Journal of the Meteorological Society of Japan, Vol. 75, No. 1B, pp. 191–209, 1997

Assimilation of Observations, an Introduction

By Olivier Talagrand

Laboratoire de Météorologie Dynamique du CNRS, Paris, France

(Manuscript received 13 November 1995, in revised form 11 March 1997)

Abstract

Assimilation of meteorological or oceanographical observations can be described as the process through which all the available information is used in order to estimate as accurately as possible the state of the atmospheric or oceanic flow. The available information essentially consists of the *observations* proper, and of the *physical laws* which govern the evolution of the flow. The latter are available in practice under the form of a *numerical model*. The existing assimilation algorithms can be described as either *sequential* or *variational*. The links between these algorithms and the *theory of statistical estimation* are discussed. The performances of present algorithms, and the perspectives for future development, are also briefly discussed.

1. Introduction

Implementation of Numerical Weather Prediction (NWP) obviously requires the specification of appropriate initial conditions. In the early stages of Numerical Weather Prediction, forty years ago, observations were synchronous in time, and bore on the same physical variables (geopotential, horizontal winds) as those used in the prediction models for describing the state of the atmospheric flow. Long before the availability of computers, meteorologists had been performing what was called the *analysis* of the meteorological situation, which consisted in correcting every day, with the new observations, the forecast from the previous day, available under the form of meteorological maps. With the advent of NWP, this task was devoted to the computer. Two-dimensional interpolation algorithms were defined by Bergthorsson and Döös (1955) and Cressman (1959). These algorithms followed an approach which has remained essentially unchanged to this day. *Background* fields defined at the grid-points of the forecasting model are interpolated to the observation locations. The differences between observations and interpolated values are then interpolated back to grid-points in order to define *corrections* to be applied to the first-guess. Eliassen (1954) and Gandin (1963), giving a statistical basis to the analysis process, defined what has become known in meteorology as *optimal interpolation*, built on the statistical covariance functions of the meteorological fields. After many improvements, and under many variants, optimal interpolation is still to this day at the heart of most procedures for defining initial conditions of numerical weather forecasts.

But specific problems also appeared very soon. It was observed that, if appropriate precautions were not taken, the early stages of forecasts produced by non-filtered models, which do not impose an *a priori*, time independent, relationship between the mass and velocity fields, exhibited unrealistic high-frequency motions, early identified as gravity wave oscillations (Hinkelmann, 1951). This led to the practice of *initialization*, performed after the analysis proper, and intended at producing "balanced" initial conditions, which would not result in unrealistic oscillations. It was also recognized that the requirement for balanced initial conditions could on occasions be in contradiction with the requirement that the initial conditions be close to the available observations. At about the same time, the celebrated work of Lorenz (1963) showed that the atmosphere possesses the remarkable property of extremely high *sensitivity to initial conditions*. This, in addition to marking the birth of the new discipline of *deterministic dissipative chaos*, showed that deterministic prediction of the atmospheric circulation was ineluctably limited in time. Meteorologists thus learnt that all the efforts they could make were in a sense doomed to failure, but also received an additional incentive for defining as accurately as possible the initial conditions of numerical weather forecasts. In the late sixties, the development of satellite observing systems, and the perspective that asynoptic observations, performed more or less continuously in time, would become more and more numerous in the future, led to the notion that the dynamical evo-

lution of the flow should be explicitly taken into account in the very definition of the initial conditions of the forecast. The word *assimilation* was coined at that time for denoting a process in which observations distributed in time are merged together with a dynamical numerical model of the flow in order to determine as accurately as possible the state of the atmosphere.

Since then, continuous progress in theory, in efficiency of numerical algorithms, as well as in available computing power, has led to a slow but steady progress in the methods for assimilation. This progress, together with improvements in the quality of the NWP models themselves (and, to a lesser extent, with improvements in the observing system), has significantly contributed to the continuous increase observed in the last decades in the quality of numerical weather forecasts. It is worth mentioning that the proportion of resources allocated to assimilation in the whole process of NWP has steadily increased over time. At the beginning of NWP, the computational cost of analysis was negligible in comparison to the cost of a 24-hour forecast. Now, the cost of the computations required by the assimilation (in addition to the cost of integrating the model over the assimilation period) is typically the cost of a 24-hour forecast. And major meteorological centres are considering to allocate daily, for a 24-hour assimilation, the equivalent of a ten-day forecast or more. This evolution has not resulted from a clearly stated voluntary choice. With hindsight, it would be more appropriately described as a progressive "natural selection" process, during which increase of the proportion of resources allocated to assimilation repeatedly and consistently proved to be beneficial.

If there is no doubt in the minds of knowledgeable people that improvements in assimilation methods have significantly contributed over the years to the improvements in the quality of numerical weather forecasts, purely objective proofs of that fact are not readily obtained, since all components of the whole NWP process have been improving simultaneously. An extremely interesting and instructive by-product of assimilation has been presented by Salstein and Rosen (1986) (see also, Oort, 1989). These authors have compared the rate of rotation of the Earth, as estimated from geodetic measurements, with the angular momentum of the atmosphere with respect to the Earth's axis of rotation, as obtained from the analyses produced by the US National Meteorological Center. After subtraction of well identified tidal components from the observed fluctuations of the rate of rotation, the latter and the atmospheric angular momentum are correlated, over periods of up to a few years, to a remarkable high degree of accuracy. This shows that non-tidal short-term fluctuations of the rotation of the Earth are essentially due to exchanges of angular momentum between the atmosphere and the solid Earth (and that, at least over short ranges, oceans play only a minor role). Another extremely interesting and instructive by-product of assimilation has been the identification, by several authors (see, *e.g.*, Vautard, 1990), of *weather regimes* in which the atmospheric circulation stabilizes over periods that can last for as long as a few weeks.

These examples show that, if assimilation of observations originated from the needs of NWP, and if the latter is still to this day the main incentive for research and improvement in that domain, assimilation has already proven to be useful for other purposes than weather prediction. Several meteorological centres are engaged in the *reassimilation* of past observations with present computing means and assimilation algorithms. These reassimilation projects, which bear on periods as long as several decades, will produce an homogeneous description of the atmospheric circulation over long periods of time (or at least as homogeneous a description as allowed by the evolving, even if slowly evolving, observing system). The long sequences of reassimilated states thus obtained will be extremely useful for climatological studies of many kinds. This is particularly important at the present time of strong concern about the possible impact of human activities on climate, making it particularly desirable to detect early and reliably any possible climate change. One can mention a major difference between assimilation intended at defining the initial conditions of a numerical weather forecast, and *a posteriori* reassimilation of past observations. In the former case, one can of course use only observations performed before, or at the latest at, the time at which one wants to estimate the state of the flow. In the case of reassimilation, there is no reason to ignore observations performed after estimation time, and it is certainly desirable to use algorithms that are capable, in a way or another, to carry the information contained in the observations both forward and backward in time.

Assimilation of observations is also rapidly developing in the field of *dynamical oceanography*. The present state of development of numerical modelling of the oceanic circulation is described in this volume by Anderson, and the present oceanographical observing system, together with expected future developments, by Busalacchi. One basic difficulty is that the quantity of available observations is, relatively speaking, much smaller for oceanography than for meteorology (Ghil and Malanotte-Rizzoli, 1991, taking into account the appropriate characteristic spatial and temporal scales, have estimated that the temporal density of oceanographic observations has so far been four orders of magnitude less than the density of atmospheric observations). This makes validation of oceanic models particularly diffi-

cult, but also makes it particularly desirable to take the best possible advantage of the available observations. Assimilation of oceanographic observations is rather different from assimilation of atmospheric observations in that the primary purpose of assimilation of oceanographic observations is not (at least up to now) to define the initial conditions of a forecast, but more modestly to produce a reasonable description of the state of the oceanic circulation. The interest for assimilation of oceanographic observations has been strongly stimulated by the development, either already effective or anticipated, of new observing systems, in particular of satellite altimetry.

Still another domain in which there exists growing interest for assimilation of observations is modelling of the *biosphere*. Numerical modelling of the oceanic or terrestrial biosphere, and of its interactions with the atmosphere or the oceans, is rapidly progressing. The full exploitation of models of the biosphere will require appropriate assimilation of the various observations, in particular satellite observations, bearing on the biosphere.

Stated in general terms, the purpose of assimilation can be described as follows: using all the available information, determine as accurately as possible the state of the atmospheric or oceanic flow. Depending on one's particular eventual goal, one may wish to determine the state of the flow at a given time, or alternatively the history of the flow over a period of time. As for the available information, it consists first of the *observations proper*. As described in the contributions by Atlas and Busalacchi in this volume, the observations vary significantly in nature and accuracy, and have a highly irregular temporal and spatial distribution. In particular, the observations can be "direct", in that they bear on the same physical quantities to be used in the desired description of the flow (typically, velocity, temperature, plus humidity for the atmosphere or salinity for the ocean). Or they can be "indirect", *i.e.* bearing on quantities that are more or less "complicated" functions (usually some forms of integrals) of the quantities chosen for describing the flow. Satellite altimetric measurements of the ocean surface, and satellite measurements of the infra-red thermal flux emitted by the atmosphere, are examples of indirect measurements. Another example is provided by acoustic tomographic measurements performed in the ocean.

The second source of information to be used in the assimilation consists of the dynamical model, and more generally of the *physical laws* governing the flow. These physical laws are fundamentally the principles of conservation of mass, energy and momentum, and a numerical model is nothing else than a numerically usable (and approximate) statement of these principles. In addition to the contribution by Anderson, relative to oceanic modelling, the present state of development of numerical models of the atmospheric circulation is described in this volume by Arakawa. It is clear that it must be possible to acquire some knowledge of the meteorological or oceanographical fields from appropriate combination of the observations and of the physical laws governing the flow. For instance, the time derivative of the surface pressure is, under the hydrostatic approximation, the vertical integral of the divergence of the horizontal wind. Observations of the surface pressure performed at a given point at successive times therefore contain information on the wind field. To mention another simple example, information on the wind field can also be obtained from the observed motion of tracers, such as humidity.

One could think that the observations on the one hand, and the physical laws that govern the flow on the other, together make up all the appropriate information, and that there is no need to add anything to these two basic sources of information. This is certainly true in principle, but the practical situation is somewhat different. It may be useful for instance to explicitly introduce in the assimilation climatological estimates of at least some of the quantities to be determined, even if it is known that climatological quantities are in the last instance determined by the physical laws governing the flow (and by the energy input to the system). A less obvious, but practically much more important example is given by *geostrophic balance*. It is known that, at least in middle latitudes, the atmospheric and oceanic flows are in approximate geostrophic balance. It has already been mentioned that starting a numerical weather forecast from initial conditions which are not in appropriate balance will result in the presence in the forecast, at least for some time, of unrealistic high frequency motions. Geostrophic balance must be a necessary consequence of the physical laws governing the flow. But it is only an asymptotic property, in the sense that any solution of the relevant equations will asymptotically tend to approximate geostrophic balance, but need not be in geostrophic balance at the initial time. And it is indeed observed that numerical models, provided they contain the nonlinearities associated with fluid advection and a reasonable form of dissipation (plus a reasonable form of energy forcing if they are to be integrated over long periods) produce solutions which tend to approximate geostrophic balance, even though the initial conditions may have been non-geostrophic. But this property is not sufficient by itself for ensuring geostrophic balance in the fields produced by assimilation of meteorological observations over a period of, say, 24 hours. Experience shows that it is necessary to explicitly introduce in the assimilation process the information that the

atmospheric flow is in approximate geostrophic balance.

Now, no information will ever be exact, whether it comes from observations, physical laws (especially if these physical laws are expressed in the form of a discretized numerical model) or from some other form of knowledge. And thus, there always will be some uncertainty on the result of the assimilation, originating from the uncertainty on the various sources of information used in the assimilation. An ideal assimilation system should therefore produce not only an estimate of the state of the flow, but also an estimate of the *associated uncertainty.*

A first difficulty one encounters in assimilation of meteorological observations is simply the numerical dimension of the problem. The number of individual scalar meteorological observations performed over a 24-hour period is at present typically on the order of 10^5. As for the dimension of the largest NWP models (*i.e.* the number of independent parameters defining in the models the state of the flow at a given time), it is now in the range $10^6 - 10^7$. NWP models typically require one hour of elapsed computer time for 24 hours of simulated time. Assimilation of meteorological observations is most often performed over periods of 24 hours, which is of course the natural thing to do for meteorological services issuing forecasts on a daily basis. In addition to at least one 24-hour integration, the assimilation algorithms which are at present considered as most efficient essentially require the solution of one or several linear systems of equations whose dimension is either the number of observations or the dimension of the model. Fitting the corresponding computational load within the narrow limits of operational NWP imposes very strong constraints on assimilation. These constraints are critical for the choice and implementation of assimilation algorithms. Another difficulty arises from the nonlinear (actually chaotic) character of the atmospheric and oceanic flows. This imposes strong limits on assimilation, just at it imposes limits on predictability. But it must be mentioned that most of the work done so far on assimilation has in effect been done within the bounds of some appropriate local linear approximation (which will be described in some detail below), so that the full effects of nonlinearity have not been incorporated yet in assimilation.

The goal of determining as accurately as possible the state of the atmospheric or oceanic flow, together with the associated uncertainty, may seem very ambitious. We shall here attempt to convince the reader that it can be achieved, at least to a reasonable degree of precision, by solving an appropriate *generalized least-squares minimization problem.* A basic reference on assimilation of meteorological observations is a book by Daley (1991), which gives a comprehensive description of existing assimilation methods. Bennett (1992), in the general context of assimilation of oceanographic observations, describes mathematical techniques, with emphasis on possibilities for the future. More specific aspects related to spatial interpolation have been studied in detail by Thiébaux and Pedder (1987) and Wahba (1990).

Before we come to the more technical aspects of assimilation of meteorological and oceanographical observations, it is worth mentioning that similar problems are encountered in many fields of science and engineering. Navigation of aircraft and spacecraft of various kinds, in which one wants to know at any time, as accurately as possible, the position and velocity of a vehicle "observed" through various instruments, is a form of assimilation of observations. All forms of "signal filtering" are also essentially of the same nature, and one basic tool of estimation theory, Kalman filtering, which is used in assimilation of meteorological and oceanographical observations, originated in electrical engineering. Many "inverse problems" also present similarities with assimilation of observations. In plasma physics, one often wants to know the internal state of a physical system from observations performed at its surface (and also to control the system through action at its surface). Solid Earth geophysics is another example. Most of what is known on the internal structure of the Earth comes from inversion of signals (mostly seismic signals) observed at its surface. All of these examples lead to estimation problems, in which one wants to infer the state of a physical system from information which may be extremely heteregeneous in origin, nature and accuracy, and which may be related only very "indirectly" to the quantities to be estimated. The interesting fact which we want to stress here is that, in spite of the diversity of the physical systems under consideration, the methods used for solving these different problems are basically very similar, even though they often have been developed independently. They are all stated, or at least can be stated, in a probabilistic setting, and aim in effect at determining some reasonable approximation to the *conditional probability distribution* function of the state of the system, given the available information (see, *e.g.*, Tarantola, 1987, or Lorenc, 1986). The numerical algorithms are often very similar, and are fundamentally independent of the "equations" governing the physical system under consideration. These equations are in effect data, introduced in the estimation process in a way that is not basically different from (and is in some cases identical with) the way the observations proper are themselves introduced. These basic similarities are extremely instructive in that they show that estimation procedures are essentially independent of the

particular properties of the physical system under consideration. But this does not mean of course that significant differences do not exist. In particular, assimilation of meteorological and oceanographical observations seems to be unique in its extremely large numerical dimensions.

2. Sequential and variational assimilation

From a purely algorithmic point of view (and independently of the underlying theory), assimilation exists at present under two forms, *sequential assimilation* and *variational assimilation*. In sequential assimilation, which is the only form to have been used so far in operational NWP, the assimilating model is integrated over the time interval over which the observations to be used are distributed. Whenever the model time reaches an instant at which observations are available, the state predicted by the model is used as a background which is "updated", or "corrected", with the new observations. The integration of the model is then restarted from the updated state, and the process is repeated until all the available observations have been used. In operational NWP, the state obtained at the end of the assimilation period is taken as the initial state for the ensuing forecast. As already mentioned, the operation which consists in correcting a background at a given time with new observations is called an analysis. Sequential assimilation is therefore an alternative sequence of *analyses performed at observation times*, and of *integrations of the model* between successive analyses. In all algorithms of sequential assimilation that have been developed so far, there is only one sweep of the model over the assimilation period, so that each individual observation is used once and only once.

One appealing feature of sequential assimilation is the constant updating it performs on the state predicted by the model : each new piece of observation is used for correcting the latest estimate of the state of the atmospheric flow. This feature makes sequential assimilation well adapted to NWP. Numerical forecasts are normally produced at the rate of one forecast every day. In order to define the initial conditions of a new forecast at time t_0, it is very natural, starting from the initial conditions of the previous forecast at time $t_0 - 24$ hr, to perform a 24-hr sequential assimilation. This approach is implemented operationally in numerous NWP centres, and has now been running continuously in some of them for more than ten years without interruption (but with constant improvement of both the model and the assimilation algorithm itself).

However, sequential assimilation also possesses a serious drawback : because precisely of the sequential character of the assimilation, each individual piece of observation influences the estimated state of the flow only at later times, and not at previous times. There is propagation of the information contained in the observations only from the past into the future, and not from the future into the past. As said in the Introduction, this is of no importance in the case of weather prediction, where one necessarily must at one stage run the model into the future. But it certainly is a disadvantage in the case of *a posteriori* reassimilation of past observations, where it seems preferable to use algorithms capable of carrying information both forward and backward in time.

Variational assimilation, on the other hand, aims at globally adjusting a model solution to all the observations available over the assimilation period. The adjustment being simultaneous, the adjusted states at all times are influenced by all the observations over the assimilation period, thereby avoiding the difficulty mentioned above. In presently existing algorithms for variational assimilation, one first defines a scalar function which, for any model solution over the assimilation interval, measures the "distance", or "misfit", between that solution and the available observations. That so-called *objective function (*or *cost function)* will typically be a sum of squared differences between the observations and the corresponding model values, *e.g.*

$$J \equiv \Sigma_j \alpha_j (\mathrm{y}_j - \mathrm{y}^{\mathrm{o}}{}_j)^2 \tag{2.1}$$

where the $\mathrm{y}^{\mathrm{o}}{}_j$'s are the observations, the y_j's are the corresponding model values, and the α_j's are numerical weights reflecting the accuracy of the various observations. One will then look for the model solution that minimizes the objective function. Since a model solution is uniquely defined by the corresponding initial conditions at the beginning of the assimilation period, these initial conditions are taken as *control variables*, *i.e.* as the variables with respect to which the minimization is effectively performed. The minimizing initial state is obtained through an iterative procedure, each step of which requires the explicit knowledge of the local values of the set of partial derivatives, or *gradient vector*, of the objective function with respect to the initial state. As will be explained below, this gradient can be determined, at a non-prohibitive numerical cost, through use of the *adjoint equations* of the assimilating model.

3. Least-squares statistical linear estimation. Generalities

Most (but not all) assimilation algorithms that have been used so far, either for research or for operational purposes, and either of the sequential or of the variational type, can be described as more or less simplified forms of *least-squares statistical linear estimation*. Many "nonlinear" applications correspond in fact to cases which are very close, in some sense, to linearity. Least-squares statistical linear

estimation is a classical tool, whose basic principles are very simple, even though practical implementation on large dimension systems can raise many problems. We will first describe the basic principles of least-squares statistical linear estimation on elementary examples, before going to present meteorological and oceanographical applications.

Let us consider the following simple estimation problem. We want to determine some unknown scalar quantity x^t from two known measurements z_1 and z_2 of the form

$$z_1 = x^t + \zeta_1 \tag{3.1a}$$

$$z_2 = x^t + \zeta_2 \tag{3.1b}$$

where ζ_1 and ζ_2 are "observational" errors. These errors are of course unknown, but we assume that the statistical performances of the instruments which have produced z_1 and z_2 are known. More precisely, we assume that these instruments are unbiased, *i.e.*

$$E(\zeta_1) = E(\zeta_2) = 0 \tag{3.2a}$$

where $E(.)$ denotes the statistical mean, and that the statistical variances of ζ_1 and ζ_2 are known

$$E({\zeta_1}^2) = {\sigma_1}^2 \quad E({\zeta_2}^2) = {\sigma_2}^2 \tag{3.2b}$$

We assume in addition, for the sake of simplicity, that the observation errors are uncorrelated

$$E(\zeta_1 \zeta_2) = 0 \tag{3.2c}$$

This will be the case if, for instance, the two observations have been obtained with different instruments.

We now want to estimate x^t as a linear combination of the two observations z_1 and z_2, *viz.*

$$x^a = a_1 z_1 + a_2 z_2 \tag{3.3}$$

where the weights a_1 and a_2 are to be determined. We first want the estimate x^a to be statistically unbiased, *i.e.* to verify the condition $E(x^a - x^t) = 0$. This will be verified if

$$a_1 + a_2 = 1 \tag{3.4}$$

We also want x^a, among all unbiased estimates, to minimize the statistical variance of the estimation error, *viz.*

$$\sigma^2 = E[(x^a - x^t)^2] \tag{3.5}$$

The solution to this simple constrained minimization problem (minimize 3.5 under constraint 3.4) is easily found to correspond to weights a_1 and a_2 which are inversely proportional to the variances of the corresponding observation errors, *i.e.*

$$a_1 = {\sigma_2}^2/({\sigma_1}^2 + {\sigma_2}^2) \quad a_2 = {\sigma_1}^2/({\sigma_1}^2 + {\sigma_2}^2) \tag{3.6}$$

In addition, the corresponding minimum of the estimation error variance σ^2 is given by the relationship

$$1/\sigma^2 = 1/{\sigma_1}^2 + 1/{\sigma_2}^2 \tag{3.7}$$

which has a simple interpretation : if one calls "precision" the inverse of an error variance, then the precision of the estimate x^a is the sum of the precisions of the observations.

The same estimate x^a can be found through a different approach : an acceptable estimate of the exact value x^t must be close to the observations, at least within the accuracy of the latter. For any value x, the "distance" between x and the observations can be measured by the following quadratic quantity

$$J(x) \equiv (x - z_1)^2/{\sigma_1}^2 + (x - z_2)^2/{\sigma_2}^2 \tag{3.8}$$

where the observational error variances ${\sigma_1}^2$ and ${\sigma_2}^2$ account for the accuracy of the observations. Now, the value of x which minimizes $J(x)$ is precisely the estimate x^a given by eqs (3.3–6). Minimization of (3.8) therefore provides another way of determining the estimate x^a, based on a (very elementary) variational principle.

Formulæ (3.3–6) generalize to the case of any number m of observations $z_j = x^t + \zeta_j (j = 1, ..., m)$. The generalization is obvious if the observation errors are uncorrelated. It is slighty less obvious, but still elementary, when the errors are correlated. But it is more realistic to consider observations which do not necessarily bear on a quantity to be estimated. Observations are rarely performed at the times and spatial locations at which estimates are sought. In addition, as already mentioned, many observations are "indirect", and do not bear on the physical quantities to be estimated. For instance, satellite-borne radiometers measure the radiative flux emitted by the Earth to outer space at different wavelengths, while what one basically wants are estimates of the atmospheric temperature and humidity fields. The measured fluxes are functions of these fields (and of other quantities, such as cloud amount and top level pressure, and surface emissivity) through the *radiative transfer equation.* Indirect measurements will become more and more numerous in the future, and assimilation methods must allow for measurements that are "complicated" functions of the physical parameters to be estimated. Statistical linear estimation can accomodate such indirect measurements (of course within the limits of linearity, which will be discussed later) as we will now proceed to show.

We assume that what we now want to estimate is a complete vector $\mathbf{x}^t$, with dimension n and components ${x^t}_i (i = 1, ..., n)$. That vector can be thought of as consisting for instance of the values of one or several meteorological fields (temperature, wind components, humidity) at a given instant at the points of a two- or three-dimensional regular array. But the developments that follow are very general, and independent of the physical nature or significance of

the parameters to be estimated. The vector $\mathbf{x}^t$ to be estimated will be called the *state vector*, since it will in general describe the state of a physical system, such as the atmosphere or the ocean. As for the available observations, they make up a vector $\mathbf{z}$, with dimension m and components $z_j (j = 1, ..., m)$. We assume that the vector $\mathbf{z}$ can be written under the form

$$\mathbf{z} = \Gamma \mathbf{x}^t + \zeta \tag{3.9}$$

where Γ is an $m \times n$ matrix which defines the link between the parameters to be estimated and the observed quantities. Γ will be called the *observation matrix*. If the observations bear on the physical field to be estimated, but are performed at points in space-time different from the points at which estimates are sought, Γ will represent some appropriate space-time interpolation. If the observations bear on "indirect" functions of the parameters to be estimated, Γ will represent an appropriate linearization of the physical and/or statistical relationship linking $\mathbf{x}^t$ and $\mathbf{z}$. In the simple example (3.1) (where $n = 1$, $m = 2$), Γ is the matrix $(1\ 1)^T$ (where the superscript ^{T} denotes transposition). As for the m-vector ζ in (3.9), its components are the errors affecting the observations. This error vector is of course unknown, but we will assume that it is statistically unbiased, *i.e.*

$$E(\zeta) = 0$$

and that the variances-covariances of its components, making up the matrix $E(\zeta\zeta^T)$, are known

$$E(\zeta\zeta^T) = \Sigma$$

To sum up, the exactly known quantities in eq. (3.9) are the observation vector $\mathbf{z}$ and the observation matrix Γ, while the observation error vector is known only through its statistical properties, and the vector $\mathbf{x}^t$ to be estimated is totally unknown. In order to estimate $\mathbf{x}^t$, we now proceed as in the example (3.1) above, and look for an estimate $\mathbf{x}^a$ which is a linear function of $\mathbf{z}$, *i.e.* is of the form

$$\mathbf{x}^a = \mathbf{A}\mathbf{z} \tag{3.10}$$

where $\mathbf{A}$ is an $n \times m$ matrix to be determined. We want the estimate $\mathbf{x}^a$ to be unbiased, *i.e.* to be such that $E(\mathbf{x}^a - \mathbf{x}^t) = 0$. This condition is verified if

$$\mathbf{A}\Gamma = \mathbf{I}_n \tag{3.11}$$

where $\mathbf{I}_n$ is the unit matrix of order n. In addition, among all the matrices verifying (3.11), we want to choose the one that minimizes the variance of the norm of the estimation error, *i.e.* the matrix that minimizes the trace (sum of the diagonal terms) of the covariance matrix $\mathbf{P}^a = E[(\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^T]$ of the estimation error. The solution to that problem is

$$\mathbf{A} = [\Gamma^T \Sigma^{-1} \Gamma]^{-1} \Gamma^T \Sigma^{-1} \tag{3.12}$$

As for the corresponding matrix $\mathbf{P}^a$, it is equal to

$$\mathbf{P}^a = [\Gamma^T \Sigma^{-1} \Gamma]^{-1} \tag{3.13}$$

$\mathbf{P}^a$ contains the variances and covariances of the estimation errors on all the components of $\mathbf{x}^t$. In particular, its diagonal terms are the variances of these estimation errors.

Formulæ (3.12–13) generalize formulæ (3.6–7). As in that previous example, there is a variational formulation to the estimation of $\mathbf{x}^t$, which now corresponds to minimizing the objective function

$$J(\mathbf{x}) \equiv [\Gamma\mathbf{x} - \mathbf{z}]^T \Sigma^{-1} [\Gamma\mathbf{x} - \mathbf{z}] \tag{3.14}$$

where $\mathbf{x}$ is an n-vector. The meaning of expression (3.14), which generalizes (3.8), should be clear: for any $\mathbf{x}$, the vector $\Gamma\mathbf{x} - \mathbf{z}$ is the difference vector between what would be observed if the vector to be estimated was equal to $\mathbf{x}$ (and if the observations were perfect) and the actual observations $\mathbf{z}$. $J(\mathbf{x})$ is the squared norm of that difference vector, weighted so as to take into account, through Σ^{-1}, the accuracies of the different observations (and also the possible correlations between the various observation errors). The estimate $\mathbf{x}^a$ is the value of $\mathbf{x}$ for which that squared norm is minimum, *i.e.* the value of $\mathbf{x}$ that would produce, if exactly observed, the values closest to the actual observations.

It must be stressed that the problem of minimizing the statistical variance of the estimation error on $\mathbf{x}^t$, and of minimizing the objective function (3.14), are *a priori* distinctly different problems, even though the fact that they lead to identical results is algebraically obvious in simple cases. In the first problem, one minimizes a quantity defined on the n-dimensional space of *state vectors*, while in the second problem one minimizes a quantity defined on the m-dimensional space of *observations*. It is certainly not *a priori* obvious that the solutions to these two problems should always be identical.

The vector $\mathbf{x}^a$ defined by Eqs. (3.10–12) is called the *Best Linear Unbiased Estimate*, or BLUE, of $\mathbf{x}^t$ from $\mathbf{z}$. The theory leading to the BLUE is standard, and has been described here in order to stress its generality, and also as an introduction to the various methods for assimilation which, although they often do not explicitly refer to the theory of statistical linear estimation, can almost always be described, as we have already said, as more or less simplified applications of that theory. In addition to the estimate $\mathbf{x}^a$, statistical linear estimation produces the covariance matrix $\mathbf{P}^a$ of the corresponding estimation error, thus fulfilling the goal assigned in the Introduction to an ideal assimilation algorithm. It can be noted that the matrix $\mathbf{P}^a$ does not depend

on the observation vector $\mathbf{z}$, but only on the observation operator Γ and on the matrix Σ, *i.e.* on the nature and accuracy of the observations. In particular, the theory of statistical linear estimation can be used for evaluating the performance of an hypothetical observing system, defined by what that system would observe and with which accuracy, but independently of any actual observations.

Implementation of statistical linear estimation requires, in addition to the knowledge of the matrices Γ and Σ, the fact that the observation error ζ be unbiased. Actually, if that error was biased, but the bias was known, it would be sufficient to first subtract the bias from the observation vector in order to obtain a new, unbiased, observation vector. The requirement that the bias is zero is therefore in effect only a requirement that the statistical mean of the observation error is known.

To sum up the results obtained so far: before one can implement statistical linear estimation, one must know what has been observed, in terms of the parameters to be estimated (this is expressed by the observation matrix), and with which accuracy (this is expressed by the mean and the covariance matrix of the observation error vector). These requirements may seem extremely demanding, since there will always be instruments, especially newly developed instruments, for which it will certainly be very difficult to assign reliable values to the observational errors. But at the same time, it is obvious that a prerequisite for a rational use of a set of observations is to know what has been observed, and with which accuracy. It is therefore a good thing that the requirement for knowledge of the nature and accuracy of the observations comes out of the mathematics of estimation theory. If the required information is not available, one will then have to compensate for it by as reasonable as possible hypotheses on the observation matrix Γ and the corresponding error covariance matrix Σ. One advantage of studying assimilation in the perspective of general estimation theory is that it forces one to explicitly formulate hypotheses which are necessarily made in one way or another.

The fact that the required knowledge on the error vector is limited to the statistical moments of only the first two orders is due to the fact that the estimate $\mathbf{x}^{\mathrm{a}}$ has *a priori* been sought under the linear form (3.10). The determination of the most general least-variance estimate would require the knowledge of the entire probability distribution function of the observation error. However, in the case when the error vector is Gaussian, the associated conditional probability distribution function for $\mathbf{x}^{\mathrm{t}}$ is also Gaussian, with expectation defined defined by (3.10–12), and covariance defined by (3.13). In the Gaussian case, the BLUE therefore entirely solves the problem of determining the conditional probability distribution function for the state vector $\mathbf{x}^{\mathrm{t}}$.

4. The sequential form of statistical linear estimation. Kalman filtering

Formulæ (3.10, 12 and 13) assume a particular form, extremely useful in many applications, when the "observation vector" $\mathbf{z}$ defined by Eq. (3.9) can be decomposed into two components $\mathbf{z} = (\mathbf{x}^{\mathrm{bT}}, \mathbf{y}^{\mathrm{oT}})^{\mathrm{T}}$, where $\mathbf{x}^{\mathrm{b}}$ is a prior estimate of the vector $\mathbf{x}^{\mathrm{t}}$, or *background* estimate, and $\mathbf{y}^{\mathrm{o}}$ is an additional set of observations, with dimension p. The background can be written as

$$\mathbf{x}^{\mathrm{b}} = \mathbf{x}^{\mathrm{t}} + \boldsymbol{\zeta}^{\mathrm{b}} \tag{4.1a}$$

where ζ^{b} is the corresponding error, while $\mathbf{y}^{\mathrm{o}}$, assumed to be associated with a $p \times n$ observation matrix $\mathbf{H}$, can be written as

$$\mathbf{y}^{\mathrm{o}} = \mathbf{H}\mathbf{x}^{\mathrm{t}} + \boldsymbol{\varepsilon} \tag{4.1b}$$

where ϵ is the associated obervation error. Formulæ (4.1) are the analogue of (3.9), the dimension of the observation vector being now $m = n + p$, the corresponding observation matrix being $\Gamma = (\mathbf{I}_n, \mathbf{H}^{\mathrm{T}})^{\mathrm{T}}$, and the corresponding observation error vector being $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^{\mathrm{bT}}, \boldsymbol{\varepsilon}^{\mathrm{T}})^{\mathrm{T}}$. As for the covariance matrix of the observation error, it is defined by

$$\Sigma = \begin{bmatrix} E(\boldsymbol{\zeta}^{b}\boldsymbol{\zeta}^{\mathrm{bT}}) & E(\boldsymbol{\zeta}^{b}\boldsymbol{\varepsilon}^{\mathrm{T}}) \\ E(\boldsymbol{\varepsilon}\boldsymbol{\zeta}^{\mathrm{bT}}) & E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}) \end{bmatrix} \tag{4.2}$$

It is important to stress that the background $\mathbf{x}^{\mathrm{b}}$ need not, and will normally not, consist of "observations" in the strict sense of the word. The background can for instance be an already known statistical or "climatological" average of the vector $\mathbf{x}^{\mathrm{t}}$, or it can be any estimate of $\mathbf{x}^{\mathrm{t}}$, obtained through whatever means may have been available: theoretical developments, or integration of a NWP model. The only important thing is that $\mathbf{x}^{\mathrm{b}}$ be numerically known, together with the corresponding statistical covariance of the error ζ^{b}.

Formulæ (3.12–13) can be applied on the above quantities in order to obtain the corresponding BLUE $\mathbf{x}^{\mathrm{a}}$ and the covariance matrix $\mathbf{P}^{\mathrm{a}}$ of the associated estimation error. The results can be put into forms which are extremely useful from both the theoretical and the numerical points of view. For simplicity, we will assume that the errors ζ^{b} and ε are statistically uncorrelated, so that the off-diagonal terms in (4.2) are zero. The matrix $E(\zeta^{\mathrm{b}}\zeta^{\mathrm{bT}})$ will be denoted $\mathbf{P}^{\mathrm{b}}$, and the matrix $E(\varepsilon\varepsilon^{\mathrm{T}})$ will be denoted $\mathbf{R}$. With these notations, Eqs. (3.10 and 12) and (3.13) can be put in the respective forms

$$\mathbf{x}^{a} = \mathbf{x}^{\mathrm{b}} + \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}) \tag{4.3}$$

and

$$\mathbf{P}^{\mathrm{a}} = \mathbf{P}^{\mathrm{b}} - \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}^{\mathrm{b}} \tag{4.4}$$

Considering first Eq. (4.3), we see that it defines the analysed state $\mathbf{x}^{\mathrm{a}}$ as the sum of the background $\mathbf{x}^{\mathrm{b}}$ and of a correction term. The latter is proportional to the vector $\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}$, *i.e.* to the difference between the additional observation vector $\mathbf{y}^{\mathrm{o}}$ and what the observation operator $\mathbf{H}$ would produce if it was applied to the background $\mathbf{x}^{\mathrm{b}}$. That difference is therefore essentially the lack of agreement between the background and the new observations. It is obvious that, if that difference happened to be exactly equal to zero, *i.e.* if the background happened to agree perfectly with the new observations, there would be no point in performing any correction on the background. And, in the linear approach followed here, the correction to be applied to the background naturally appears as a linear function of the difference vector $\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}$. The corresponding matrix $\mathbf{K} = \mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}$, which is called the *gain matrix*, is simply the matrix which, taking into account the respective accuracies of the background and of the observations, as defined by the covariance matrices $\mathbf{P}^{\mathrm{b}}$ and $\mathbf{R}$, produces the best estimate, in the sense of the minimum of variance, of the state vector $\mathbf{x}^{\mathrm{t}}$.

The vector $\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}$ is called the vector of *residuals*, or the *innovation* vector. That second denomination, which comes from general estimation theory, is extremely suggestive, because the vector $\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}$ effectively describes *all the new information* contained in the additional observation vector $\mathbf{y}^{\mathrm{o}}$.

Another remark can be made about Eq. (4.3). Its numerical implementation requires the inversion of the matrix $\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}$, which is of dimension p. The direct use of Eq. (3.12) requires (at least) the inversion of the matrix $\Gamma^{\mathrm{T}}\Sigma^{-1}\Gamma$, which is of dimension n. If $p \ll n$ (which is usually the case in meteorological problems, where the number of observations available at a given time is normally much smaller than the dimension of the model state vector), use of formula (4.3) is much more economical. Also, formula (4.3) does not require the covariance matrix Σ to be invertible, contrary to what Eq. (3.12) does (at least apparently). If for instance, the additional observations $\mathbf{y}^{\mathrm{o}}$ are perfect ($\mathbf{R} = 0$), Eq. (4.3) can still be used. It is interesting to mention that, in this case of exact observations, the analysed state $\mathbf{x}^{\mathrm{a}}$ will be exactly compatible with the observations, in the sense that it will verify the equality $\mathbf{H}\mathbf{x}^{\mathrm{a}} = \mathbf{y}^{\mathrm{o}}$.

As for Eq. (4.4), it too has a clear significance. It defines the analysis error covariance matrix $\mathbf{P}^{a}$ as the background error covariance matrix $\mathbf{P}^{\mathrm{b}}$ minus a correction matrix. The latter is symmetric, with non-negative eigenvalues, which implies that the analysis error variance on any parameter is at most equal to the corresponding background error variance. The second term on the right-hand side of Eq. (4.4) therefore represents the gain brought about by the additional observations $\mathbf{y}^{\mathrm{o}}$ on the accuracy with which the state vector $\mathbf{x}^{\mathrm{t}}$ is known.

We mention another expression for $\mathbf{P}^{a}$, directly obtainable from (3.13)

$$(\mathbf{P}^{a})^{-1} = (\mathbf{P}^{\mathrm{b}})^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$$

This expression is generally not of much use for numerical applications, but its analogy with (3.7) is obvious: it also expresses that the "precision" of the analysis is the sum of the precisions of the background on the one hand, and of the additional observations on the other.

Finally, the variational problem associated with the estimation of $\mathbf{x}^{\mathrm{t}}$ from the background $\mathbf{x}^{\mathrm{b}}$ and the additional observation vector $\mathbf{y}^{\mathrm{o}}$ is easily seen from (3.14) to correspond to the minimization of the objective function

$$J(\mathbf{x}) \equiv [\mathbf{x} - \mathbf{x}^{\mathrm{b}}]^{\mathrm{T}}(\mathbf{P}^{\mathrm{b}})^{-1}[\mathbf{x} - \mathbf{x}^{\mathrm{b}}] + [\mathbf{H}\mathbf{x} - \mathbf{y}^{\mathrm{o}}]^{\mathrm{T}} \times \mathbf{R}^{-1}[\mathbf{H}\mathbf{x} - \mathbf{y}^{\mathrm{o}}] \tag{4.5}$$

where $\mathbf{x}$ is, as before, an n-vector. The objective function is the sum of two terms, one measuring the distance to the background $\mathbf{x}^{\mathrm{b}}$, the other measuring the distance to the additional observation vector $\mathbf{y}^{\mathrm{o}}$. These two terms are weighted by the inverse covariance matrices of the corresponding errors.

Formulæ (4.3–4) express the most general form of optimal interpolation, which, as already mentioned, is at the heart of most of the operational analysis techniques. It is most usually implemented in the following way: a background $\mathbf{x}^{\mathrm{b}}$ produced by the assimilating model for the analysis time is combined with a vector of observations $\mathbf{y}^{\mathrm{o}}$ at the same time through some approximate form of Eq. (4.3). The integration of the model is then restarted from the analyzed state $\mathbf{x}^{\mathrm{a}}$ until the next observation time, at which a new analysis is performed. This forms the basis of sequential assimilation, as it is implemented at present, with many variants, in operational NWP. Implementation of Eq. (4.3) requires the preliminary knowledge of the covariance matrices $\mathbf{P}^{\mathrm{b}}$ and $\mathbf{R}$ of the forecast and observation errors respectively. Reliable specification of those matrices, especially of $\mathbf{P}^{\mathrm{b}}$, raises a number of problems, which will not be discussed here. Let us only mention that $\mathbf{P}^{\mathrm{b}}$ is usually modelled on the basis of a number of simple hypotheses on the shape and spatial extension of the corresponding covariance functions. It is also commonly assumed that the forecast errors on geopotential and wind are geostrophically related in middle and high latitudes, which allows unambiguous determination of all required covariances from the knowledge of the covariance of geopotential forecast error

only. A number of simplifications are also made in order to reduce the computational cost of Eqs. (4.3) and especially (4.4). Indeed, exact implementation of those formulæ would be much too costly for operational NWP. Concerning (4.3), its implementation requires at least solving one linear system of dimension p. In a typical meteorological situation, p can be of the order of $10^4 - 10^5$. This is much too large for the constraints of operational NWP. In order to reduce the corresponding cost, only observations located in the vicinity of a given model gridpoint are used when performing the analysis at that point. That "selection" of observations is certainly legitimate in the sense that observations performed at a large distance of a given point must have a small influence on the analysed fields at that point. However, experience shows that it nevertheless introduces spatial noise in the analysed fields, which must be then filtered out by *ad hoc* smoothing. As for Eq. (4.4), it is not implemented in its totality, but only the diagonal terms of the matrix $\mathbf{P}^{\mathrm{a}}$, *i.e.* the variances of the analysis error, are usually computed.

Experience also shows that, in spite of the geostrophic link assumed between the wind and geopotential forecast errors, the fields produced by optimal interpolation are still contamined by unrealistic ageostrophic noise. This noise must be filtered out through additional appropriate initialization procedures, already alluded to in the Introduction (for more information on this aspect, see Temperton, 1988, and references therein).

Optimal interpolation, as implemented in present operational NWP, produces results which are on the whole quite satisfactory. Description of its performances can be found in a number of articles or reports. An article by Lorenc (1981), although it is not very recent, contains a clear description of the basic principles of operational optimal interpolation and of the main properties of the results it produces. Among recent developments, the National Meteorological Center (Washington, USA) has introduced in operations an implementation of optimal interpolation, based not on Eq. (4.3), but on an iterative minimization of the objective function (4.5) (Parrish and Derber, 1992). The results show improvement of the quality of the analysis for a lower computational cost. This is probably largely due to the suppression of the need for selecting observations in the vicinity of each analysis point. In addition, an appropriate definition of the matrix $\mathbf{P}^{\mathrm{b}}$ eliminates the need for initialization. It therefore appears that it is numerically more efficient to perform optimal interpolation, not by direct use of Eq. (4.3) and explicit exact solution of one (or several) linear system of equations, but by iterative (and approximate) minimization of the corresponding objective function (4.5). At the time of writing, the European Centre for Medium-range Weather Forecasts was also planning to introduce soon a similar "three-dimensional variational analysis".

But there is much more to formulæ (4.3–4) than analysis at a given time, and those formulæ are at the basis of the technique of *Kalman filtering* which, in the linear context considered here, is the appropriate approach to sequential assimilation. Let us first consider the following situation. A vector $\mathbf{z}$ of observations, of the form (3.9), has been processed through eqs (3.10–12) and (3.13) in order to produce the corresponding BLUE, which we will denote $\mathbf{x}^{\mathrm{a}-}$, together with the covariance matrix of the associated estimation error, which we will denote $\mathbf{P}^{\mathrm{a}-}$. At that stage, an additional vector of observations $\mathbf{y}^{\mathrm{o}}$, of the form (4.1b), becomes available, and one wants to determine the BLUE $\mathbf{x}^{\mathrm{a}}$ of $\mathbf{x}^{\mathrm{t}}$ from the combined observation vector $\mathbf{v} = (\mathbf{z}^{\mathrm{T}}, \mathbf{y}^{\mathrm{oT}})^{\mathrm{T}}$. Is it necessary to restart the computations from the beginning, or is it possible to take advantage of the computations that have already been performed and have led to $\mathbf{x}^{\mathrm{a}-}$ and $\mathbf{P}^{\mathrm{a}-}$? It must be clear from the foregoing developments, and it is easy to verify that, if the observation error vector ε associated with $\mathbf{y}^{\mathrm{o}}$ is uncorrelated with the estimation error vector $\mathbf{x}^{\mathrm{a}-} - \mathbf{x}^{\mathrm{t}}$, then the BLUE $\mathbf{x}^{\mathrm{a}}$ and the associated covariance matrix of estimation error are given by formulæ (4.3) and (4.4), $\mathbf{x}^{\mathrm{b}}$ and $\mathbf{P}^{\mathrm{b}}$ being replaced by $\mathbf{x}^{\mathrm{a}-}$ and $\mathbf{P}^{\mathrm{a}-}$ respectively. There is no need therefore for restarting the computations from the original $\mathbf{z}$, and one can take advantage of the already performed computations. In particular, if the dimension p of the additional observation vector $\mathbf{y}^{\mathrm{o}}$ is small in comparison to the dimension n of $\mathbf{x}^{\mathrm{t}}$, the numerical gain of using formulæ (4.3–4) rather than restarting the entire computations is obvious.

Formulæ (4.3–4) therefore provide a way for constantly updating with new observations the latest estimate of the state of the system under observation. If the errors associated with the successive observations are mutually uncorrelated, the estimate obtained at any stage of the process will always be the BLUE of $\mathbf{x}^{\mathrm{t}}$ from the observations already introduced into the process, and there will be no loss in the accuracy of the estimate resulting from the sequential character of the procedure.

In the case of assimilation of observations, an additional complication comes from the fact that the observations to be assimilated are distributed over a time period over which the state of the system is itself evolving. In order to take the temporal dimension into account, and in agreement with the linear approach followed so far, we consider a system whose state evolves in time according to the linear equation

$$\mathbf{x}^{\mathrm{t}}{}_{k+1} = \mathbf{M}\mathbf{x}^{\mathrm{t}}{}_{k} + \boldsymbol{\eta}_k \tag{4.6}$$

where $\mathbf{x}^{t}{}_{k}$ denotes the state of the system at time k, while $\mathbf{M}$ is a known *transition matrix* expressing the time variation of the state vector between times k and $k+1$. As for the term η_k, it represents contributions to the temporal variations of the state vector which are not represented by the transition matrix. One can consider that the transition matrix is the numerical model available for performing the assimilation, while the term η_k represents the accumulated effect, on the variation of the state vector between times k and $k+1$, of the various processes not represented by $\mathbf{M}$. The term η_k will accordingly be called the *model error*. It will be considered as an unbiased random vector, uncorrelated in time, with known covariance matrix $\mathbf{Q}$

$$E(\eta_k \eta_l{}^{\mathrm{T}}) = \mathbf{Q}\delta_{kl} \tag{4.7}$$

where δ_{kl} is the classical Kronecker symbol (in order to keep notations simple, we do not allow for an explicit time dependence of $\mathbf{M}$ and $\mathbf{Q}$; introducing such a dependence is straightforward, and would not modify the essence of what is to follow).

We assume in addition that observations of the general form (4.1b) are available at the successive instants $k = 0, 1, ..., N$:

$$\mathbf{y}^{o}{}_{k} = \mathbf{H}\mathbf{x}^{t}{}_{k} + \varepsilon_k \tag{4.8}$$

The corresponding observation errors ε_k are supposed as before to be unbiased, to be uncorrelated in time and have covariance matrix $\mathbf{R}$ (with again no explicit time dependence for $\mathbf{H}$ and $\mathbf{R}$). This leads to the expression

$$E(\varepsilon_k \varepsilon_l{}^{\mathrm{T}}) = \mathbf{R}\delta_{kl}$$

In addition, the model and observation errors are supposed to be mutually uncorrelated

$$E(\varepsilon_k \eta_l{}^{\mathrm{T}}) = 0$$

We now assume that the BLUE $\mathbf{x}^{a}{}_{k}$ of the state $\mathbf{x}^{t}{}_{k}$ of the system at time k from all observations up to time k has been determined, together with the covariance matrix $\mathbf{P}^{a}{}_{k}$ of the corresponding estimation error. The BLUE of the state of the system at time $k+1$ from all observations up to time k can be shown to be equal to

$$\mathbf{x}^{b}{}_{k+1} = \mathbf{M}\mathbf{x}^{a}{}_{k} \tag{4.9}$$

As for the covariance matrix of the corresponding estimation error $\mathbf{x}^{b}{}_{k+1} - \mathbf{x}^{t}{}_{k+1}$, it is equal to

$$\begin{aligned}\mathbf{P}^{b}{}_{k+1} &\equiv E[(\mathbf{x}^{b}{}_{k+1} - \mathbf{x}^{t}{}_{k+1})(\mathbf{x}^{b}{}_{k+1} - \mathbf{x}^{t}{}_{k+1})^{\mathrm{T}}] \\ &= E[(\mathbf{M}\mathbf{x}^{a}{}_{k} - \mathbf{M}\mathbf{x}^{t}{}_{k} - \boldsymbol{\eta}_k)(\mathbf{M}\mathbf{x}^{a}{}_{k} - \mathbf{M}\mathbf{x}^{t}{}_{k} - \boldsymbol{\eta}_k)^{\mathrm{T}}] \\ &= E\{[\mathbf{M}(\mathbf{x}^{a}{}_{k} - \mathbf{x}^{t}{}_{k}) - \boldsymbol{\eta}_k][\mathbf{M}(\mathbf{x}^{a}{}_{k} - \mathbf{x}^{t}{}_{k}) - \boldsymbol{\eta}_k]^{\mathrm{T}}\} \\ &= \mathbf{M}\mathbf{P}^{a}{}_{k}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}\end{aligned} \tag{4.10}$$

where the various non-correlation hypotheses have been used. The first term on the last line represents the error at time $k+1$ resulting from the error at time k, while the second term is the contribution of the model error η_k between times k and $k+1$.

At this stage, it is easy to introduce the observation vector $\mathbf{y}^{o}{}_{k+1}$ at time $k+1$: one simply has to use formulæ (4.3–4), $\mathbf{x}^{b}$ and $\mathbf{P}^{b}$ being replaced by $\mathbf{x}^{b}{}_{k+1}$ and $\mathbf{P}^{b}{}_{k+1}$ respectively, and $\mathbf{y}^{o}$ being replaced by $\mathbf{y}^{o}{}_{k+1}$. This leads to the BLUE $\mathbf{x}^{a}{}_{k+1}$ of the state of the system at time $k+1$, from all the observations up to, and including, time $k+1$

$$\begin{aligned}\mathbf{x}^{a}{}_{k+1} &= \mathbf{x}^{b}{}_{k+1} + \mathbf{P}^{b}{}_{k+1} \\ &\times \mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{b}{}_{k+1}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}(\mathbf{y}^{o}{}_{k+1} - \mathbf{H}\mathbf{x}^{b}{}_{k+1})\end{aligned} \tag{4.11}$$

and to the covariance matrix of the corresponding estimation error, *viz.*

$$\begin{aligned}\mathbf{P}^{a}{}_{k+1} &= \mathbf{P}^{b}{}_{k+1} - \mathbf{P}^{b}{}_{k+1} \\ &\times \mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{b}{}_{k+1}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P}^{b}{}_{k+1}\end{aligned} \tag{4.12}$$

The sequential process defined by Eqs. (4.9) to (4.12) is called *Kalman filtering* (Kalman, 1960). At any stage, Kalman filtering produces the BLUE of the state of the system under observation, using all observations up to estimation time. It also produces the covariance matrix of the corresponding estimation error.

Kalman filtering has been applied to many different problems. A general description of the theory of Kalman filtering and of its properties can be found in, *e.g.*, Jazwinski (1970). In the case of assimilation of meteorological or oceanographical observations, one can see that, if one accepts the linear hypotheses which underlie Eqs. (4.6) and (4.8), Kalman filtering fulfills the goal assigned in the Introdution to an ideal assimilation system : namely, to use all the available information in order to produce the most accurate possible description of the state of the flow, together with the uncertainty resulting from the uncertainties on the various sources of information. In the present case, the available information consists, not only of the observations $\mathbf{y}^{o}{}_{k}$ (Eq. 4.8), but also of the model (4.9) (and of the initial estimate $\mathbf{x}^{b}{}_{0}$ from which the whole process must be started). As for the associated uncertainties, they are defined by the covariance matrices $\mathbf{R}$ and $\mathbf{Q}$ (and the initial covariance matrix $\mathbf{P}^{b}{}_{0}$). Kalman filtering consistently combines all these elements in order to produce the BLUE (4.11) and the associated covariance matrix (4.12).

The application of Kalman filtering to assimilation of meteorological and oceanographical observations has been studied by a number of authors, in particular Ghil and collaborators (see, *e.g.*, Ghil, 1989, or Ghil and Malanotte-Rizzoli, 1991). Experiments performed with various linear systems have produced convincing results as to the capability of

the method for effectively extracting the information contained in the observations and the model. One major difficulty with Kalman filtering in the context of assimilation of meteorological and oceanographical observations is its numerical cost. Writing the first term on the last line of Eq. (4.10) under the form $\mathbf{M}(\mathbf{M}\mathbf{P}^{\mathrm{a}}{}_k)^{\mathrm{T}}$ shows that the corresponding computations require two successive matrix multiplications by $\mathbf{M}$. Now, the multiplication of one vector by $\mathbf{M}$ corresponds to one integration of the model between times k and $k+1$ (Eq. 4.9). Implementation of Eq. (4.10) therefore requires $2n$ integrations of the model, where n is, as above, the dimension of the state vector of the model. With values of n on the order of $10^6 - 10^7$, this goes largely beyond the possibilities of assimilation for operational NWP, or even of *a posteriori* assimilation. In operational NWP, the computation (4.10) is replaced by a simple multiplication of the variances of the analysis errors by an *a priori* specified coefficient (typically, 1.5 for a 6hr-forecast), the associated correlations being modelled independently, as already mentioned above. The corresponding numerical cost is negligible, but that procedure amounts to ignoring the influence of the particular meteorological situation under consideration, and especially of the particular instabilities that may develop, on the evolution of the forecast error. This certainly is one of the major weaknesses of present operational assimilation methods, which, in the perspective taken here, can be described as degraded but economical forms of Kalman filtering. Comparisons of the results produced by variational assimilation and by algorithms similar to operational algorithms (Rabier *et al.*, 1993) suggest that a more accurate description of the evolution of forecast error might substantially improve the quality of assimilations.

Now, the correlation between forecast errors at points located a large distance apart must be negligible, and a large proportion of the entries of covariance matrices such as $\mathbf{P}^{\mathrm{a}}{}_k$ must have zero or negligible values. This should allow to reduce the cost of computation (4.10). In addition, it is known that the most rapidly amplifying modes in the evolution of the forecast error are geostrophic modes (see, *e.g.*, Lacarra and Talagrand, 1988), so that it should be possible to restrict computation (4.10) to a subset of all the model modes. These ideas have been exploited by several authors (see, *e.g.*, Cohn and Parrish, 1991, Dee, 1991, or Bouttier, 1994) in order to reduce the cost of computation (4.10). Much active research is now being done on the problem of defining algorithms for describing the temporal evolution of the forecast error that are both economical enough for practical implementation, and accurate enough for improving on present operational methods of sequential assimilation.

5. The variational form of statistical linear estimation

We will now restrict ourselves to the case where the model is supposed to be perfect, *i.e.* $\eta_k = 0$ in Eq. (4.6), so that the exact evolution of the flow reduces to

$$\mathbf{x}^{\mathrm{t}}{}_{k+1} = \mathbf{M}\mathbf{x}^{\mathrm{t}}{}_k \tag{5.1}$$

The variational form (3.14) of the estimation problem defined by Eqs. (4.8) and (5.1) leads to the objective function

$$J(\mathbf{x}) \equiv \sum\nolimits_{0\le k\le N} [\mathbf{H}\mathbf{x}_k - \mathbf{y}^{\mathrm{o}}{}_k]^{\mathrm{T}} \mathbf{R}^{-1} [\mathbf{H}\mathbf{x}_k - \mathbf{y}^{\mathrm{o}}{}_k] \tag{5.2}$$

where $\mathbf{x} = (\mathbf{x}_k{}^{\mathrm{T}})^{\mathrm{T}}$ is a sequence of model states at successive times, linked by the model equation (5.1). $J(\mathbf{x})$ is the sum of the model-minus-observations squared differences, weighted by the inverse of the observation error covariance matrices. Minimizing the objective function (5.2) under the constraint (5.1) will produce at any time k the BLUE of the real state $\mathbf{x}^{\mathrm{t}}{}_k$ of the system at time k, from *all the available observations*, *i.e.* from observations performed before, at, and after time k. In particular, the state at the end of the assimilation period will be the same as the state produced by Kalman filtering (under the assumption of an exact model, *i.e.* under the condition that $\mathbf{Q} = 0$ in Eq. 4.10). The variational form of statistical estimation therefore provides a way to globally adjust a model to observations distributed in time.

But we can also note that the assumption of linearity, necessary to establish the link with the theory of statistical linear estimation and with Kalman filtering, is by no means necessary for a variational problem of type (5.1–2). One can very well consider the problem of minimizing an objective function of the form (5.2), under a constraint of the form (5.1), where the matrices $\mathbf{M}$ and $\mathbf{H}$ are replaced by nonlinear operators. Indeed, numerical models of the atmospheric or oceanic flows are nonlinear, and many observations are nonlinearly related to the atmospheric or oceanic variables one wants to estimate. For an already mentioned example, infrared radiances measured by satellites are related to the temperature and humidity profiles of the emitting atmospheric columns through the radiative transfer equation, which is strongly nonlinear. We will therefore drop for the time being the hypothesis of linearity (only to come back to it later in order to show that it is often justified in some sense) and consider the problem of minimizing a nonlinear objective function (5.2) (*i.e.* an objective function with nonlinear observation operators) under a nonlinear constraint of the form (5.1). In order to stress that we are now dealing with nonlinear operators, we will

use the notations M and H instead of $\mathbf{M}$ and $\mathbf{H}$ respectively.

There basically exist two methods for solving a constrained minimization problem. The principle of the first method is obvious, and consists in *reducing the constraint*, *i.e.* in eliminating some of the constrained variables so as to transform the problem into an unconstrained problem. In the present case, one can note that a model solution (5.1) is uniquely defined by the specification of the corresponding initial condition $\mathbf{x}_0$. The objective function J can therefore be considered as a function of $\mathbf{x}_0$ only, upon which no constraint is imposed, so that one is led to a problem of unconstrained minimization with respect to $\mathbf{x}_0$. The second method, whose principle is much less obvious, consists in associating unknown coefficients, called *Lagrange multipliers*, with the constraints of the problem, and to form the corresponding *Lagrangian*. In the present case, there are N constraints (5.1), each of which of dimension n, and the set Λ of Lagrange multipliers consists of N vectors $\Lambda_k (k = 0, 1, ..., N-1)$, each of dimension n. The associated Lagrangian reads

$$L(\mathbf{x}, \Lambda) = J(\mathbf{x}) + \sum\nolimits_{1 \leq k < N} {\Lambda_k}^{\mathrm{T}} [\mathbf{x}_{k+1} - M\mathbf{x}_k]$$

A well-known theorem then says that the minima of the constrained minimization problem (5.1–2) correspond to the stationary points of the Lagrangian $L(\mathbf{x}, \Lambda)$, considered as a function of the independent variables $\mathbf{x}$ and Λ.

The method of *adjoint equations*, which is a classical tool of control theory (Lions, 1971), seems to be by far the most efficient way for numerically solving the minimization problem (5.1–2). Interestingly enough, the method of adjoint equations can be derived by either reducing the constraint (5.1) so as to use only the initial state $\mathbf{x}_0$ as independent variable (see, *e.g.*, Le Dimet and Talagrand, 1986, or Talagrand and Courtier, 1987), or alternatively by using the technique of Lagrange multipliers (see, *e.g.*, Thacker and Long, 1988). Assuming for instance that we want to solve problem (5.1–2) as a problem of unconstrained minimization with respect to the initial state $\mathbf{x}_0$, it is necessary, in order to even start solving the problem, to be able to relate the variations of the initial state $\mathbf{x}_0$ to the corresponding variations of the objective function J. For a given initial state, these variations are related through the local vector of partial derivatives, or *gradient vector*, of the objective function with respect to the components of the initial state. In particular, if one is able to numerically compute the gradient for a given initial state, it will be possible to feed that gradient into a standard minimization algorithm which will determine the minimizing initial state through successive iterations. In most situations, it will of course be impossible to establish explicit analytical expressions for the gradient. It is possible to numerically (and approximately) determine the gradient through explicit finite perturbations of the initial state, but this would be much too costly for practical implementation : it would require to compute the objective function, *i.e.* to effectively integrate the model over the assimilation period, as many times as there are independent components in the initial state. The method of adjoint equations provides a way for computing the gradient at a numerical cost which is at most a few times the cost of one computation of the distance function. The principle of the method is extremely simple. Let us consider a computer code (or part of a code) which, starting from some input vector $\mathbf{u}$ with components $\mathbf{u}_i (i = 1, ..., q)$, produces an output vector $\mathbf{v}$ with components $\mathbf{v}_j (j = 1, ..., r)$. The process can be described by the equation

$$\mathbf{v} = G(\mathbf{u}) \tag{5.3}$$

where G stands for all the computations that lead from $\mathbf{u}$ to $\mathbf{v}$. For a given perturbation *du* on the input, the resulting perturbation $\delta\mathbf{v}$ on the output is equal to first order to

$$\delta\mathbf{v} = G'\delta\mathbf{u} \tag{5.4}$$

where G' is the matrix of local partial derivatives, or *jacobian matrix*, of the components of $\mathbf{v}$ with respect to the components of $\mathbf{u}$. Eq. (5.4) is called the *tangent linear equation* to (5.3). Let now $J(\mathbf{v})$ be a scalar function of the output $\mathbf{v}$. The gradient of J with respect to $\mathbf{u}$ is given by the chain rule

$$\frac{\partial \mathbf{J}}{\partial \mathbf{u}_i} = \sum_{j=1}^{r} \frac{\partial \mathbf{v}_j}{\partial \mathbf{u}_i} \frac{\partial \mathbf{J}}{\partial \mathbf{v}_j} \quad i = 1, ..., q$$

or, in transparent matrix notation

$$\nabla_{\mathbf{u}} J = {G'}^{\mathrm{T}} \nabla_{\mathbf{v}} J \tag{5.5}$$

where, as before, the superscript $^{\mathrm{T}}$ denotes transposition.

The adjoint method is based on a systematic use of formula (5.5). More precisely, let us suppose that the process G is the composition of a number of more elementary processes, namely

$$G = G_c o ... o G_2 o G_1$$

the jacobian G' will be product of the elementary jacobians

$$G' = G'_M ... G_2' G_1'$$

and the transpose ${G'}^{\mathrm{T}}$ will be the product of the elementary transposes, taken in reversed order

$${G'}^{\mathrm{T}} = {G_1'}^{\mathrm{T}} {G_2'}^{\mathrm{T}} ... {G_M'}^{\mathrm{T}}$$

This shows that, in order to numerically determine the gradient $\nabla_{\mathbf{u}} J$ with respect to the input $\mathbf{u}$, it is sufficient to proceed backwards through the direct computations and, at every step, to perform the corresponding transpose, or *adjoint* computations. The total cost of one adjoint computation (5.5) will generally be of the same order of magnitude as the cost of one direct computation (5.3). (It can be shown, see *e.g.*, Morgenstern, 1984, that the total operation count of one adjoint computation can be reduced to at most 4 times the total operation count of the corresponding direct computation, this ratio being reduced to 2 if one considers only multiplications and divisions). This is of course much more economical than direct perturbations of the input vector.

In the case of the determination of the gradient of the (nonlinear) objective function (5.2) with respect to the initial state $\mathbf{x}_0$ of the (nonlinear) assimilating model (5.1), the adjoint computations reduce to integrating the equation (see, *e.g.*, Talagrand and Courtier, 1987)

$$\delta' \mathbf{x}_k = M'^{\mathrm{T}} \delta' \mathbf{x}_{k+1} + H'^{\mathrm{T}} \mathbf{R}^{-1} [H \mathbf{x}_k - \mathbf{y}^{\mathrm{o}}{}_k] \quad (5.6)$$

backwards in time, starting from the "final" state $\delta' \mathbf{x}_{N+1} = 0$. In this equation, M' and H' are the jacobians of the respective nonlinear model and observation operators M and H. The gradient of the objective function with respect to the initial state $\mathbf{x}_0$ is equal to $2\delta' \mathbf{x}_0$.

It is seen that the basic model solution $\mathbf{x}_k$ under consideration explicitly appears in the adjoint equation (5.6) in the quantity $H\mathbf{x}_k - \mathbf{y}^{\mathrm{o}}{}_k$, which, except for its sign, is the innovation vector of Eqs. (4.3) and (4.11). This means that the basic solution will have to be computed, and kept in memory, before the adjoint integration can be performed. In the general case of nonlinear operators M and H, the basic solution will also be necessary for determining the jacobians M' and H'. Saving the basic solution in memory may entail large core requirements, which constitute one important feature of the adjoint method.

The method of adjoint equations for performing variational assimilation of meteorological observations seems to have been first suggested by Penenko and Obraztsov (1976), who applied it to a simple, small-dimensional linear problem. Since then, a large number of experiments have been performed on (usually nonlinear) models of increasing complexity and dimensions, and with various types of observations. Experiments have been performed on both meteorological and oceanographical examples (for meteorological applications see, *e.g.*, Lewis and Derber, 1985, Talagrand and Courtier, 1987, Derber, 1987, Courtier and Talagrand, 1987, 1990, Lorenc, 1988, Thépaut and Courtier, 1991, Navon *et al.*, 1991, Rabier and Courtier, 1992; for oceanographical applications see, *e.g.*, Thacker and Long, 1988, Sheinbaum and Anderson, 1990a and b, Greiner and Perigaud, 1994). The first general conclusion that can be drawn from these experiments is that variational assimilation works in that it is capable of minimizing the objective function. Also, and contrary to what happens in sequential assimilation, there is propagation of information, as should be, both forward and backward in time. However, when the objective function contains only terms measuring the misfit between individual observations and model values, the minimization solution tends to contain unrealistic "noise", often under the form of small-scale oscillations and/or of ageostrophic gravity waves. The minimizing solution is physically realistic only if appropriate terms, measuring the energy contained in the small scales of the flow, or its ageostrophy, are added to the objective function (see, *e.g.*, Courtier and Talagrand, 1990, Thépaut and Courtier, 1991). Indeed, the need for adding terms intended at avoiding unrealistic oscillations in the estimated fields is by no means restricted to meteorological or oceanographical applications, but is almost universal in problems where fields are estimated through a variational method. Such terms are often called "smoothing", "penalizing" or still "regularizing" terms. But it must be stressed that the need for appropriate smoothing is not restricted to variational methods. It is also present in statistical estimation when implemented through Kalman filtering which, as already said, must lead to the same final result as variational algorithms. In present operational optimal interpolation, the requirement for appropriate smoothing is satisfied on the one hand through the presence of the background $\mathbf{x}^{\mathrm{b}}$, which defines what the analyzed field must be in data-void areas, and on the other hand through the "initialization" process, which filters out unrealistic gravity wave oscillations.

The most recent experiments of variational assimilation of meteorological observations have been performed with multilevel primitive equation models similar, but not identical yet, to the models used in NWP (Thépaut *et al.*, 1993). The remaining differences lie in the resolution, which is still coarser in variational assimilation experiments (typically one order of magnitude less points in the horizontal than in operational models), and in the representation of many "physical" processes, such as convection and water phase changes, which are still absent from variational assimilation. These recent experiments confirm the results previously obtained, and show in particular that variational assimilation, because it explicitly uses the evolution equations of the system, is able to propagate the information contained in the observations much more accurately than operational optimal interpolation.

Variational assimilation, like Kalman filtering, is therefore able to assimilate observations in a way

that is exactly consistent with the dynamics of the system, as described by the model equations. But, as in the case of Kalman filtering, and in spite of the fact that the adjoint equations are by far the most efficient way for computing the gradient of the objective function, the computational price to be paid is heavy : in addition to the necessity of storing in memory the model solution produced by the direct integration which must be performed before each adjoint integration, minimization of the objective function typically requires from 10 to 30 iterations of the minimization algorithm. Each iteration itself requires one integration of the model over the assimilation period, followed by one adjoint integration. The cost of one adjoint integration is about twice the cost of one direct integration, so that one minimization typically requires the equivalent of between 30 and 100 integrations of the model over the assimilation period. This is of course very costly, but it now seems it will be possible to operationally implement in the coming years simplified forms of variational assimilation, in which the assimilation will be performed at a somewhat lower resolution than the full NWP model (Courtier *et al.*, 1994).

The present situation as concerns assimilation methods is therefore rather clear. In addition to the relatively simple, rather *ad hoc*, but economical and basically satisfactory operational algorithms, there exist two broad classes of algorithms that are capable, in the general framework of statistical linear estimation, of consistently extracting the information contained in the observations on the one hand and in the physical laws expressed by the assimilating model on the other : Kalman filtering and variational assimilation. In order to implement these algorithms, one must express the observations under the general form (3.9), *i.e.* one must know what has been measured, in terms of the parameters to be estimated, and with which accuracy. Now, exact implementation of either of these two classes of algorithms is numerically costly, and a large part of the research being done at present on assimilation is in effect directed at determining the most cost-efficient simplications that can be made on them. This task may indeed be with us for a long time : no end is foreseen to the increase in the power of computers and to the deep modifications in their structures, nor to the development of new observing systems and of more realistic models. Changes in any of these aspects may radically modify any conclusion one may have reached as to the most efficient way to perform assimilation.

We will now briefly comment on the relative advantages and disadvantages of Kalman filtering and variational assimilation. As already mentioned, both algorithms will lead to the same final state at the end of the assimilation period in the case of linear observations (*i.e.* of a linear observation operator $\mathbf{H}$) and of a perfect (*i.e.* $\eta_k = 0$) linear model (5.1). A basic difference between the two algorithms is that Kalman filtering carries information only from the past into the future, while variational assimilation carries information in both time directions. On the other hand, variational assimilation, contrary to Kalman filtering, does not take into account the fact that the assimilating model, like the observations, will never be perfect and will always contain errors. But it must also be said that these differences are only true of these algorithms as they have been described here and as they have been most usually implemented so far in meteorological and oceanographical applications. As concerns Kalman filtering, there exists a procedure, called Kalman *smoothing* (see, *e.g.*, Anderson, 1979), which allows, once a first pass has been performed over the assimilation period, to proceed backward in time so as to obtain, at any intermediate time k, the BLUE of the state of the system at time k from all available observations, performed before, at or after time k. We will not describe here the theory of Kalman smoothing, which is related to the theory of adjoint equations, and will only refer to Gaspar and Wunsch (1989) as a simple but instructive example of an application of Kalman smoothing to an oceanographical problem. And, as concerns variational assimilation, it can incorporate model errors: it suffices to impose the model equation (5.1) not as a constraint to be exactly satisfied by the sequence $\mathbf{x}_k$ of assimilated states, but (to use the vocabulary introduced by Sasaki, 1970) as a "weak constraint" to be satisfied only approximately. This can be done by modifying the objective function (5.2) to

$$J(\mathbf{x}) \equiv \sum_{0\leq k<N} [H\mathbf{x}_k - \mathbf{y^o}_k]^{\mathrm{T}} \mathbf{R}^{-1} [H\mathbf{x}_k - \mathbf{y^o}_k] + \sum_{0\leq k<N} [\mathbf{x}_{k+1} - M\mathbf{x}_k]^{\mathrm{T}} \mathbf{Q}^{-1} [\mathbf{x}_{k+1} - M\mathbf{x}_k] \tag{5.7}$$

where $\mathbf{Q}$ is, as in (4.7) the covariance matrix of the model error. The meaning of the second sum on the right-hand-side of (5.7) must be clear : it simply expresses that the difference $\mathbf{x}_{k+1} - M\mathbf{x}_k$ must not be considered as exactly zero, as it would be if the model was exact, but equal to zero only within the uncertainty defined by the matrix $\mathbf{Q}$. Accordingly, the sequence of states must be considered as unconstrained, and the minimization of J must be performed with respect to the entire sequence $\mathbf{x} = (\mathbf{x}_k{}^{\mathrm{T}})^{\mathrm{T}}$. It is not difficult to see from the variational form (3.14) of statistical linear estimation, that the sequence of states minimizing (5.7) is made up of the BLUEs, at all times k, of the state of the system, from all observations over the entire assimilation period. It results in particular that minimization of (5.7) must lead to the same sequence

of assimilated states as Kalman filtering, followed by Kalman smoothing. For additional information on weak constraint variational assimilation, see Bennett (1992) and Bennett *et al.* (1993).

Another difference between Kalman filtering and variational assimilation is that the former produces, in addition to the BLUE of the state of the system, the covariance matrix of the corresponding estimation error, while the latter produces only the BLUE. In this sense, only Kalman filtering fulfills the goal assigned to an ideal assimilation system in the Introduction. This of course is obtained at the already mentioned much higher cost of Kalman filtering. Now, it is easy to verify from eq. (3.13) and (3.14) that the covariance matrix of the estimation error is the inverse of the hessian (matrix of second derivatives) of the objective function. The question thererefore arises whether the inverse hessian can be computed in variational assimilation, at least to a sufficient degree of accuracy, at a lower cost than in Kalman filtering. Indeed, some minimization algorithms, of the so-called quasi-Newton type (see, *e.g.*, Gill *et al.*, 1982), do compute an approximate inverse hessian in the course of the minimization. The problem of the determination of the estimation error in variational assimilation is the subject of active research (Fisher, *pers. com.*).

To conclude with the theoretical and methodological aspects of assimilation, we will discuss a point which we have so far left in some obscurity, namely the validity of the linear hypothesis which underlies the theory of Kalman filtering and which, although not necessary for variational assimilation, gives it a special significance and facilitates the understanding and analysis of the results it produces. Not only are the equations governing the atmospheric and oceanic flows strongly nonlinear, but their nonlinearity is at the origin of one of the most important properties of these flows, namely their chaotic character. This character imposes stringent limits on the predictability of these flows, and one can legitimately wonder whether a linear hypothesis is legitimate in the context of assimilation. Considering first the equations for Kalman filtering, let us assume that the model and observation operators are nonlinear, and accordingly denoted M and H respectively. If the difference $\mathbf{x}^{\mathrm{a}}{}_k - \mathbf{x}^{\mathrm{t}}{}_k$ is small enough, the quantity $M\mathbf{x}^{\mathrm{a}}{}_k - M\mathbf{x}^{\mathrm{t}}{}_k$ (second line of Eq. 4.10) can be approximated by $M'(\mathbf{x}^{\mathrm{a}}{}_k - \mathbf{x}^{\mathrm{t}}{}_k)$, where M' is the jacobian matrix of the operator M, taken at point $\mathrm{x}^{\mathrm{a}}{}_k$. Eq. (4.10) accordingly becomes

$$\mathbf{P}^{\mathrm{b}}{}_{k+1} = M'\mathbf{P}^{\mathrm{a}}{}_k M'^{\mathrm{T}} + \mathbf{Q} \tag{5.8}$$

Similarly, if the difference $\mathbf{x}^{\mathrm{t}}{}_{k+1} - \mathbf{x}^{\mathrm{b}}{}_{k+1}$ is small enough so that the innovation vector $\mathbf{y}^{\mathrm{o}}{}_{k+1} - H\mathbf{x}^{\mathrm{b}}{}_{k+1} = H\mathbf{x}^{\mathrm{t}}{}_{k+1} - H\mathbf{x}^{\mathrm{b}}{}_{k+1} + \varepsilon_{k+1}$ can be approximated by $H'(\mathbf{x}^{\mathrm{t}}{}_{k+1} - \mathbf{x}^{\mathrm{b}}{}_{k+1}) + \varepsilon_{k+1}$, calculations show that Eqs. (4.11) and (4.12) can be respectively replaced by

$$\begin{aligned}\mathbf{x}^{\mathrm{a}}{}_{k+1} &= \mathbf{x}^{\mathrm{b}}{}_{k+1} \\ &+\mathbf{P}^{\mathrm{b}}{}_{k+1}H'^{\mathrm{T}}[H'\mathbf{P}^{\mathrm{b}}{}_{k+1}H'^{\mathrm{T}} + \mathbf{R}]^{-1} \\ &(\mathbf{y}^{\mathrm{o}}{}_{k+1} - H\mathbf{x}^{\mathrm{b}}{}_{k+1})\end{aligned} \tag{5.9}$$

and

$$\begin{aligned}\mathbf{P}^{\mathrm{a}}{}_{k+1} &= \mathbf{P}^{\mathrm{b}}{}_{k+1} \\ &-\mathbf{P}^{\mathrm{b}}{}_{k+1}H'^{\mathrm{T}}[H'\mathbf{P}^{\mathrm{b}}{}_{k+1}H'^{\mathrm{T}} + \mathbf{R}]^{-1} \\ &H'\mathbf{P}^{\mathrm{b}}{}_{k+1}\end{aligned} \tag{5.10}$$

In these equations, H has been replaced by the jacobian H', except in the expression for the innovation vector. The algorithm defined by eqs (5.8 to 10), to which the nonlinear analogue to (4.9) must be added, is called *extended Kalman filtering* (see, *e.g.*, Jazwinski, 1970). It is valid whenever the differences between the real and estimated states of the system are small enough to allow local linearizations as just described.

A similar argument holds for variational assimilation. For a linear model and linear observation operators, the objective function (5.2) will be a quadratic function of the initial state $\mathbf{x}_0$. For a nonlinear model or a nonlinear observation operator, the objective function will not be quadratic, but will remain approximately quadratic in a neighbourhood of its minimum. If the initial uncertainty on the state of the system (defined for example by the point from which the minimization process is initiated) is small enough to ensure that, at any stage of the assimilation, the estimated state of the system will always lie within that neighbourhood, the theoretical nonlinearity of the objective function will have no practical effect. In particular, the minimizing solution will be the BLUE of the state of the flow.

The linear hypothesis made previously, and the associated logic of statistical linear estimation, including in particular the equivalence between sequential and variational assimilation, will therefore be valid if the differences between the real and estimated states of the system are always small enough to allow the local linearizations described above. The validity of this so-called *tangent linear approximation* has been checked systematically in a number of situations. For instance, Lacarra and Talagrand (1988) have shown on a barotropic model that, for realistic amplitudes of the error on the initial state of the flow, a linear approximation for the evolution of the forecast error is valid up to about 48 or 72 hours. Similarly, Thépaut and Moll (1990) have shown that, within the uncertainty existing in practice on the atmospheric profiles of temperature and humidity, the tangent linear hypothesis is valid for the radiance observations performed by the TIROS Operational Vertical Sounder (TOVS) carried by the

satellites of the NOAA series. In addition, many results have indirectly confirmed the validity of the tangent linear hypothesis. This means that sequential or variational assimilation can confidently be expected to produce reasonable estimates of the state of the flow. But it certainly does not mean that there do not remain problems. There must exist limitations to the validity of the tangent linear hypothesis, due for instance to the presence, in the physically most realistic models, of processes capable of inducing sharp variations in the model fields. These limitations have so far not been clearly identified, and research on fully nonlinear assimilation is only starting (see, *e.g.*, Miller *et al.*, 1994, or Pires *et al.*, 1996). But there is also no doubt that much development work remains to be done within the context of statistical linear estimation.

These considerations apply primarily to the atmosphere. Concerning the ocean, Evensen (1992, 1994) has shown strong evidence that the tangent linear approximation may not always be valid, essentially because the temporal density of observations is too low. In the context of sequential assimilation, Evensen suggests the use of an "Ensemble Kalman filtering", in which the temporal evolution of the estimation error covariance matrix is computed, not through a formula of form (4.10), but through an ensemble of forecasts performed with the fully nonlinear model.

6. Assimilation of "indirect" observations

It has been shown above that, in order to implement statistical estimation, it is necessary to know, for each individual observation, what has been measured, and with which accuracy. A rather general practice so far, when dealing with "indirect" satellite observations, has been to first "invert" them to "geophysical variables", such as for instance temperatures and humidities, and then to use the inverted fields as observations in the assimilation algorithm. This is commonly done, for instance, for radiance observations, which are inverted to produce estimates of the atmospheric temperature and humidity profiles. Now, such a preliminary inversion is by no means necessary. It does not avoid the basic need for the definition of an appropriate observation operator and for the specification of the associated observation error. And it usually requires a background which itself depends on the other available observations. This leads to interdependence of the errors associated with the various "observations" used in the assimilation. The problems raised by such an interdependence can be solved in the context of statistical estimation, but it certainly seems preferable to avoid them in the first place. For these reasons, and also in order to define a systematic approach to be followed for any type of observations, the tendency is now to avoid as much as possible preliminary processing of the observations before the assimilation, and to introduce raw observations in the assimilation algorithm with an appropriate observation operator. In the case of radiance measurements, the associated problems have been studied, among others, by Eyre (1989a and b) and Thépaut and Moll (1990). Several meteorological services are taking steps to directly incorporate radiance measurements in their operational assimilation algorithms. The same general trend is followed for all types of measurements : for instance, radial winds measured by Doppler effect either from ground-based radars or from satellite-borne lidars can be assimilated through an observation operator which reduces to the computation of the wind component along the appropriate direction (see respectively, *e.g.*, Sun *et al.*, 1991, and Courtier *et al.*, 1992). It is presumably the same approach which will be followed for the assimilation of observations made by future observing systems.

7. Conclusions

Our primary purpose in these notes was to describe the principles that lie at the basis of assimilation of observations. These principles are those of statistical linear estimation, and it has been shown that they lead to a generalised least-squares approach, amounting to minimizing a measure of the difference between the available observations and the state to be estimated. The word "observations" must be taken here in a very broad sense, to include all information available in quantitative form, and in particular the equations governing the assimilating model. Two classes of algorithms, sequential and variational assimilation, can be used for actually performing the required computations. Both algorithms are costly, and neither of them can be considered at the present stage as intrinsically superior.

Whatever the algorithm used, it is necessary to specify, for each individual piece of information used in the assimilation, the relationship of that particular piece of information with the variables to be estimated, and the accuracy of that relationship. This amounts to expressing the available information under the general form (3.9), where the matrix Γ must in the most general case be replaced by a nonlinear operator Γ. The latter expresses the relationship between the available information and the variables to be estimated. As for the corresponding accuracy, it is defined, in the basically linear approach described here, by the first and second order statistical moments of the error ζ (mean $E(\zeta)$ and covariances $S = E(\zeta\zeta^{\mathrm{T}})$). Any assimilation algorithm requires hypotheses, either explicit or implicit, on what Γ and the statistical moments of ζ are.

This leads us to our final remark. Estimating Γ and the statistical moments of ζ is by itself an estimation problem, *a priori* no easier than estimat-

ing the state of the atmospheric flow at a given time. One difference is that those quantities can be estimated by appropriate statistical accumulation. And, as concerns Γ, the physics of the measurement process is of course also fundamental. It has been repeatedly mentioned in these notes that statistical estimation, in addition to the BLUE of the state of the sytem, produces the covariances of the associated estimation errors. These covariances depend on Γ and Σ (Eq. 3.13). Any disagreement between the predicted analysis-minus-observations differences and the *a posteriori* effectively observed differences must therefore be due to inaccurate estimation of $E(\zeta)$ and/or Σ, and must be usable for improving the corresponding estimates. As for the observation operator Γ, it is in principle possible to determine it as the statistical minimizer of the innovation vector. Work along these lines of *adaptive filtering* has been done by Daley (1992), Dee (1993), Hoang Hong *et al.* (1995) and Blanchet (pers. com.).

References

Anderson, B.D.O., 1979, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, 357 pp.

Bennett, A.F., 1992, *Inverse Methods in Physical Oceanography*, Cambridge University Press, Cambridge, United Kingdom, 346 pp..

Bennett, A.F., L.M. Leslie, C.R. Hagelberg and P.E. Powers, 1993, Tropical Cyclone Prediction Using a Barotropic Model Initialized by a Generalized Inverse Method, *Mon. Wea. Rev.*, **121**, 1714–1729.

Bergthorsson, P. and B. Döös, 1955, Numerical weather map analysis, *Tellus*, **5**, 329–340.

Bouttier, F., 1994, A Dynamical Estimation of Forecast Error Covariances in an Assimilation System, *Mon. Wea. Rev.*, **122**, 2376–2390.

Cohn, S.E. and D.F. Parrish, 1991, The Behavior of Forecast Error Covariances for a Kalman Filter in Two Dimensions, *Mon. Wea. Rev.*, **119**, 1757–1785.

Courtier, P., P. Gauthier, F. Rabier, P. Flamant, A. Dabas, F. Lieutaud and H. Renault, 1992, *Study of preparation of the use of Doppler wind lidar information in meteorological assimilation systems*, Final report, ESA Contract 8850/90/HGE-I, ESA, Paris.

Courtier, P. and O. Talagrand, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation. I : Numerical results, *Q. J. R. Meteorol. Soc.*, **113**, 1329–1347.

Courtier, P. and O. Talagrand, 1990, Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations, *Tellus*, **42A**, 531–549.

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.

Cressman, G.P. 1959, An operational objective analysis scheme, *Mon. Wea. Rev.*, **87**, 367–374.

Daley, R., 1991, *Atmospheric Data Analysis*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, Cambridge, 457 pp.

Daley, R., 1992, The Lagged Innovation Covariance : A Performance Diagnostic for Atmospheric Data Assimilation, *Mon. Wea. Rev.*, **120**, 178–196.

Dee, D.P., 1991, Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.*, **117**, 365–384.

Dee, D.P., 1993, A simple scheme for tuning forecast error covariance parameters, in *Variational assimilation, with special emphasis on three-dimensional aspects*, ECMWF, Reading, England, 191–205.

Derber, J.C., 1987, Variational four dimensional analysis using quasi-geostrophic constaints, *Mon. Wea. Rev.*, **115**, 998–1008.

Eliassen, A., 1954, *Provisional report on calculation of spatial covariance and autocorrelation of the pressure field*, Report no 5, Videnskaps-Akademiets Institutt for Vaer-Og Klimaforskning, Oslo, Norway, 12 pp.. Reprinted in Bengtsson, L., M. Ghil and E. Källén, (editors), 1981, *Dynamic Meteorology. Data Assimilation Methods*, Springer Verlag, New York, USA, 330 pp., 319-328.

Evensen, G., 1992, Using the Extended Kalman Filter with a Multilayer Quasigeostrophic Ocean Model, *J. Geophys. Res.*, **97** (C11), 17,905–17,924.

Evensen, G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, **99** (C5), 10,143–10,162.

Eyre, J., 1989a, Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS, *Q. J. R. Meteorol. Soc.*, **115**, 1001–1026.

Eyre, J., 1989b, Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. II: Application to TOVS data, *Q. J. R. Meteorol. Soc.*, **115**, 1027–1037.

Gandin, L.S., 1963, *Objective analysis of meteorological fields*, Gidrometeor. Izd., Leningrad (in Russian), (English Translation by Israel Program for Scientific Translations, Jerusalem, 1965).

Gaspar, P. and C. Wunsch, 1989, Estimates from Altimeter Data of Barotropic Rossby Waves in the Northwestern Atlantic Ocean, *J. Phys. Oceanogr.*, **19**, 1821–1844.

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. in Geophys.*, **33**, 141–266.

Gill, P.E., W. Murray and M.H. Wright, 1982, *Practical Optimization*, Academic Press, London.

Greiner, E. and C. Perigaud, 1994, Assimilation of Geosat Altimetric Data in a Nonlinear Reduced-Gravity Model of the Indian Ocean. Part 1: Adjoint Approach and Model-Data Consistency, *J. Phys. Oceanogr.*, **24**, 1783–1804.

Hinkelmann, K., 1951, Der Mechanismus des meteorologischen Lämes, *Tellus*, **3**, 285–296.

Hoang Hong S., P. De Mey, O. Talagrand and R. Baraille, *Assimilation of Altimeter Data in a Multilayer Quasi-Geostrophic Ocean Model by Simple*

Nonlinear Adapttive Filter, Proceedings, International Symposium on Assimilation of Observations in Meteorology and Oceanography, World Meteorological Organization, Tokyo, Japan, March 1995.

Jazwinski, A.H., 1970, *Stochastic Processes and Filtering Theory*, Academic Press, New-York, 376 pp.

Kalman, R.E., 1960, A new approach to linear filtering and prediction problems, *J. Basic Eng.*, **82D**, 35–45.

Lacarra, J.F. and O. Talagrand, 1988, Short-range evolution of small perturbations in a barotropic model, *Tellus*, **40A**, 81–95.

Le Dimet, F.X. and O. Talagrand, 1986, Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects, *Tellus*, **38A**, 97–110.

Lewis, J.M. and J.C. Derber, 1985, The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus*, **37A**, 309–322.

Lions, P.L., 1971, *Optimal control of systems governed by partial differential equations*, Springer-Verlag, Berlin, 396 pp.

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

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

Lorenc, A., 1988, Optimal nonlinear objective analysis, *Q. J. R. Meteorol. Soc.*, **114**, 205–240.

Lorenz, E.N., 1963, Deterministic Nonperiodic Flow, *J. Atmos. Sci.*, **20**, 130–141.

Miller, R.N., M. Ghil and F. Gauthiez, 1994, Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems, *J. Atmos. Sci.*, **51**, 1037–1056.

Morgenstern, J., 1984, *How to compute fast a function and all its derivatives. A variation on the theorem of Baur-Stressen*, Report No 49, Laboratoire CNRS 168, Universit 罪 de Nice, Nice, France, 5 pp.

Navon, I.M., X. Zou, K. Johnson, J. Derber and J. Sela, 1991, *Variational data assimilation with an adiabatic version of the NMC spectral model*, Report FSU-SCRI-91-13, The Florida State University, Tallahassee, Florida, USA, 43pp.

Oort, A.H., 1989, Angular Momentum Cycle in the Atmosphere-Ocean-Solid Earth System, *Bull. Amer. Meteor. Soc.*, **70**, 1231–1242.

Parrish, D. and D. Derber, 1992, The National Meteorological Center's Spectral Statistical-Interpolation Analysis System, *Mon. Wea. Rev.*, **120**, 1747–1763.

Penenko, V.V. and N.N. Obraztsov, 1976, A variational initialization method for the fields of the meteorological elements (English translation), *Soviet Meteorol. Hydrol.*, no 11, 1–11.

Pires, C., R. Vautard and O. Talagrand, 1995, On extending the limits of variational assimilation in nonlinear chaotic systems, *Tellus*, **48A**, 96–121.

Rabier, F. and P. Courtier, 1992, Four-dimensional assimilation in the presence of baroclinic instability, *Q. J. R. Meteorol. Soc.*, **118**, 649–672.

Rabier, F., P. Courtier, J. Pailleux, O. Talagrand and D. Vasiljevic, 1993, A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis, *Q. J. R. Meteorol. Soc.*, **119**, 845–880.

Salstein, D.A. and R.D. Rosen, 1986, Earth Rotation as a Proxy for Interannual Variability in Atmospheric Circulation, 1860-Present, *J. Climate App. Meteor.*, **25**, 1870–1877.

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

Sheinbaum, J. and D.L.T. Anderson, 1990a, Variational Assimilation of XBT Data. Part I, *J. Phys. Oceanogr.*, **20**, 672–688.

Sheinbaum, J. and D.L.T. Anderson, 1990b, Variational Assimilation of XBT Data. Part II : Sensitivity Studies and Use of Smoothing Constraints, *J. Phys. Oceanogr.*, **20**, 689–704.

Sun, J., D.W. Flicker and D.K. Lilly, 1991, Recovery of Three-Dimensional Wind and Temperature Fields from Simulated Single-Doppler Radar Data, *J. Atmos. Sci.*, **48**, 876–890.

Talagrand, O. and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation. I : Theory, *Q. J. R. Meteorol. Soc.*, **113**, 1311–1328.

Tarantola, A., 1987, *Inverse Problem Theory*, Elsevier, Amsterdam, The Netherlands, 613 pp.

Temperton, C., 1988, Implicit normal mode initialization, *Mon. Wea. Rev.*, **116**, 1013–1031.

Thacker, W.C. and R.B. Long, 1988, Fitting dynamics to data, *J. Geophys. Res.*, **93**, 1227–1240.

Thépaut, J.N. and P. Courtier, 1991, Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model, *Q. J. R. Meteorol. Soc.*, **117**, 1225–1254.

Thépaut, J.-N., R.N. Hoffman and P. Courtier, 1993, Interactions of Dynamics and Observations in a Four-Dimensional Variational Assimilation, *Mon. Wea. Rev.*, **121**, 3393–3414.

Thépaut, J. N., and P. Moll, 1990, Variational inversion of simulated TOVS radiances using the adjoint technique, *Q. J. R. Meteorol. Soc.*, **116**, 1425–1448.

Thiébaux, H.J. and M.A. Pedder, 1987, Spatial Objective Analysis, with Applications in Atmospheric Sciences, Academic Press, New York, United States.

Vautard, R., 1990, Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors, *Mon. Wea. Rev.*, **118**, 2056–2081.

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