the linearization (12a,b) will always depend on the background vector $\mathbf{x}^{\mathrm{b}}$ about which it is carried out. Most current implementations use $\mathbf{x}^{\mathrm{g}} = \mathbf{x}^{\mathrm{b}}$ and some use the terms "background" and "(first) guess" interchangeably. We shall see in the next (sub)section that the distinction does become useful.

4.2 Incremental 4D-Var

We introduce the increment $\delta\mathbf{x}(t_0) \equiv \mathbf{x}(t_0) - \mathbf{x}^{\mathrm{g}}(t_0)$; within the tangent linear approximation, 4D-Var is equivalent to first order to its incremental formulation:

$$\begin{aligned} J[\delta\mathbf{x}(t_0)] = &\frac{1}{2}\{(\delta\mathbf{x}(t_0) - [\mathbf{x}^{\mathrm{b}}(t_0) - \mathbf{x}^{\mathrm{g}}(t_0)]\}^{\mathrm{T}} \\ &\times\mathbf{B}_0^{-1}\{\delta\mathbf{x}(t_0) - [\mathbf{x}^{\mathrm{b}}(t_0) - \mathbf{x}^{\mathrm{g}}(t_0)]\} \\ &+\frac{1}{2}\sum_{i=0}^{n}[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i]^{\mathrm{T}} \\ &\times\mathbf{R}_i^{-1}[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i], \end{aligned} \tag{14}$$

with $\delta\mathbf{x}(t_i) = \mathbf{M}(t_i, t_0)\delta\mathbf{x}(t_0)$ and $\mathbf{d}_i = \mathbf{y}_i^{\mathrm{o}} - H_i[\mathbf{x}^{\mathrm{g}}(t_i)]$. The linearization M' is carried out about the forecast (1) started from $\mathbf{x}^{\mathrm{g}}(t_0)$. The analysis is then obtained by adding the analysis increments $\delta\mathbf{x}^{\mathrm{a}}(t_0)$ to the background,

$$\mathbf{x}^{\mathrm{a}}(t_0) = \mathbf{x}^{\mathrm{g}}(t_0) + \delta\mathbf{x}^{\mathrm{a}}(t_0),$$

where $\delta\mathbf{x}^{\mathrm{a}}(t_0)$ is obtained by minimizing (14).

To reduce the computational cost of the minimization, a linear simplification operator $\mathbf{S}$ is introduced to act on the initial states $\mathbf{x}^{\mathrm{b}}(t_0)$ and $\mathbf{x}^{\mathrm{g}}(t_0)$. It can be a projection operator [see, for instance Ghil (1989) and references there] — such as spectral truncation to large-scale or other selected modes, in which case $\mathbf{S}^2 = \mathbf{S}$ — or some averaging operator for a finite-difference model. In any case, we define $\delta\mathbf{w} \equiv \mathbf{S}\delta\mathbf{x}$ and $\mathbf{S}$ is meant to be rank-deficient, in order to reduce the dimension of the (approximate) minimization problem to be solved, so that one needs to define an appropriate generalized inverse, denoted by $\mathbf{S}^{-\mathrm{I}}$. Simplified dynamics $\mathbf{L}$ is introduced to approximate $\mathbf{SMS}^{-\mathrm{I}}$, as well as a simplified observation operator $\mathbf{G}$, approximating $\mathbf{HS}^{-\mathrm{I}}$. Both are obtained through linearization in the vicinity of the trajectory $\mathbf{x}^{\mathrm{g}}(t_i)$. The simplified minimization problem becomes

$$\begin{aligned} J[\delta\mathbf{w}(t_0)] = &\frac{1}{2}\{\delta\mathbf{w}(t_0) - \mathbf{S}[\mathbf{x}^{\mathrm{b}}(t_0) - \mathbf{x}^{\mathrm{g}}(t_0)]\}^{\mathrm{T}} \\ &\times\mathbf{B}_{(\mathbf{w})}^{-1}(t_0)\{\delta\mathbf{w}(t_0) - \mathbf{S}[\mathbf{x}^{\mathrm{b}}(t_0) - \mathbf{x}^{\mathrm{g}}(t_0)]\} \\ &+\frac{1}{2}\sum_{i=0}^{n}[\mathbf{G}_i\delta\mathbf{w}(t_i) - \mathbf{d}_i]^{\mathrm{T}} \\ &\times\mathbf{R}_i^{-1}[\mathbf{G}_i\delta\mathbf{w}(t_i) - \mathbf{d}_i], \end{aligned} \tag{15}$$

with $\delta\mathbf{w}(t_i) = \mathbf{L}(t_i, t_0)\delta\mathbf{w}(t_0)$ and $\mathbf{B}_{(\mathbf{w})}$ approximating $\mathbf{SB}_{(\mathbf{x})}\mathbf{S}^{\mathrm{T}}$. Here $\mathbf{L}(t_i, t_0) = \Pi_{j=0}^{i-1}\mathbf{L}_j$ and the analysis is obtained through

$$\mathbf{x}^{\mathrm{a}}(t_0) = \mathbf{x}^{\mathrm{g}}(t_0) + \mathbf{S}^{-\mathrm{I}}\delta\mathbf{w}^{\mathrm{a}}(t_0). \tag{16}$$

Remarks:

1. When an ambiguity may arise from the use of different spaces for the model or observation states (*e.g.*, temperatures vs. radiances in the atmosphere, sound arrival times vs. densities in the ocean, fewer observations p_i than grid variables n, or the use of preconditioning or simplification transformations), a subscript may be added to the matrices $\mathbf{P}$ or $\mathbf{B}$. To distinguish this from discrete-time and -space indices, it is recommended to enclose these subscripts in parentheses, *e.g.*,

$$\mathbf{B}_{(\mathbf{w})}(t_i) \cong \mathbf{SB}_{(\mathbf{x})}(t_i)\mathbf{S}^{\mathrm{T}}. \tag{17}$$

2. The well-known duality that holds, for linear dynamics and observations, between the Kalman filter and an optimal control problem (Kalman, 1960; Ghil and Malanotte-Rizzoli, 1991) can be used to combine incremental 4D-Var with the EKF for another promising suboptimal filter.

3. The minimization is not performed in practice in the $\delta\mathbf{w}$ space: a change of variables is introduced to improve the conditioning of the minimization problem. Given the change of variable $\delta\mathbf{w} = \mathbf{Uv}$ and its appropriately defined generalized inverse $\mathbf{v} = \mathbf{U}^{-\mathrm{I}}\delta\mathbf{w}$, the minimization is performed in the $\mathbf{v}$ space (Courtier *et al.*, 1994).

5. Conclusions

A consistent set of symbols has been introduced for advanced assimilation methods, based on both the sequential and variational approach. The duality of the approaches demands and also facilitates the use of unified notation, such as the matrix $\mathbf{B}$ that is used as an approximation to the correct error-covariance matrix $\mathbf{P}$ in sequential methods, while $\mathbf{B}^{-1}$ is used as an *a priori* set of weights in variational ones.

Table I. Notations by type

1. Vectors (bold lower-case Roman)

Symbol	Definition	Remarks
$\mathbf{d}$	innovation vector	**D91**; $\mathbf{h}$ in B92; η in GM91
$\mathbf{x}$	state vector	Any appropriate notation can be used to represent specific variables: **B92**, **C94**, **L86**; $\mathbf{s}$ in D91; $\mathbf{w}^{a,f,t}$ in GM91
$\delta\mathbf{x}$	incremental state vector	δu in C94
$\mathbf{y}^o$	observations	any appropriate notation can be used to represent specific variables: **L86**; $\mathbf{d}$; f_o in D91, $\mathbf{y}$ in C94; $\mathbf{w}^o$ in GM91
$\mathbf{y}$	estimated observation values, calculated from the state vector $\mathbf{x}$	**L86**
$\boldsymbol{\eta}$	model error ("system noise")	$\mathbf{q}$ in B92; $\boldsymbol{\varepsilon}$ in D91; $\mathbf{b}^t$ in GM91
$\boldsymbol{\varepsilon}$	observational error	**B92**, **C94**, **D91**; $\mathbf{b}^o$ in GM91
$\mathbf{w}$	state vector in simplified 4D-Var	
$\delta\mathbf{w}$	incremental state vector in simplified 4D-Var	
$\mathbf{v}$	preconditioned state variable	

2. Operator (upper-case Italic)

Symbol	Definition	Remarks
H	observation operator	**C94**, **D91**, **GM91**; K in L86; $\mathcal{L}$ in B92
J	cost (penalty) function	**C94**, **GM91**; I in D91; $\mathcal{I}$ in B92
M	model operator	**C94**, **D91**; $\mathcal{L}$ in B92; N in GM91
N	scheme to advance variances in OI	
S	simplification operator	

3. Matrix (bold upper-case Roman)

Symbol	Definition	Remarks
$\mathbf{B}$	approximate error covariance	**C94**, **D91**, **L86**; $\mathbf{S}$ in GM91
$\mathbf{C}$	time-independent correlation matrix in OI	**D91**, **GM91**
$\mathbf{D}$	empirical forecast-error variance in OI	
$\mathbf{E}$	covariance of instrumental observation errors	$\mathbf{O}$ in L86
$\mathbf{F}$	covariance of representativeness errors in the observations	**L86**
$\mathbf{G}$	simplified observation operator	
$H' \equiv \mathbf{H}$	linearized observation operator	**C94**
$\mathbf{I}$	identity matrix	
$\mathbf{K}$	gain matrix	**B92**, **C94**, **D91**, **GM91**
$\mathbf{L}$	symplified model operator	
$M' \equiv \mathbf{M}$	linearized dynamics operator	$\mathbf{R}$ in C94, $\boldsymbol{\Psi}$ in GM91
$\mathbf{P}$	forecast-error covariance	**D91**, **GM91**, $\mathbf{c}$ in B92
$\mathbf{Q}$	model-error covariance	**B92**, **D91**, **GM91**
$\mathbf{R}$	observational error covariance	**GM91**; $\mathbf{w}$ in B92; $\mathbf{O}$ in C94, D91
$S' \equiv \mathbf{S}$	simplification matrix	
$\mathbf{U}$	preconditioning matrix	

4. Superscripts*

Symbol	Definition	Remarks
$(\cdot)^{-1}$	inverse	
$(\cdot)^{-\mathrm{I}}$	generalized inverse	to be defined unambiguously (*e.g.*, B92)
$(\cdot)^{\mathrm{T}}$	transpose	
$(\cdot)^{*}$	adjoint	used only when not equivalent to the transpose
$(\cdot)^{\mathrm{OI},\ldots}$	indicates the specific suboptimal scheme used to approximate the EKF	**GM91**
$(\cdot)'$	linearized (operator)	
$(\cdot)^{\mathrm{a}}$	analysis	**C94**, **GM91**; $(\cdot)^{+}$ in B92; $(\cdot)^{\mathrm{a}}$ in D91
$(\cdot)^{\mathrm{b}}$	background	**C94**, $(\cdot)_{B)}$ in D91.
$(\cdot)^{\mathrm{f}}$	forecast	**C94**, **GM91**; $(\cdot)^{\mathrm{f}}$ in D91
$(\cdot)^{\mathrm{g}}$	(first) guess in iteration	**C94**
$(\cdot)^{\mathrm{o}}$	observed	**GM91**; **d** in B92; $(\cdot)_0$ in D91; **y** in C94; $\mathbf{y}^{\mathrm{o}}$ in L86. (see also Table I.1.)
$(\cdot)^{\mathrm{t}}$	true	**GM91**; $(\cdot)_t$ in C94 and L86; $(\cdot)_T$ in D91.

* upper-case Roman for mathematical operations, or specific assimilation schemes (if necessary); lower-case Roman for the nature of vectors or matrices

5. Subscripts

Usage:	1.	to describe discrete-time index;
	2.	to denote the corresponding space-related indices on a finite-difference grid or for spectral coefficients; and
	3.	(in parenthesis) to refer to the corresponding vector when used for a covariance matrix.

Table 1 lists the symbols defined in this article. Among many earlier works that have made important contributions to the field of data assimilation, only a handful are referenced here to show other conventions in wide use prior to this article. B92, C94, D91, GM91, and L86 are abbreviations for Bennett (1992), Courtier *et al.* (1994), Daley (1991), Ghil and Malanotte-Rizzoli (1991), and Lorenc (1986), respectively, listed in alphabetical order; bold lettering for a reference indicates that it uses the same notation as recommended here.

Acknowledgments

We would like to thank the 205 participants at the International Symposium for highlighting the need for unified notation. The data-assimilation work of MG and KI is supported by NASA grant NAG5-713 and ONR grant N00014-93-1-0673. The Appendix was written in response to comments from A.F. Bennett. We also benefited greatly from comments by R.M. Errico, C. Wunsch and two anonymous referees. Carole Edis at ECMWF and Kristin Hartman at UCLA typed the manuscript. This is publication no. 4858 of UCLA's Institute of Geophysics and Planetary Physics.

Appendix

Notation for distributed−parameter systems

The evolution of atmospheric and oceanic flows is governed by systems of partial differential, rather than ordinary difference equations: PDEs rather than OΔEs. In practice, of course, the nonlinearity of these PDEs precludes their analytic solution, except in a few simple cases, and discretization — in both space (P $\rightarrow$ O) and time (D $\rightarrow$ Δ) — is necessary to obtain and study solutions. Still, the PDEs are used in various theoretical investigations of the flow evolution. To the extent that data assimilation is in a phase of rapid theoretical development, at least minimal guidance on notation at these frontiers might do some good and certainly little harm.

In the engineering literature, finite-dimensional systems, in discrete (OΔEs) or continuous (ODEs) time, are referred to as *lumped-parameter* systems, while infinite-dimensional ones (PDEs) are referred to as *distributed-parameter* systems. In the mathematical and geophysical literature, the term "parameter" is reserved usually for a constant in the equations, rather than a dependent variable, so one

talks about *field* estimation, rather than *vector* estimation: the instantaneous state of the system is described by a (set of) field(s) — *i.e.*, by a function(s) of one or more independent space variables — rather than by a vector.

On the whole, continuous dependence on time tends to simplify notation, as well as the analysis of such theoretical issues as observability, controllability and stability for sequential estimation. The counterpart of Eqs. (1,2) in the main text, for a time-continuous finite-dimensional system (ODEs) is:

$$\dot{\mathbf{x}}^{\mathrm{f}} = \mathcal{M}(\mathbf{x}^{\mathrm{f}}, t), \tag{A.1}$$

$$\mathbf{y}^{\mathrm{o}}(t) = \mathcal{H}(\mathbf{x}^{\mathrm{t}}, t) + \boldsymbol{\varepsilon}, \tag{A.2}$$

where $\dot{(\ )} \equiv d(\ )/dt$, $\mathcal{M}$ and $\mathcal{H}$ are nonlinear operators in continuous time, and the interpretation of $\mathbf{x}$, $\mathbf{y}$, and $\boldsymbol{\varepsilon}$ is accordingly different.

The real geofluid is assumed, in the sequential approach, to evolve according to a stochastic differential equation:

$$d\mathbf{x}^{\mathrm{t}} = \mathcal{M}(\mathbf{x}^{\mathrm{t}}, t)dt + d\boldsymbol{\eta}^{\mathrm{t}}; \tag{A.3}$$

here $\boldsymbol{\eta}$ is a Wiener process (Brownian motion) with independent, normally distributed increments, having mean zero and covariance matrix $\mathbf{Q} \equiv E[\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}]$. The expectation operator E is recommended over the bracket or overbar notation, $E[\mathbf{b}^{\mathrm{T}}] \equiv \langle \mathbf{a}\mathbf{b}^{\mathrm{T}} \rangle \equiv \overline{\mathbf{a}\mathbf{b}^{\mathrm{T}}}$, for greater legibility and to emphasize the integral character of this operator (with respect to a suitably defined probability measure). The covariance matrices of $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are still denoted by $\mathbf{P}$ and $\mathbf{R}$, for simplicity.

Since the observations $\mathbf{y}^{\mathrm{o}}$ do occur, typically, at discrete times, the EKF forecast step between two arbitrarily spaced observation times is

$$\dot{\mathbf{x}}^{\mathrm{f}} = \mathcal{M}(\mathbf{x}^{\mathrm{f}}, t), \tag{A.4a}$$

$$\dot{\mathbf{P}}^{\mathrm{f}} = \mathbf{M}^{\mathrm{f}}\mathbf{P}^{\mathrm{f}} + (\mathbf{P}^{\mathrm{f}})^{\mathrm{T}}\mathbf{M}^{\mathrm{f}} + \mathbf{Q}, \tag{A.4b}$$

with $\mathbf{M}(t) = \mathcal{M}'(t)$ (*e.g.*, Ide and Ghil, 1997a,b; Jazwinski, 1970). The linearization in the EKF itself is about $\mathbf{x}^{\mathrm{f}}(t)$, and various suboptimal filters with linearization about $\mathbf{x}^{\mathrm{b}}(t)$ may be considered. Since it is unlikely that both time-continuous and time-discrete versions of the EKF will be used in the same paper, it seems advisable to minimize the number of distinct types of font used. For those interested in the theoretical developments, the distinction between $\mathbf{M}(t_{i+1}, t_i)$ in Section 4.2 and $\mathbf{M}(t)$ here should be obvious.

When observations do occur, at $t = t_j$, say, an analysis step is performed:

$$\mathbf{x}^{\mathrm{a}}(t_j) = \mathbf{x}^{\mathrm{f}}(t_j) + \mathbf{K}_j\{\mathbf{y}_j^{\mathrm{o}} - \mathcal{H}[\mathbf{x}^{\mathrm{f}}(t_j)]\},$$

$$\mathbf{P}^{\mathrm{a}}(t_j) = (\mathbf{I} - \mathbf{K}_j\mathbf{H}_j)\mathbf{P}^{\mathrm{f}}(t_j). \tag{A.5}$$

Here $\mathbf{H}_j = \mathcal{H}'$ and the gain matrix $\mathbf{K}_j$ is given by

$$\mathbf{K}_j = \mathbf{P}^{\mathrm{f}}(t_j)\mathbf{H}_j^{\mathrm{T}}[\mathbf{H}_j\mathbf{P}^{\mathrm{f}}(t_j)\mathbf{H}_j^{\mathrm{T}} + \mathbf{R}_j]^{-1}. \tag{A.6}$$

The details, as usual, are in the references, but the parallelism between Eqs. (A.1)–(A.6) here and (1)–(6) in Section 3 is helpful. Some care needs to be exercised, however, in the reinterpretation of the symbols, since stochastic processes in continuous time, especially nonlinear ones, differ in many respects from those in discrete time (*e.g.*, Miller *et al.*, 1994, and references there). To the extent that no such subtleties arise in the time-continuous strong-constraint variational approach, the counterpart of Eqs. (11)–(13) is left as an exercise to the interested reader [**Hint**: Keep $\mathbf{B}_0$ and $\mathbf{M}(t)$, replace summation by integration and $(\cdot)^{\mathrm{T}}$ by $(\cdot)^*$].

Continuous dependence on space is important to keep in mind when analyzing in depth issues of interpolation and subgrid-scale variability (Bennett, 1992). Unfortunately, substantial mathematical difficulties arise even in linear (Curtain and Pritchard, 1978) or deterministic (Lions, 1971) systems of PDEs. Considerable attention goes to the appropriate function spaces in which the fields of interest lie, and the duals of these function spaces on which the adjoint operators that arise in minimization are defined.

Staying on the purely formal level, like much of the distributed-parameter literature in engineering, one can rewrite Eqs. (A.1)–(A.6) for a field $\mathbf{u}(\mathbf{r}, t)$, where $\mathbf{u}$ now designates the dependent variables – such as velocity, geopotential, temperature and so on — and $\mathbf{r}$ the independent space variables. The model dynamics $\mathcal{M}$ becomes a nonlinear partial differential operator, $\mathcal{M} = \mathcal{M}[\partial_{\mathbf{r}}^{\alpha}, \mathbf{u}(\mathbf{r}, t), \mathbf{r}, t]$, where $\partial_{\mathbf{r}}^{\alpha}$ denotes partial derivatives with respect to the components of $\mathbf{r}$ up to a certain order, $|\alpha| \leq m$. The observation operator $\mathcal{H}[\mathbf{u}(\mathbf{r}, t), \mathbf{r}, t]$ maps typically an infinite-dimensional into a finite-dimensional space (Wahba, 1990):

$$\mathbf{v}_k^{\mathrm{o}}(\mathbf{r}, t) = \mathcal{H}_k[(\mathbf{u}(\mathbf{r}, t), \mathbf{r}, t] + \mathbf{w}(\mathbf{r}, t). \tag{A.7}$$

Here $\mathbf{v}$ is used to denote the observed field(s). Observations occur at discrete locations $\mathbf{r}_k$; alternatively, the observations can be, or include, a finite number of integrals over the domain, at certain radiative wavelengths (in the atmosphere) or sound frequencies (in the ocean). In the latter case, the label k refers to the radiative or sound channels in which the observation is made.

The covariance matrices $\mathbf{P}$ and $\mathbf{Q}$ of Eqs. (1)–(6) and (A.1)–(A.6) become symmetric (or self-adjoint), positive-definite operators. Their spectral, eigenvalue-eigenfunction decomposition is given by the infinite-dimensional version of the Karhunen-Loève theorem (*e.g.*, Loève, 1978).

The objective function $J[\mathbf{x}(t_0)]$ in Section 4.1 becomes a functional $\mathcal{J}[\mathbf{u}(\mathbf{r}, t_0)]$ and its minimization involves the adjoint PDE for a given linearization

of $\mathcal{M}(\mathbf{u}, \mathbf{x}, t)$ (Lions, 1971; Marchuk, 1975). The combined complexities of nonlinearity and stochastic perturbations are challenging but the motivation for addressing them in the PDE setting is strongly enhanced by the promises of increased resolution in both models and observations.

References

Bennett, A.F., 1992: *Inverse Methods in Physical Oceanography.* Cambridge University Press, Cambridge/New York, 346 pp.

Bennett, A.F., B.S. Chua and L.M. Leslie, 1997: Generalized inversion of a global numerical weather prediction model, II: analysis and implementation. *Meteor. Atmos. Phys.*, in press.

Bucy, R.S. and P.D. Joseph, 1987: *Filtering for Stochastic Processes with Applications to Guidance* (2nd ed.). Chelsea, New York, 217 pp.

Carton, J.A. and E.C. Hackert, 1989: Applications of multi-variate statistical objective analysis to the circulation in the tropical Atlantic. *Dyn. Atm. Oceans*, **13**, 491–515.

Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1387.

Curtain, R.F. and A.J. Pritchard, 1978: *Infinite Dimensional Linear Systems Theory.* Lecture notes in Control and Information Sciences, **8**, Springer-Verlag, New York, 297 pp.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, Cambridge, England, 457 pp.

Derber, J.C. and A. Rosati, 1989: A global ocean data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347.

Evensen, G., 1994: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97**, 17905–17924.

Gandin, L.S., 1963: *Objective Analysis of Meteorological Fields.* Gidrometeorol. Izd. Leningrad (in Russian), (English Translation by Israel Program for Scientific Translations, Jerusalem, 1965).

Gelb, A. (Ed.), 1974: *Applied Optimal Estimation.* M.I.T. Press, Cambridge, MA, 374 pp.

Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. *Dyn. Atmos. Oceans*, **13**, 171–218.

Ghil, M. and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141–266.

Gill, P.E., W. Murray and M.H. Wright, 1987: *Practical Optimization.* Academic Press, London.

Ide, K. and M. Ghil, 1997a: Extended Kalman filtering for vortex systems. Part I: Methodology and point vortices. *Dyn. Atmos. Oceans*, in press.

Ide, K. and M. Ghil, 1997b: Extended Kalman filtering for vortex systems. Part II: Rankine vortices and observing-system design. *Dyn. Atmos. Oceans*, in press.

Jazwinski, A.H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, London, 376 pp.

Kalman, R.E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME*, **82D**, 35–45.

Lions, J.L., 1971: *Optimal Control of Systems Governed by Partial Differential Equations.* Springer-Verlag, Berlin, 396 pp.

Loève, M., 1978: *Probability Theory.* Springer-Verlag, Berlin, 413pp.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701–721.

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.

Marchuk, G.I., 1975: Formulation of theory of perturbations for complicated models. *Appl. Math. Optimization*, **2**, 1–33.

McPherson, R.D., K.H. Bergman, R.E. Kistler, G.E. Rasch and D.S. Gordon, 1979: The NMC operational global data assimilation system. *Mon. Wea. Rev.*, **107**, 1445–1461.

Ménard, R. and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221–237.

Miller, R.N., M. Ghil and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056.

Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research, 1991: *Four-Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences.* National Academy Press, Washington, D.C., 78 pp.

Rutherford, I.D., 1972: Data assimilation by statistical interpolation for forecast error fields. *J. Atmos. Sci.*, **29**, 809–815.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875–883.

Thiébaux, H.J. and M.A. Pedder, 1987: *Spatial Objective Analysis.* Academic Press, London, 299 pp.

Todling, R. and S.E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.*, **122**, 2530–2557.

Wahba, G. 1990: *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics, Philadelphia, 169 pp.