



Centre for Environmental Systems Research,
School of Environment & Life Science.

Four-Dimensional Data Assimilation and Numerical Weather Prediction

Fathalla A. Rihan & Chris G. Collier

CERS Data Assimilation Technical Report No. 1
July 2003.

Available from CESR, Salford University,
and at the WWW:
<http://www.ties.salford.ac.uk/frihan/>

Four-Dimensional Data Assimilation and Numerical Weather Prediction

Fathalla A. Rihan¹ Chris G. Collier²

^a*School of Environmental and Life Sciences, The University of Salford, Manchester M5 4WT, England.*

Abstract

All forecast models, whether they represent the state of the weather, the spread of a disease, or levels of economic activity, contain unknown parameters. These parameters may be the model's initial conditions, its boundary conditions, or other tunable parameters which have to be determined. *Four dimensional variational data assimilation* (4D-Var) is a method of estimating this set of parameters by optimizing the fit between the solution of the model and a set of observations which the model is meant to predict. The *four dimensional* nature of 4D-Var reflects the fact that the observation set spans not only three dimensional space, but also a time domain.

Although the method of 4D-Var described in this report is not restricted to any particular system, the application described here has a Numerical Weather Prediction (NWP) model at its core, and the parameters to be determined are the *initial conditions* of the model.

The purpose of this report is to give a survey covering assimilation of Doppler radar wind data into a high-resolution NWP model. Some associated problems, such as sensitivity to small variations in the initial conditions or due to small changes in the background variables, and biases due to nonlinearity are also studied.

Key words: NWP, Data Assimilation, 3D-Var, 4D-Var, Adjoint, Doppler Radar, Sensitivity, Nonlinear Bias.

Contents

1	Introduction	3
2	What is Data Assimilation?	5
2.1	The cost function	6
2.2	Sequential Assimilation– 3D-Var	6

¹ F.Rihan@salford.ac.uk

² C.G.Collier@salford.ac.uk

3	Variational Assimilation–4D-Var	9
3.1	Adjoint method for minimizing the cost function	11
4	Assimilation of Radar Wind Data	12
4.1	3D-Var for Doppler radial winds	13
4.2	4D-Var for Doppler radial winds and reflectivities	15
5	Numerical Implementation Model	17
5.1	Basic equations	18
5.2	Numerical algorithm	19
6	Why are Weather Forecasts Sometimes Wrong?	20
6.1	Sensitivity to initial conditions	21
6.2	Biases due to nonlinearity	24
7	Conclusion	25
	References	26

1 Introduction

Weather forecast models use observations and numerical methods to represent the atmosphere. However, to predict weather more than 24 hours in advance, forecasters rely on Numerical Weather Predictions (NWP) that are generated by integrating the governing dynamical equations in time. All possible atmospheric information is collected for a given time, and then diagnosed to produce regular descriptions of the atmosphere at that time. This analysis is applied, as the initial conditions, to the initial value differential equations that reflect the physical behaviour of the atmosphere, integrating those equations to predict future events. One of the major challenges in NWP is obtaining accurate initial conditions. A small error in the initial conditions affects the forecast in space and time.

Doppler radars, which provide observations of radial velocity and reflectivity of hydrometeors with spatial resolutions of a few kilometers every 3-10 min both in clear air and inside heavy rainfall regions, are practically the only instrument capable of sampling the four-dimensional structure of severe storm flows. However, consideration needs to be given to identify the optimum way of assimilating these observations into numerical weather prediction models. In this study we survey these needs, and some methods that have been used

in assimilating such data to produce as accurately as possible the initial state of the atmospheric model.

Of course, the fact that weather predictions may be wrong demonstrates that there are problems with the models, the initial conditions, or both. The error in NWP comes from two main sources: One is the inherent inaccuracies in the models used to describe the atmosphere. The second comes from the sensitivity of the output to small changes in the initial conditions [22]; further discussion of these issues is given in §6.

The goal of Four Dimensional (space and time) Data Assimilation (4DDA) is to incorporate actual observations (satellite, radar, ship, land surface, balloon) into mathematical and computational models in order to create a unified, complete description of the atmosphere. This can be used by community-wide scientists to study important phenomena associated with environmental global change; such as greenhouse-gas research, ozone-loss research, or atmospheric pollution. In other words, 4DDA is a process whereby a state forecast and observations are combined to form a best estimate, or *analysis* of the state.

Optimization of the NWP initial conditions can be regarded as a class of inverse problem. Indeed, we assume that the forward problem is soluble, i.e. given a set of initial conditions, a set of forward models (governing equations) can be run to predict the observations. The most important forward model is the core NWP forecast model which gives the evolution in time of the model's state. Examples of such predicted observations commonly used in a NWP context are the meteorological variables (wind, temperature and humidity) interpolated from the model grid to the position of an instrument (in a field station, or on a sonde or aircraft). Of course, the *predicted observations* may or may not match the *actual observations*; this depends on the choice of initial conditions (and on the suitability of the models). We are here interested in solving the inverse problem which can be posed as follows: *what set of initial conditions will seed the models to best predict the known observations?* Such inverse problems are in general much harder to solve than the forward analogue. Indeed, the inverse problem relies on the existence of the forward models themselves which are run many times in an iterative fashion to give the analysis (see [28,35]).

Although a considerable progress has been made in assimilation of Doppler radar winds, this area remains important and demands further investigations. This report carries out a survey of four-dimensional variational data assimilation through adjoint methods (see §2 & §3). An extension to assimilating Doppler radar winds into atmospheric model is presented in §4, and numerical implementation is discussed in §5. Issues associated with data assimilation and NWP, such as sensitivity analysis and nonlinear bias, are also considered (see §6).

2 What is Data Assimilation?

Recently, there has been growing interest in *data assimilation* in NWP (see, e.g., [5,16]). As indicated in the introduction to this report, NWP is an initial-boundary value problem: given an estimate of the present state of the atmosphere, the model simulates (forecasts) its evolution. The problem of determining the initial conditions for a forecast model is very important and complex, and has become a science in itself. Clearly, specification of proper initial conditions and boundary conditions for the numerical dynamical models is essential in order to have a well-posed³ problem and then a good forecast model. Hence *Data assimilation* can be described as the process through which all the available information is used in order to determine, as accurately as possible, the state of the atmospheric flow on a regular grid. Basically, current atmospheric data assimilation systems use two sources of data: observations, and a recent forecast valid at the current time. The most recent forecast is called the *background* (sometimes called the "first guess") field because it is the best guess of the state of atmosphere before any observations are taken. Data assimilation for NWP uses the average of two pieces of data⁴ representing the observations and the background⁵, combined with the atmospheric model, in order to produce the best possible model initial state.

There are two broad classes of data assimilation methods: *sequential* data assimilation and *variational* data assimilation. Firstly, sequential data assimilation involves an analysis produced by combining a forecast background and the observation available at a given time. The numerical method is then integrated forward to the next observation time, starting from the analysis initial conditions. Each new observational element is used for correcting the last estimate, so that the best fit of the model solution to observations is achieved at the end of the assimilation period through the propagation of information from the past in a sequential manner. Secondly, the variational assimilation seeks an optimal fit of the model solution to observations over an assimilation period by adjusting the estimation states in this period simultaneously. To achieve this goal, a numerical model is used to link the state of the atmosphere at different times. The estimated states over the assimilation period are influenced by all the observations distributed in time. The information is propagated both from the past into the future and from the future into the past; we shall discuss this further in §3.

The *variational approach* has been extensively used in data assimilation for meteorological models and shows promising results for NWP [13,14,34]. This

³ A well-posed initial/boundary problem has a unique solution that depends continuously of the initial/boundary conditions.

⁴ Each piece of data should be weighted according to its accuracy.

⁵ In most cases the data is sparse and indirectly related to the model variables. In order to make the model a well posed problem it is necessary to rely on some *background* information in the form of the *a priori* estimate of the model state.

approach includes *three-dimensional variational* data assimilation (3D-Var) and the *four-dimensional variational* data assimilation (4D-Var). The 4D-Var (four dimensional variational) searches for an optimal set of model parameters (e.g., optimal initial state of the model) which minimizes the discrepancies between the model forecast and time distributed observational data over the assimilation window. A practical implementation of the minimization process requires a fast and accurate evaluation of the gradient of a cost function which may be provided by adjoint modelling; we shall discuss this further in §3.1.

2.1 The cost function

The assimilation problem includes unknown errors that come from the model, the background, and the observations. Assume that these errors are independent, then the total probability density function (pdf) is a product of the component pdf's of the model P_m , the background P_b , and the observation P_o

$$P = P_m P_b P_o \approx \exp(-\log P_m - \log P_b - \log P_o). \quad (1)$$

The maximum of the probability density function can be expressed in term of a cost function \mathcal{J} as follows:

$$\min \mathcal{J} = \min(-\log P_m - \log P_b - \log P_o) \equiv \min(\mathcal{J}_m + \mathcal{J}_b + \mathcal{J}_o). \quad (2)$$

This means that the maximum likelihood approach to solving the inverse problem requires the minimization of a *cost function* (see next). A detailed description of the various assumptions used by data assimilation techniques, including probabilistic interpretation may be found in [5].

2.2 Sequential Assimilation– 3D-Var

If all the variable information and associated uncertainty are given and the errors are assumed to be an unbiased Gaussian distribution, an idealized equation for finding the optimal estimate of the atmospheric flow can be derived based on Bayesian statistics where the probability density functions are given by

$$P_b = \frac{1}{\sqrt{2\pi|\mathbf{B}|}} \exp\left(-\frac{1}{2}(\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x})\right), \quad (3)$$

$$P_o = \frac{1}{\sqrt{2\pi|\mathbf{E}|}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathcal{H}[\mathbf{x}])^T \mathbf{E}^{-1}(\mathbf{y} - \mathcal{H}[\mathbf{x}])\right). \quad (4)$$

If we ignore the model's error, the statistical cost function $\mathcal{J}[\mathbf{x}]$ takes the form (see [21])

$$\begin{aligned}\mathcal{J}[\mathbf{x}] &= \mathcal{J}_b + \mathcal{J}_o \\ &= \frac{1}{2}(\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathcal{H}[\mathbf{x}])^T \mathbf{E}^{-1}(\mathbf{y} - \mathcal{H}[\mathbf{x}]),\end{aligned}\quad (5)$$

where \mathbf{x} denotes the analysis vector, \mathbf{x}_b the background vector, and \mathbf{y} the observation vector. The first term \mathcal{J}_b is a measure of the distance of the initial state from the background estimate and the second term \mathcal{J}_o is a measure of the distance between the model trajectory and observations over the assimilation window. \mathbf{B} and \mathbf{E} are background and observation error covariance matrices, respectively (T denotes the matrix transpose). Since \mathbf{x} and \mathbf{y} are different variables and on different grids, the observation operator \mathcal{H} represents an analytical function⁶ that relates the model variables to the observation variable and a transformation between the different grid meshes⁷. It should also be noted that radar observations are concentrated within a certain area of the radar, and data voids are often present in the model domain. A background estimate \mathbf{x}_b can be used to fill these data voids and to provide a first guess for the minimization procedure.

The nonlinear observation operation \mathcal{H} can be linearized as

$$\mathcal{H}[\mathbf{x} + \delta\mathbf{x}] = \mathcal{H}[\mathbf{x}] + \mathbf{H}\delta\mathbf{x}.\quad (6)$$

The matrix \mathbf{H} , denoted the *linear* observation operator with elements $h_{ij} = \partial\mathcal{H}_i/\partial x_j$, transforms vectors in model space into their corresponding values in observation space. Its transpose or adjoint \mathbf{H}^T transforms vectors in observation space into vectors in model space. Therefore, we can expand the second term of (5), the observation differences, assuming that the analysis is a close approximation to the truth and therefore to the observation. Linearizing \mathcal{H} around the background value \mathbf{x}_b gives

$$\mathbf{y} - \mathcal{H}[\mathbf{x}] = \mathbf{y} - \mathcal{H}[\mathbf{x}_b + (\mathbf{x} - \mathbf{x}_b)] = (\mathbf{y} - \mathcal{H}[\mathbf{x}_b]) - \mathbf{H}(\mathbf{x} - \mathbf{x}_b).\quad (7)$$

The *analytical equation* (exact solution) that defines the minimum of \mathcal{J} (the best analysis $\mathbf{x} = \mathbf{x}_a$) can be obtained by taking the first order derivative of \mathcal{J} and setting it to zero ($\nabla_x \mathcal{J}[\mathbf{x}_a] = 0$), which yields the best analysis:

⁶ for instance, the relation between Cartesian components u, v, w of the wind velocity from the model and the radial velocity v_r from Doppler-radar observations.

⁷ In practice there are fewer observations than variables in the model and they are irregularly disposed, hence the only way to compare the observations with the state vector is through the use of an observation operator \mathcal{H} .

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}(\mathbf{y} - \mathcal{H}[\mathbf{x}_b]) \quad (8)$$

where

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{E})^{-1}. \quad (9)$$

The vector $(\mathbf{y} - \mathcal{H}[\mathbf{x}_b])$ in Eq. (8) is often referred to as the observation *innovation vector* (or departure vector) and the matrix \mathbf{W} is called the *gain matrix*. Eq. (8) shows that the analysis is the sum of prior information and a correction term, which is proportional to the difference between the new information brought in from the observation and the prior information (background). Directly solving (8) & (9) is not an easy task due to the large dimension of the matrices involved. Simplification and approximation needs to be carried out in order to reduce the computation. Different approximations lead to different techniques.

In practice, the 3D-Var technique attempts to find an approximate solution of the minimum of Eq. (5) by iteratively minimizing the cost function. The commonly used minimization algorithm is a form of conjugate gradient or quasi-Newton iterative methods (see [11]).

The 3D-Var cost function (5) can be written as an incremental formula⁸

$$\mathcal{J}[\delta\mathbf{x}] = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}[\mathbf{H}\delta\mathbf{x} - \mathbf{d}]^T\mathbf{E}^{-1}[\mathbf{H}\delta\mathbf{x} - \mathbf{d}], \quad (10)$$

where $\delta\mathbf{x} = (\mathbf{x}_a - \mathbf{x}_b)$ is the *analysis increment vector* in the analysis space, and $\mathbf{d} = (\mathbf{y} - \mathcal{H}[\mathbf{x}_b])$ is the *observational innovation vector* in the observation space. The model state vector \mathbf{x} may include the wind components, temperature, humidity and the surface pressure.

In the iterative procedure, the cost function and its gradient are computed at each iteration and used to define the best descending direction towards the optimal direction. The error covariance of the forecast background \mathbf{B} are usually modelled based on some simple hypotheses on the shape and spatial extension of the pre-assumed covariance functions. Gao *et al.* [8] applied this technique to the convective-scale and demonstrated its potential in wind retrieval from Doppler-radar observations; see also [19].

We may note from Eq. (8) an important property: the result obtained by the 3D-Var approach (finding the optimal analysis that minimizes the distance to the observations weighted by the *inverse of the error variance*) is equivalent to the result obtained by the Optimal Interpolation (OI) approach (finding the optimal *weights* that minimizes the *analysis error variance* through a least

⁸ Minimizing the cost function in a full forecast resolution is sometimes out of the bounds of computational divisibilities. Thus variational problem can be solved at lower resolution using the incremental formula.

squares approach) [21]. However, in the 3D-Var approach the *control variable* is the *analysis* \mathbf{x}_a , not the *weights* as in OI. Both of these methods do not include the evolution of the model in the assimilation. Despite the fact that there is a formal equivalence between 3D-Var and OI techniques, there are some advantages of the 3D-Var over the OI. In 3D-Var, the cost function is minimized using global minimizing algorithms, as a result it makes unnecessary many of simplifying approximations required by OI. In addition, there is no data selection, in 3D-Var, all available data are used simultaneously. This avoids jumpiness in the boundaries between regions that have selected different observations. For further discussions and comments about variational and OI techniques, we may refer to [16].

3 Variational Assimilation–4D-Var

The 4D-Var assimilation technique can be considered as an extension of 3D-Var in such a way that all observations distributed within a time window ($0 \leq i \leq N$) are taken into account in defining the cost function. A numerical model that is supposed to represent the evolution of the estimate vector \mathbf{x} is used as *a priori* information. Assume the nonlinear model equation is given in the form

$$\frac{\partial \mathbf{x}}{\partial t} = F(\mathbf{x}), \quad (11)$$

where F stands for the mathematical functions involved in the dynamical nonlinear model. If the observational error covariance \mathbf{E} is dependent of time, the *cost function* which measures the misfit between the model variables and both a prior estimate (background) and the observations is reformulated in 4D-Var to take the form

$$\begin{aligned} \mathcal{J}[\mathbf{x}(t_0)] &= \frac{1}{2} [\mathbf{x}_b(t_0) - \mathbf{x}(t_0)]^T \mathbf{B}^{-1} [\mathbf{x}_b(t_0) - \mathbf{x}(t_0)] \\ &\quad + \frac{1}{2} \sum_{i=0}^N [\mathbf{y}(t_i) - \mathcal{H}[\mathbf{x}(t_i)]]^T \mathbf{E}_i^{-1} [\mathbf{y}(t_i) - \mathcal{H}[\mathbf{x}(t_i)]] \\ &= \mathcal{J}_b + \mathcal{J}_o, \quad \mathcal{J}[\mathbf{x}_a(t_0)] = \min_{\mathbf{x}(t_0)} \mathcal{J}[\mathbf{x}(t_0)]. \end{aligned} \quad (12)$$

The control variable (the variable with respect to which the cost function is minimized) is the initial state of the model $\mathbf{x}(t_0)$, whereas the analysis at the end of the time interval is given by integration ($t_0 \leq t_n \leq t_N$) of the nonlinear model (11) to give

$$\mathbf{x}(t_n) = M[\mathbf{x}(t_0)]. \quad (13)$$

Thus the model is used as a *strong* constraint, i.e., the analysis has to satisfy the model equations. In other words, 4D-Var seeks an initial condition such

that the forecast best fits the observations within the assimilation interval. Thus 4D-Var analysis is performed using a continuous cycling procedure. The length of the assimilation cycle is $[0, 2\tau]$, consisting of assimilation period $[0, \tau]$ and forecasting period $[\tau, 2\tau]$ (see Fig. 1). An optimal initial condition is obtained from each cycle using data within the assimilation period. The analysis fields at the final time of the assimilation window are also used as first guess fields and background for the next analysis sequence. We may note that in 3D-Var, we apply the assimilation at only one point, say at $t = 0$ or $t = \tau/2$.

In practice, solving the minimization problem (12) can represent a major computing problem. Iterative minimization schemes require the estimation of the cost function gradient, $\nabla \mathcal{J}$, with respect to the control variable. The first term, \mathcal{J}_b , of the cost function is not a complicated term. The evaluation of the second term \mathcal{J}_o would seem to require N integrations of the forecast model from the analysis time to each of the observation times i , and even more for the computation of the gradient $\nabla \mathcal{J}_o$. Consequently, due to the complex nature of the 4D-Var for dealing with the large dimension problems, the *adjoint technique* is introduced to efficiently calculate the gradient of the cost function with respect to the control variable.

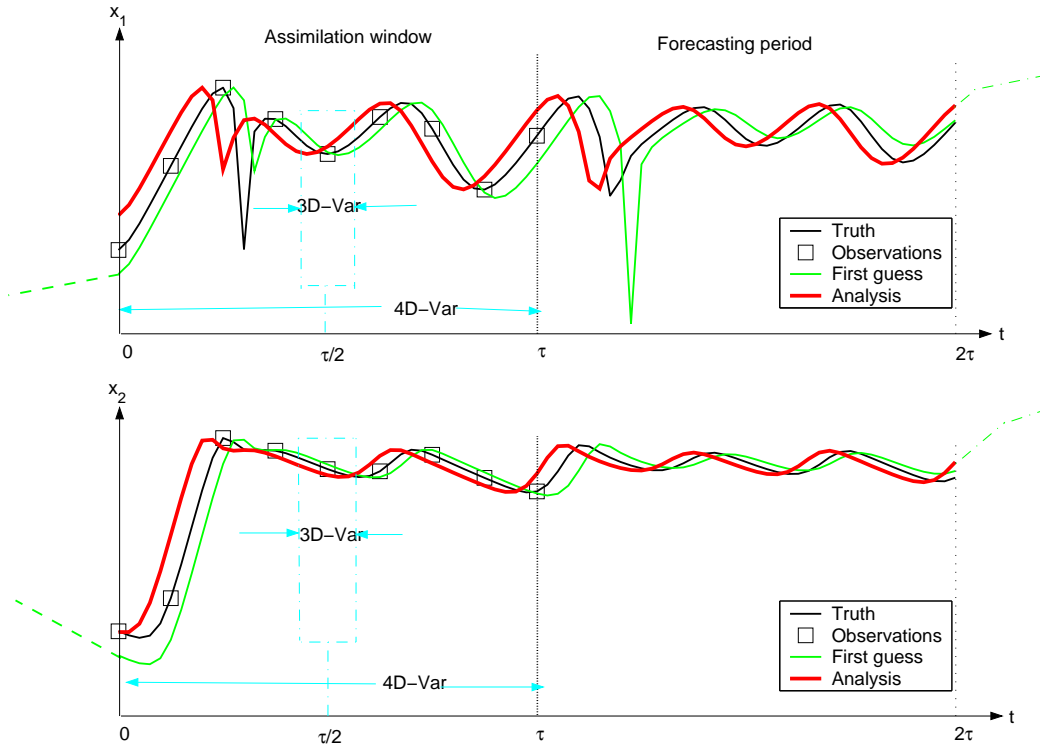


Fig. 1. Adjusting the initial conditions of the models using 3D-Var/4D-Var for the best fit. Every cycle, 2τ , a 3D-Var/4D-Var is performed to assimilate the most recent observations in $[0, \tau]$, using a segment of the previous forecast as background. This updates the initial model trajectory for the next cycle. This graph plots one solution of Lorenz equations, with true observations generated from a model run.

3.1 Adjoint method for minimizing the cost function

Adjoint methods are increasingly being developed for use in meteorology, data assimilation and sensitivity analysis. In data assimilation tasks, the adjoint model is used to compute (in reverse mode) the gradient of the cost function with respect to the control variable [10,17]. The main goal of using the adjoint method is to avoid repeat computing of the gradient of the cost function every iteration of the 4D-Var minimization routine during a forward integration. It has the advantages, especially for large systems, that the adjoint model reduces the run over time, and the computed gradient is exact.

Remark 1 *If we have a quadratic function $F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x}$, where \mathbf{A} is a symmetric matrix, then $\nabla F(\mathbf{x}) = \mathbf{A}\mathbf{x}$, and $\delta F = (\nabla F)^T \delta \mathbf{x}$.*

Using the above Remark, the gradient of the background component of (12) \mathcal{J}_b with respect to $\mathbf{x}(t_0)$ is given by

$$\nabla \mathcal{J}_b = \mathbf{B}^{-1}[\mathbf{x}(t_0) - \mathbf{x}_b(t_0)] \equiv \mathbf{B}^{-1} \delta \mathbf{x}(t_0). \quad (14)$$

However the gradient of the observational term of (12) \mathcal{J}_o is more complicated because, $\mathbf{x}(t_i)$ is a function of the control variable, $\mathbf{x}(t_i) = M_i[\mathbf{x}(t_0)]$.

Remark 2 *For a given nonlinear model (13), if we introduce a perturbation in the initial conditions $\delta \mathbf{x}(t_0)$, neglecting terms of order $[\delta \mathbf{x}(t_0)]^2$, then*

$$\mathbf{x}(t) + \delta \mathbf{x}(t) = M[\mathbf{x}(t_0) + \delta \mathbf{x}(t_0)] \approx M[\mathbf{x}(t_0)] + \frac{\partial M}{\partial \mathbf{x}} \delta \mathbf{x}(t_0) \quad (15)$$

so that the initial perturbation evolves like

$$\delta \mathbf{x}(t) = \mathbf{L}(t_0, t) \delta \mathbf{x}(t_0). \quad (16)$$

The Jacobian $\mathbf{L}(t_0, t) = \frac{\partial M}{\partial \mathbf{x}}$ is the tangent linear operator that propagates the perturbation from t_0 to t . If there are N steps between t_0 and $t_N = t$, this matrix is equal the product of matrices of corresponding each time step:

$$\mathbf{L}(t_0, t_N) = \mathbf{L}(t_{N-1}, t_N) \dots \mathbf{L}(t_1, t_2) \mathbf{L}(t_0, t_1) := \prod_{j=N-1}^0 \mathbf{L}(t_j, t_{j+1}). \quad (17)$$

Therefore, the adjoint (transpose of the linear tangent model) is given by

$$\mathbf{L}^T(t_N, t_0) = \prod_{j=0}^{N-1} \mathbf{L}^T(t_{j+1}, t_j). \quad (18)$$

Eq. (18) shows that the adjoint model advances a perturbation backwards in time, from the final to initial time.

From Remark 2 (Eq. (16)) we have

$$\frac{\partial(\mathbf{y}(t_i) - \mathcal{H}[\mathbf{x}(t_i)])}{\partial \mathbf{x}(t_0)} = \frac{\partial \mathcal{H}[\mathbf{x}(t_i)]}{\partial \mathbf{x}(t_i)} \frac{\partial M[\mathbf{x}(t_0)]}{\partial \mathbf{x}(t_0)} = \mathbf{H}_i \mathbf{L}(t_0, t_i). \quad (19)$$

Therefore, by using Remark 1 and Eq. (19), the gradient of the observation cost function with respect to the control variable $\mathbf{x}(t_0)$ is

$$\nabla \mathcal{J}_o = \sum_{i=0}^N \mathbf{L}^T(t_i, t_0) \mathbf{H}_i \mathbf{E}_i^{-1} [\mathcal{H}[\mathbf{x}(t_i)] - \mathbf{y}(t_i)]. \quad (20)$$

We note that parts of the backward adjoint integration are common to several time intervals, the summation in (20) can be arranged more conveniently. Assume for example, the assimilation period is between $t = t_0$ and $t = t_N$, i.e. we have $(N + 1)$ observations. We compute during the forward integration the weighted observation increments

$$\bar{\mathbf{d}}_i = \mathbf{H}_i \mathbf{E}_i^{-1} [\mathcal{H}[\mathbf{x}(t_i)] - \mathbf{y}(t_i)] := \mathbf{H}_i \mathbf{E}_i^{-1} \mathbf{d}_i.$$

The adjoint model $\mathbf{L}^T(t_i, t_{i-1}) := \mathbf{L}_{i-1}^T$ applied on a vector *advances* it from t_i to t_{i-1} . Then one can write (20) as

$$\nabla \mathcal{J}_o = \bar{\mathbf{d}}_0 + \mathbf{L}_0^T \left(\bar{\mathbf{d}}_1 + \mathbf{L}_1^T \left(\bar{\mathbf{d}}_2 + \dots + \mathbf{L}_{N-2}^T \left(\bar{\mathbf{d}}_{N-1} + \mathbf{L}_{N-1}^T \bar{\mathbf{d}}_N \right) \right) \right). \quad (21)$$

From (14) and (21) we obtain the gradient of the cost function, and the minimize algorithm modifies appropriately the control variable $\mathbf{x}(t_0)$. After this, a new forward integration and new observational increments are computed and the process is repeated.

From the above the 4D-Var can then be written in an incremental form with the cost function defined as

$$\begin{aligned} \mathcal{J}[\delta \mathbf{x}_0] = & \frac{1}{2} \delta \mathbf{x}_0^T \mathbf{B}^{-1} \delta \mathbf{x}_0 + \\ & \frac{1}{2} \sum_{i=1}^N [\mathbf{H}_i \mathbf{L}(t_0, t_i) \delta \mathbf{x}_0 - \mathbf{d}_i]^T \mathbf{E}_i^{-1} [\mathbf{H}_i \mathbf{L}(t_0, t_i) \delta \mathbf{x}_0 - \mathbf{d}_i], \end{aligned} \quad (22)$$

where, the observational increment, $\mathbf{d}_i := [\mathcal{H}[\mathbf{x}(t_i)] - \mathbf{y}(t_i)]$ and $\mathbf{x}_0 := \mathbf{x}(t_0)$.

4 Assimilation of Radar Wind Data

With the advent of operationally available Doppler radial wind data interest has increased in assimilating these data into NWP models. This interest

has been accelerated by the increasing use of limited area high resolution numerical models for *cloud scale prediction*. The limited area models require observations with high spatial-temporal resolution for determining the initial conditions. Doppler radar wind measurements are one possible source of information, albeit over limited areas within about 100 km of each radar site [6]. The *resolution* of *raw data* is however much higher than the resolution of the numerical models, and these data must be preprocessed, to be representative of the characteristic scale of the model, before the analysis: When several observations are too *close* together then they will be more *correlated* and as a result the forecast error correlations at the observation points will be *large*. In contrast, the individual (single) observations are less dependent and they will be given more weight in the analysis than observations that are close together. To reduce the representativeness error, as well as the computational cost, one may use (i) the vertical profiles of horizontally averaged wind in the form of Velocity Azimuth Display (VAD) technique, (ii) the observations of spare resolution, or (iii) calculate spatial averages from the raw data to generate the so called *super-observations*⁹. The generated data correspond more closely to the horizontal model resolutions than do the raw observations (see [1]).

4.1 3D-Var for Doppler radial winds

The 3D-Var has been used operationally to assimilate radar wind information in the form of VAD wind profiles (see [23]). Recently, 3D-Var, for the Swedish Meteorological and Hydrological Institute (SMHI) High Resolution Limited Area Model (HIRLAM), has been developed for assimilating Doppler radar wind data either as radial *super-observations* or as VAD wind profiles (see [12,18,19]).

Under the assumption that the background and observation errors are Gaussian, random and independent of each other, the optimal estimate of the radial wind in the analysis space is given by the increment cost function (10),

$$\mathcal{J}[\delta\mathbf{x}] = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}[\mathbf{H}\delta\mathbf{x} - \mathbf{y} + \mathcal{H}\mathbf{x}_b]^T\mathbf{E}^{-1}[\mathbf{H}\delta\mathbf{x} - \mathbf{y} + \mathcal{H}\mathbf{x}_b], \quad (23)$$

where \mathbf{x} is the state vector of the analysis (i.e., the estimated radial winds), \mathbf{x}_b is the state variable of the background radial winds, and \mathbf{y} denotes the observed radial winds in the observation space. Some constructions of the background and observation error covariance matrices \mathbf{B} and \mathbf{E} are given in [5,36]. To avoid the computationally overwhelming problem of inverting the covariance matrix \mathbf{B} in the minimization of the cost function (23) and to accelerate the convergence of the minimization algorithm, a pre-conditioning

⁹ *Super-observations* are spatial averages of raw measurements with different resolutions.

of the minimization problem is needed (see [20]). The purpose of the pre-conditioning is to ensure that the Hessian matrix of the background error term (the second derivative of \mathcal{J}_b with respect to the control variables) is an identity matrix. This can be achieved by defining a variable \mathbf{U} to be applied to the assimilation increment $\delta\mathbf{x}$ ($\mathbf{U}\delta\mathbf{x} \equiv \mathcal{X}$) such that it transforms the forecast error ϵ in the model space into $\tilde{\epsilon}$, a variable of an *identity* covariance matrix (i.e., $\langle \tilde{\epsilon}, \tilde{\epsilon}^T \rangle = \mathbf{I}$, where $\langle \cdot, \cdot \rangle$ is an inner product). This change of variable can be written as $\epsilon = \mathbf{U}^{-1}\tilde{\epsilon}$. Thus

$$\mathbf{B} = \langle \epsilon, \epsilon \rangle = \mathbf{U}^{-1} \langle \tilde{\epsilon}, \tilde{\epsilon}^T \rangle \mathbf{U}^{-T}, \quad \text{or} \quad \mathbf{B}^{-1} = \mathbf{U}^T \mathbf{U}. \quad (24)$$

This leads to a new representation of the incremental cost function of the form

$$\mathcal{J}[\mathcal{X}] = \frac{1}{2} \mathcal{X}^T \mathcal{X} + \frac{1}{2} [\mathbf{H}\mathbf{U}^{-1}\mathcal{X} - \mathbf{y} + \mathcal{H}(\mathbf{x}_b)]^T \mathbf{E}^{-1} [\mathbf{H}\mathbf{U}^{-1}\mathcal{X} - \mathbf{y} + \mathcal{H}(\mathbf{x}_b)], \quad (25)$$

where $\mathcal{X} = \mathbf{U}\delta\mathbf{x}$. With this cost function (25), no inversion of \mathbf{B} is needed.

4.1.1 Observation operator for radar radial winds

The radar wind *observation operator* \mathbf{H} produces the model counterpart of observed quantity that is presented to the variational assimilation. In the case of a *horizontal* wind observation from VAD profiles (see Fig. 2), the *observation operator* consists of a simple interpolation of the model wind field to the location of the observation. In the case of a *radar radial wind* super-observation, which is not a model variable, the *observation operator* involves: (i) a bilinear interpolation of the NWP model horizontal wind components u and v to the observation location; (ii) a projection of the interpolated NWP model horizontal wind, at the point of measurement, towards the radar beam using the formula

$$v_h = u \cos \theta + v \sin \theta, \quad (26)$$

where θ is the *azimuth* angle of the radar beam; and (iii) the v_h is finally projected in the slanted direction of the radar beam as

$$v_r = v_h \cos(\phi + \alpha), \quad \text{where} \quad \phi = \arctan \left(\frac{r \cos \alpha}{r \sin \alpha + d + h} \right), \quad (27)$$

where α is the *elevation* angle of the radar beam (see [19]). The formula for ϕ takes approximately into account the curvature of the Earth. In the term ϕ , r is the range, d is the radius of the Earth and h is the height of the radar above the sea level.

Some assumptions are, however, built into the standard formulation of the observation operator (27). First, the radar beam broadening is not taken into

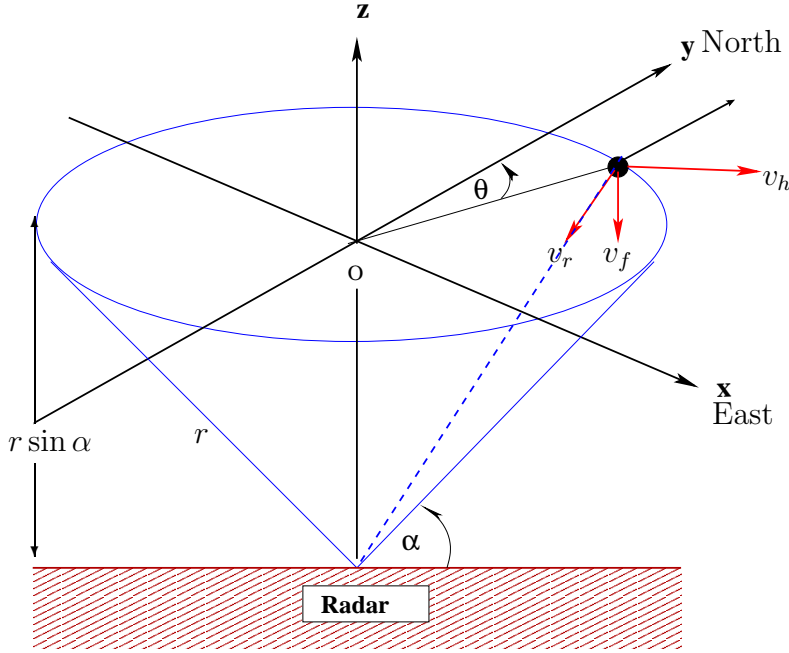


Fig. 2. Geometry for scan of velocities on a VAD circle

account. Second, the bending of the radar beam due to the hydrolapse in the boundary layer is not properly taken into account. Third, it is assumed that there is no mean velocity towards the radar due to the vertical motion of the precipitation, resulting in validity of measurements only for low elevation angles. This implicit assumption is embedded into (26) where only the NWP model horizontal wind is included.

One possible solution to relax the first assumption is to introduce a weighted average, using a Gaussian beam pattern, for the vertical interpolation of model horizontal wind components u , v of (26) to the observation location (see [27]). Then to model the broadening of the radar beam in the observation operator, one can use the Gaussian weight function

$$w = \frac{1}{2\pi} \exp\left(-\frac{(z - z_0)^2}{k}\right) \quad (28)$$

in the vertical instead of linear interpolation when defining the model horizontal components u and v to the observation height. Where, in formula (28), z is the model level height and z_0 is the observation height. The k -term defines the width of the filter response function.

4.2 4D-Var for Doppler radial winds and reflectivities

The 4-dimensional Variational Doppler Radar Analysis System (VDRAS) has also been developed, at the National Center for Atmospheric Research (NCAR), to assimilate *radial winds* and *reflectivities*, from single or multiple Doppler radars, by adding a penalty term \mathcal{J}_p to (12); see [32,31]. A cloud scale non-hydrostatic numerical model was used to represent the evolution of the motion

in the atmosphere. The data were interpolated from the original spherical polar geometry to the 3D Cartesian model grid. It was assumed that the *Doppler radar observation error correlations can be neglected*, i.e., the observational error covariance is diagonal. The cost function (12) may be reformulated as:

$$\begin{aligned} \mathcal{J}[\mathbf{x}(t_0)] = & [\mathbf{x}_b(t_0) - \mathbf{x}(t_0)]^T \mathbf{B}^{-1} [\mathbf{x}_b(t_0) - \mathbf{x}(t_0)] \\ & + \sum_{\Omega, T} \left[\eta_v [\mathcal{F}(v_r) - v_r^0]^2 + \eta_Z [\mathcal{F}(Z) - Z^0]^2 \right] + \mathcal{J}_p. \end{aligned} \quad (29)$$

The quantities v_r^0 and Z^0 are the observed radial velocities and reflectivities, respectively, while v_r and Z are their model counterparts. \mathcal{F} is the observation operator¹⁰ that projects the model variables (i.e, u , v , and w) from their Cartesian model grids to the observation variables (i.e, radial velocity v_r) in the data grids (see Eq. (31)). The coefficients η_v and η_Z give a measure of the weight¹¹ given to radial velocity and reflectivity observations, respectively. The summation is over the spatial domain Ω and the temporal domain T . Finally, the additional term \mathcal{J}_p is a penalty term that enforces the spatial and temporal smoothness of the analysis. Further discussion on the determination of \mathcal{J}_p and the coefficients in (29) is given in [29,30].

The two variables, radial velocity v_r and reflectivity Z are not direct model prognostic variables but can be computed using the model outputs of Cartesian velocity and *rainwater variable* q_r . The relation between Z and q_r , by assuming the Marshal-Palmer distribution of raindrop size, is given in the relation (see [32])

$$Z = 43.1 + 17.5 \log(\rho q_r). \quad (30)$$

While the relation between the model radial velocity v_r and the model Cartesian velocity components (u, v, w), is given by the formula:

$$v_r = \frac{x - x_{rad}}{r'} u + \frac{y - y_{rad}}{r'} v + \frac{z - z_{rad}}{r'} (w - V_{Tm}). \quad (31)$$

Here r' is the distance between a grid point (x, y, z) and the location of the radar wind observation $(x_{rad}, y_{rad}, z_{rad})$. V_{Tm} is the terminal velocity of the rain, that is estimated from the reflectivity data throughout the relation (see [32])

$$V_{Tm} = 5.4 (p_0/\bar{p})^{0.4} (\rho q_r)^{0.125}, \quad (32)$$

¹⁰ In the case of direct assimilation of raw data, the observation operator relates the model variables to observation variables and a transformation between different grid meshes.

¹¹ The constants η_v , η_Z reflect the relative precision of the radial velocity v_r and the reflectivity Z observation, respectively.

where p_0, \bar{p} are the base-state pressure and the pressure at the ground, respectively. The quantity ρ is the density of the air, and q_r is a rainwater variable.

We should mention the fact that, due to the poor vertical resolution of the radar data, a vertical interpolation of the data from the constant elevation levels to model Cartesian levels can result in large errors. For this reason a direct assimilation of the Plan Position Indicator (PPI)¹² data with no vertical interpolation was recommended. This may be an optimal interpolation within the context of the 4D-Var formulation (see [29]). In this case the observation operation \mathcal{F} is formulated to map the data from the model vertical levels to the elevation angle levels via the formula

$$v_{r,e} = \mathcal{F}(v_r) = \frac{\sum e^{-z^2/2\beta^2} v_r \Delta z}{\sum e^{-z/2\beta^2} \Delta z}, \quad (33)$$

where $v_{r,e}$ is the radial velocity on an elevation angle level and Δz is the model vertical grid spacing. β is the beam half-width and z is the distance from the center of radar beam. The summation is over the model grid points that lie in a radar beam.

Another major challenge in radar data analysis is to provide an analysis with a smooth transition between regions with and without radar data, or between data-dense and data-sparse regions. A good analysis should be able to fit to the observations while maintaining smoothness in space and time. Both *background term* and the *penalty term* in the cost function may help to smooth the analysis.

5 Numerical Implementation Model

NWP can be summarized in the following three steps: The first is to collect all atmospheric observations for a given time. Second, those observations are diagnosed and analyzed (that represented in data assimilation) to produce a regular, coherent spatial representation of the atmosphere at that time. This analysis becomes the initial condition for time integration of NWP model based on the governing differential equations of the atmosphere. These equations are in general partial differential equations of which the most important are equations of motion, the first law of thermodynamics, and the mass and humidity conservation equations. Finally, these equations are solved numerically to predict the future states of the atmosphere.

¹² PPI product takes all data collected during a 360 degree azimuth scan using the same elevation angle and projects it down onto a plane.

5.1 Basic equations

The primary set of equations that governs the evolution of the atmosphere (see [7,15] for details) is given by

$$\frac{d\mathbf{v}}{dt} = \alpha \nabla p - \nabla \phi + \mathbf{F} - 2\mathbf{v}_g \times \mathbf{v}, \quad (34a)$$

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}), \quad (34b)$$

$$p\alpha = RT, \quad (34c)$$

$$Q = C_p \frac{dT}{dt} - \alpha \frac{dp}{dt}, \quad (34d)$$

$$\frac{\partial \rho q}{\partial t} = -\nabla \cdot (\rho \mathbf{v} q) + \rho(E - C). \quad (34e)$$

These equations have seven unknowns: velocity $\mathbf{v}=(u, v, w)$, temperature T , pressure p , density $\rho = \frac{1}{\alpha}$, and evaporation ratio q . The first equation represents the *conservation of momentum, or Newton's second law*, where F , ϕ , and \mathbf{v}_g are, respectively, the frictional force, the geopotential¹³, and the angular velocity of the parcel of the air. The second is the *continuity equation or equation of conservation mass*. The third is the equation of state of perfect gases, where R is the gas constant for air. The fourth expresses the *thermodynamic energy equation* applied to a parcel of air with heat rate Q per unit mass at constant pressure C_p . The potential temperature θ is defined by $\theta = T(p_0/p)^{R/C_p}$, where p_0 is a reference pressure (say, 1000 hPa). Whereas the fifth equation (34e) represents the *equation of conservation of water vapour mixing ratio* q , where E and C are the evaporated and condensed amounts of water, respectively.

However, in order to assimilate a time series of radar observations (radial and reflectivity) from single or multi-Doppler radars, a cloud-scale NWP model is used. Sun and Crook [33] reduced the above system into four prognostic equations: three for the velocity components (u, v, w) , and the fourth for the potential temperature θ . An equation governing the evolution of reflectivity was also included to assimilate the reflectivity data. The system takes the form

$$\frac{d\mathbf{v}}{dt} = \delta_{i3} \frac{g\theta}{\Theta_0} \mathbf{k} - \nabla \tilde{p} + \nu \nabla^2 \mathbf{v}, \quad (35a)$$

¹³The *geopotential* of a unit mass relative to the *sea level*, numerically is the *work* that would be done in lifting the unit mass from sea level to the height at which the mass is located.

$$\frac{\partial \theta}{\partial t} = -w \frac{\partial \Theta}{\partial z} + k \nabla^2 \theta, \quad (35b)$$

$$\frac{\partial Z}{\partial t} = \mu \nabla^2 Z, \quad (35c)$$

where \tilde{p} is the perturbation pressure divided by a reference density, Θ is the horizontal mean potential temperature, Θ_0 is the reference potential temperature, and Z is the reflectivity. The quantity ν is the eddy viscosity, k is the thermal diffusivity, and μ is the diffusivity of reflectivity, which are assumed to be constant.

The mass continuity equation is written as

$$\nabla \cdot \mathbf{v} = 0, \quad (35d)$$

and the perturbation pressure \tilde{p} is diagnosed through the Poisson equation

$$\nabla^2 \tilde{p} = -\nabla \cdot (\mathbf{v} \cdot \nabla \mathbf{v}) + \frac{g}{\Theta_0} \frac{\partial \theta}{\partial z}. \quad (35e)$$

5.2 Numerical algorithm

Suppose that the governing atmospheric equations (35)–(35e) are expressed in the vector form

$$\begin{aligned} \frac{\partial \mathbf{x}(t)}{\partial t} &= F(\mathbf{x}(t)), \\ \mathbf{x}(t_0) &= \mathbf{x}_0. \end{aligned} \quad (36)$$

The vector \mathbf{x} contains all the prognostic variables.

The following algorithm outlines the steps of estimating (the first order approximation of) the cost function, and its gradient.

- (1) Discretize Eq. (36), using a suitable numerical method such as a finite difference or spectral method. This gives a nonlinear model solution, that depends only on the initial conditions,

$$\mathbf{x}(t_n) = M[\mathbf{x}(t_0)], \quad (37)$$

where $\mathbf{x}(t_n)$ represents the model forecast at time t_n and $\mathbf{x}(t_0)$ is the initial condition vector of the prognostic variables, and M is the time integration of the numerical scheme from the initial condition to the time t_n .

- (2) Evaluate the first-order variation of $\mathbf{x}(t_n)$ throughout the formula

$$\delta\mathbf{x}(t_n) = \mathbf{L}(t_0, t_n)\delta\mathbf{x}(t_0), \quad (38)$$

where $\mathbf{L}(t_0, t_n) = \frac{\partial M}{\partial \mathbf{x}}$, is the tangent linear operator.

- (3) Find the adjoint model, by taking the transpose of the tangent linear operator, so that

$$\delta\mathbf{x}^*(t_0) = \mathbf{L}^T(t_n, t_0)\delta\mathbf{x}(t_n), \quad (39)$$

where $\delta\mathbf{x}^*(t_0)$ represents the adjoint variable, and $\mathbf{L}^T(t_n, t_0)$ is the operator of the adjoint model.

- (4) Evaluate the weighted observation increments

$$\bar{\mathbf{d}}_n = \mathbf{H}_n \mathbf{E}_n^{-1} [\mathcal{H}[\mathbf{x}(t_n)] - \mathbf{y}(t_n)] \equiv \mathbf{H}_n \mathbf{E}_n^{-1} \mathbf{d}_n, \quad (40)$$

where $\mathbf{H}_n = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t_n)}$, and \mathcal{H} is the observation operator that relates the model variables to the observations.

- (5) Estimate the gradient of the cost function \mathcal{J} , (14) & (21), throughout:
- (i) Initialize $gradient = 0$.
 - (ii) for $k = N, 0, -1$ do
 - $gradient = L_{k-1}^T [\bar{\mathbf{d}}_k + gradient]$ ($L_{-1}^T = 1$).
 - (iii) $gradient = \mathbf{B}^{-1} \delta\mathbf{x} + gradient$.
- (6) The control variables are then adjusted (by the minimization routine, applying quasi-Newton iterative methods subject to constraints (36)) so as to reduce the value of the cost function;
- 7- When no further reduction in the value cost function $\mathcal{J}[\mathbf{x}(t_0)]$ is possible, *the best fit values* of the initial conditions $\mathbf{x}(t_0)$ have been found. The governing equations of the atmosphere (36) are then forward integrated in time to predict future states from the present state.
- 8- The process is repeated for the new cycle.

Due to the linearized approximation, an exact adjoint is generated and the cost function is quadratic and the minimization process speeds up.

6 Why are Weather Forecasts Sometimes Wrong?

In general, forecast skill increases not only by increasing model resolution, but also by improving the numerical models and the method of solution. However, even if we had a forecast model that represented atmosphere processes perfectly, we would never be able to predict the state of atmosphere accurately for *long lead times*¹⁴. This occurs, because the *non-linear dynamical systems*

¹⁴ A chaotic behaviour occurs (and leads to an unpredictable long-term evolution) when solving deterministic, nonlinear, dynamical systems that exhibit sensitivity to

that describe the atmospheric behaviour are *sensitive* to small changes in initial conditions.

6.1 Sensitivity to initial conditions

Lorenz [22] has shown the sensitivity dependence on the initial measured state of the weather when he obtained two solutions of forecast models which were integrated with slightly different initial conditions. He demonstrated visually that there was structure in his chaotic weather model, and, when plotted in three dimensions, the atmospheric motion fell onto a butterfly-shaped set of points; see Fig. 3.

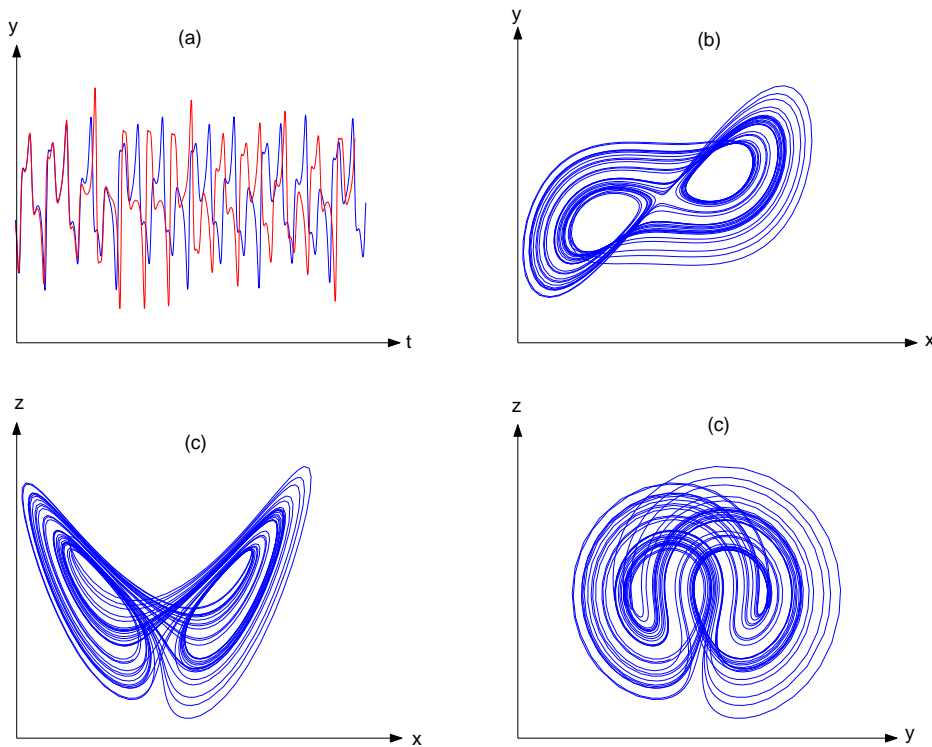


Fig. 3. Shows the trajectory of a chaotic motion, for the Lorenz system [22] that is sensitive to its initial values: $x' = \sigma(y - x)$; $y' = -xz + rx - y$; $z' = xy - bz$ (where $\sigma = 10$, $b = 8/3$, and $r = 28$). (a) plots two solutions of the component y with slightly change in the initial value. (b) & (c) show also how a system in chaotic motion is completely unpredictable.

To what lead time forecasts remain skillful depends on how small errors in the initial conditions, boundary conditions, or model specifications grow to affect the state output or the forecast. Because errors tend to grow rapidly in processes that occur at smaller spatial-scales, then forecasts for small scale processes may be predictable only for few hours. However, forecasts of large

 initial conditions.

scale processes can be predicted for perhaps 2 week ahead. Thus, when solving the forward problem, it is very important to assess the sensitivity of the state output variables of the dynamic system with respect to small changes in the initial conditions. A knowledge of how the *state variables* can vary with respect to small changes in the *initial data* can yield insights into the behaviour of the model and assist the modelling process to determine (for example) the most sensitive area. The sensitivity analysis, of the dynamic system, entails finding the partial derivative of the state variable (or the analysis) with respect to the parameters, which is a big challenge in a large nonlinear systems; see [3,4].

It should also be noted that the predictability of the atmospheric state depends mainly on the accuracy of the *parameter estimates* (the control variables), when solving the inverse problem. Since the ultimate goal is to produce an analysis that gives the best forecast, it is desirable to have information about the effect on the *analysis system* (or the estimates) due to perturbing the observations (or noisy data), or small changes in the *background*.

We give here a simple *linear* case of sensitivity in the *analysis* due to small changes in the *observations* and the *background*. Assume that the analysis (8) is expressed in this form

$$\mathbf{x}_a = \mathbf{W}\mathbf{y} + [\mathbf{I} - \mathbf{W}\mathbf{H}]\mathbf{x}_b, \quad (41)$$

where \mathbf{H} is the jacobian matrix to the linearized forward operator $\mathcal{H}[\mathbf{x}]$. If the analysis is projected at the observation locations so that $\hat{\mathbf{x}}_a = \mathbf{H}\mathbf{x}_a$, then

$$\hat{\mathbf{x}}_a = \mathbf{H}\mathbf{W}\mathbf{y} + [\mathbf{H} - \mathbf{H}\mathbf{W}]\mathbf{H}\mathbf{x}_b. \quad (42)$$

Here $\mathbf{H}\mathbf{x}_b = \hat{\mathbf{x}}_b$ is the projected background at the observation locations. Then $\hat{\mathbf{x}}_a$ is a weighted mean of \mathbf{y} and $\hat{\mathbf{x}}_b$.

Thus, the sensitivity of the analysis $\hat{\mathbf{x}}_a$ to observations is

$$\mathbf{S} \equiv \frac{\partial \hat{\mathbf{x}}_a}{\partial \mathbf{y}} = \mathbf{W}^T \mathbf{H}^T. \quad (43)$$

While sensitivity of the analysis to the background in the observation space is

$$\frac{\partial \hat{\mathbf{x}}_a}{\partial \hat{\mathbf{x}}_b} = \mathbf{I} - \mathbf{W}^T \mathbf{H}^T. \quad (44)$$

Sensitivity functions (43) & (44) are used to estimate the global and partial influences due to small changes in the observations and the background, in the observation space. Relative sensitivity functions ($0 \leq \text{tr}(\mathbf{S})/m \leq 1$, where m is the number of observations) give an insight to the modelers to determine

the most informative and sensitive area, where the assimilated data is dense or sparse; short- or long-range.

The *adjoint method*, is the most common method calculating the *sensitivity functions*, and has been used to calculate the *sensitivity forecast errors* $\mathcal{J}[\mathbf{x}_0]$ that depend on the dynamic system (36) and the *initial conditions* (see [9,24]). By defining, a change $\delta\mathbf{x}_0$ in the initial conditions \mathbf{x}_0 , this will lead to a change in the forecast error \mathcal{J} given by

$$\delta\mathcal{J} = \langle \nabla\mathcal{J}_o, \delta\mathbf{x}_0 \rangle, \quad (45)$$

where $\langle ., . \rangle$ is defined as an inner product, and $\nabla\mathcal{J}_o$ is defined in terms of the adjoint in the formula (20). Then one can say that, in regions where the gradient $\nabla\mathcal{J}_o$ is large, a change in the initial conditions would have created a large impact on the forecast error. Similarly, in regions where the gradient is small, such a change in the initial conditions would have affected the subsequent forecast error very little.

Adjoint equations can also be used to formulate systematically a formula for sensitivity of the state variable $\mathbf{x}(t)$ of the dynamic model (36) to small variations in the initial data (see [26]):

Theorem 1 *If $\mathbf{W}(t)$ is an n -dimensional adjoint function which satisfies the differential equation*

$$\begin{aligned} \frac{d\mathbf{W}(t)}{dt} &= -\frac{\partial F^T}{\partial \mathbf{x}(t)} \mathbf{W}(t), \quad t \leq t^*, \\ \mathbf{W}(t) &= 0, \quad t > t^*; \quad \mathbf{W}(t^*) = [0, \dots, 0, 1_{ith}, 0 \dots, 0]^T, \end{aligned} \quad (46)$$

then the sensitivity coefficients $\frac{\partial \mathbf{x}(t)}{\partial \mathbf{x}_0}$ for the dynamic system (36) can be expressed by the formula

$$\frac{\partial x_i(t^*)}{\partial \mathbf{x}_0} = \mathbf{W}(0). \quad (47)$$

PROOF:

For simplicity in Eq. (36), we write $F(\mathbf{x}(t)) = F$.

Small variations in the initial data (the system parameters) causes a perturbation in the system state in (36). Then small variations $\delta\mathbf{x}_0$, result in a variation $\delta\mathbf{x}(t)$ which satisfies (for first-order) the equation

$$\begin{aligned} \delta\mathbf{x}'(t) &= \frac{\partial F}{\partial \mathbf{x}} \delta\mathbf{x}(t) \\ \delta\mathbf{x}(0) &= \delta\mathbf{x}_0. \end{aligned} \quad (48)$$

If we multiply both sides of (48) by $\mathbf{W}^T(t)$ (the transpose of the function $\mathbf{W}(t)$) and integrate both sides with respect to t over the interval $[0, t^*]$, we obtain

$$\mathbf{W}^T(t^*)\delta\mathbf{x}(t^*) - \mathbf{W}^T(0)\delta\mathbf{x}(0) - \int_0^{t^*} \mathbf{W}'^T(t)\delta\mathbf{x}(t)dt = \int_0^{t^*} \mathbf{W}^T(t)\frac{\partial F}{\partial\mathbf{x}(t)}\delta\mathbf{x}(t)dt. \quad (49)$$

Eq. (49), after some manipulation, can be rewritten in the form

$$\mathbf{W}^T(t^*)\delta\mathbf{x}(t^*) - \mathbf{W}^T(0)\delta\mathbf{x}(0) = \int_0^{t^*} \left[\mathbf{W}'(t) + \frac{\partial F^T}{\partial\mathbf{x}}\mathbf{W}(t) \right]^T \delta\mathbf{x}(t)dt, \quad t \leq t^*. \quad (50)$$

Under the assumptions given in (46) the above equation takes the form

$$\delta x_i(t^*) = \mathbf{W}^T(0)\delta\mathbf{x}(0) \quad t \leq t^*. \quad (51)$$

When $\delta\mathbf{x}(0) \rightarrow 0$, we obtain the sensitivity coefficients (47) and the Theorem is proved. \square

6.2 Biases due to nonlinearity

In this subsection we only address the *bias*¹⁵ that comes due to *nonlinearity* of the model rather than *biases* that come due to *data- and background-errors*. In general, when the model state variable $\mathbf{x}(t, \mathbf{p})$ (here $\mathbf{p} \equiv \mathbf{x}_0$) is linear in all its parameters, then the cost function $\mathcal{J}[\mathbf{x}(t, \mathbf{p})]$ has a unique and global minimum, and finding it is usually a straightforward task. When the predictions are governed by models that are non-linear in the parameter estimates, then the *least squares* approach usually leads to a nonlinear minimization problem. Numerical algorithms for the nonlinear *least squares* approach are generally iterative procedures for searching the parameter estimates and require initial starting values. An obvious difficulty is that there may exist several local minima, and finding the global minimum is not guaranteed (see Fig. 4). To decrease the effect of nonlinearity, the choice of cost function should be made with certain practical issues in mind. For more details about the nonlinearity effects in parameter estimations, we may refer to [2,25].

Thus we may conclude that our estimate of the initial state will not be exactly the same as the true state. This *bias* depends on the degree of the nonlinearity of the structural model. This also adds another challenge in numerical weather prediction.

¹⁵ Bias is defined by the difference between the analysis state and the true state.

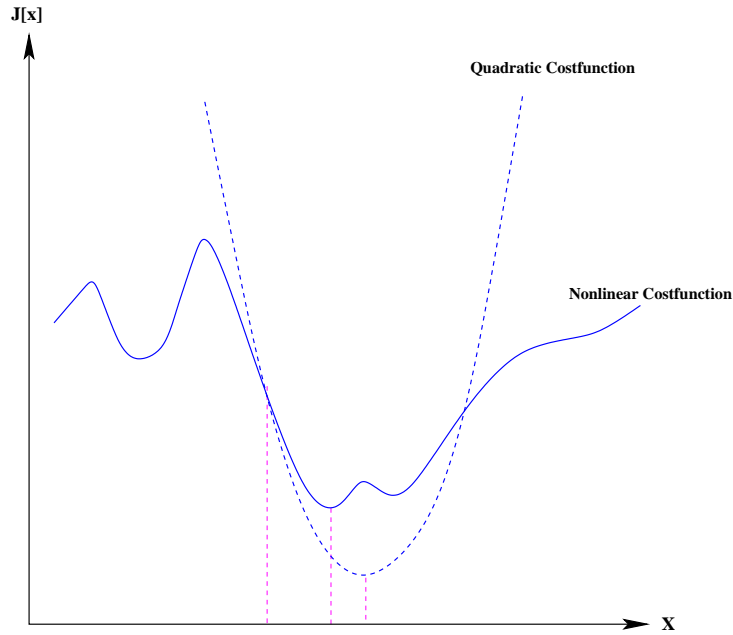


Fig. 4. Iterative solution for quadratic and nonlinear cost function.

7 Conclusion

The aim of this report was to investigate the progress of data assimilation, using 3D-Var and 4D-Var approaches, for numerical weather prediction. The essential difference between 3D-Var and 4D-Var is that 4D-Var includes the dynamic evolution of the model in the assimilation window, while 3D-Var assimilates at a particular point within the window. One limitation of using 3D-Var is that it can only be used with a short assimilation window, as the assimilation occurs at a specific point. However, 4D-Var has a considerable extra computational cost, compared to 3D-Var. We may reduce this cost by using the incremental method with low resolution. Numerical aspects concerning the estimation of the cost function and its gradient have been discussed. Some related problems, associated with data assimilation, such as nonlinearity and sensitivity of the forecast to possible small errors in initial conditions, random observation errors, and the background states have also been discussed.

In this study, a special emphasis has been given to assimilating Doppler radar winds into a NWP model. Variational methods were reviewed for their applicability to the data of Doppler radar winds. It has been shown that the variational methods offer a flexible methodology of using Doppler radar wind data, as well as the use of various constrains through the definition of the cost function. In addition, these methods combine interpolation and analysis into a single step. The analysis is performed more naturally and directly in a Cartesian coordinate system, and only interpolation from regular Cartesian grids to irregular radar observation points is needed. It has also been shown that using radar data directly at observation locations avoids an interpolation from an irregular radar coordinate system to a regular Cartesian system, which is

often a source of error, especially in the presence of data voids.

A particular challenge in the forecasting of the time evolution of atmospheric system is the nonlinearity of the system and the corresponding sensitivity of the initial conditions.

Acknowledgment

This project is sponsored by the National Environmental Research Council (NERC) through the Universities Weather Research Network (UWERN), a part of the NERC Centers for Atmospheric Science (NCAS).

References

- [1] P.P. Alberoni, Pre-processing of doppler radar data. Prepared under the European COST-717 action and available from <http://pub.smhi.se/cost717/> (2000).
- [2] D.M. Bates and D.G. Watts, *Nonlinear Regression Analysis and its Applications* (New York: John Wiley & Sons, 1988).
- [3] D.G. Cacuci, Sensitivity theory for nonlinear systems. I: Nonlinear functional analysis approach, *J.Math Phys.*, **22** (1981a) 2794–2802.
- [4] D.G. Cacuci, Sensitivity theory for nonlinear systems. II: Extensions to additional classes of response, *J.Math Phys.*, **22** (1981a) 2803–2812.
- [5] R. Daley, *Atmospheric Data Analysis*, (Cambridge University Press, 1996).
- [6] R.J. Doviak and D.S. Zrnic, *Doppler Radar and Weather Observations*. 2nd edition, (Academic Press, 1993).
- [7] G.J. Haltiner and R.T. Williams, *Numerical prediction and dynamic meteorology*, (Johon Wiley and Sons, New York, 1980).
- [8] J. Gao, A. Shapiro, and K. Dreogemeier, A variational method for analysis of three-dimensional wind fields from two Doppler radars, *Mon. Wea. Rev.*, **127** (1999) 2128–2142.
- [9] R. Gelaro, R. Buizza, T.N. Palmer, and E. Klinker, Sensitivity analysis of forecast errors and consture of perturbations using singular vectors, *J. Atms. Sc.*, **55** (1998) 1012–1037.
- [10] R. Giering, T. Kaminski, Recipes for adjoint code construction, *ACM Transactions on Mathematical Software* **24** (1998) 437-474.
- [11] J.C. Gilbert and C. Lemaréchal, Some numerical experiments with variable storage quasi-Newton algorithms, *Math. Prog.* **B25** (1989) 407–435.

- [12] N. Gustafsson, L. Berre, S. Hrnquist, X.-Y. Huang, M. Lindskog, B. Navascus, K. S. Mogensen and S. Thorsteinsson, Three-dimensional variational data assimilation for a limited area model. Part I: General formulation and the background error constraint, *Tellus A*, **53** (2001) 425-446.
- [13] P. Courtier, Variational methods. In: Ghil et al. (Eds.), *Data Assimilation in Meteorology and Oceanography: Theory and Practice*. Special issue, *J. Met. Soc. Japan*, **75**. Distributed by Universal Academic Press, (1997) 101-108.
- [14] K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc, Unified notation for data assimilation: Operational, sequential and variational, *J. Met. Soc. Japan*, **75** (1997) 181-189.
- [15] I.N. James, *Introduction to circulating atmospheres*, (Cambridge University press, Cambridge, 1994).
- [16] E. Kalnay, *Atmospheric modeling, data assimilation and prectipility*, (Cambrige Press, 2003).
- [17] F. Le Dimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoreical aspects, *Tellus* **38A** (1986) 97-110.
- [18] M. Lindskog, M. Gustafsson, B. Navascus, K. S. Mogensen, X.-Y. Huang, X. Yang, U. Andrae, L. Berre, Thorsteinsson and J. Rantakokko, Three-dimensional variational data assimilation for a limited area model. Part II: Observation handling and assimilation experiments, *Tellus A*, **53** (2001) 447-468.
- [19] M. Lindskog, H. Jrvinen, D. B. Michelson, Development of Doppler radar wind data assimilation for the HIRLAM 3D-Var, *COST-717 working document at: <http://www.smhi.se/cost717/>* (2002).
- [20] A.C. Lorenc, Development of an operational variational scheme, *J. Meteorol. Soc. Japan* **75** (1997) 339-346.
- [21] A.C. Lorenc, Analysis methods for numerical weather prediction, *Q.J.R. Meteorol. Soc* **112** (1986) 1177-1194.
- [22] E.N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.*, **20** (1963) 130-141.
- [23] D.F. Parrish and R. J. Purser, Anisotropic covariances in 3dvar: Application to hurricane Doppler radar observations, HIRLAM Workshop report. (HIRLAM 4 Workshop on Variational Analysis in Limited Area Models, Meteo-France, Toulouse, 23-25 February, (1998).
- [24] F. Rabier, E. Klinker, P. Courtier and A. Hollingsworth, Sensitivity of forecast errors to initial conditions, *J. R. Meteorol. soc.*, **122** (1996) 121-150.
- [25] D.A. Ratkowsky, *Nonlinear Regression Modeling, A Unified Practical Approach* (New York: Marcel Dekker 1983).
- [26] F.A. Rihan, Sensitivity analysis of dynamic systems with time lags, *J. Comput. Appl. Math.* **151** (2003) 445-462.

- [27] K. Salonen, Observation operator for Doppler radar radial winds in HIRLAM 3D-Var, *Proceedings of ERAD* (2002) 405-408.
- [28] R. Sneider, The role of nonlinearity in inverse problems, *Inverse Problems*, **14** (1998) 387-404.
- [29] J. Sun and N. A. Crook, Real-time low-level wind and temperature analysis using single WSR-88D data, *Wea. Forecasting*, **16** (2001a) 117-132.
- [30] J. Sun and N. A. Crook, Assimilation and forecasting experiments on supercell storms. Part I: experiments with simulated data. 14th Conference on Numerical Weather Prediction. Ft. Lauderdale, Florida, *Amer. Meteor. Soc.* (2001b) 142–146.
- [31] J. Sun, and N. A. Crook, Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part II: retrieval experiments of an observed Florida convective storm, *J. Atmos. Sci.* **55** (1998) 835–852.
- [32] J. Sun, and N. A. Crook, Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments, *J. Atmos. Sci.*, **54** (1997) 1642–1661.
- [33] J. Sun, and N. A. Crook, Wind and thermodynamic retrieval from single-Doppler measurements of a gust front observed during Phoenix II, *Mon. Wea. Rev.*, **122** (1994) 1075–1091.
- [34] O. Talagrand, Assimilation of observations. An introduction. In: Ghil et al. (Eds.), *Data Assimilation in Meteorology and Oceanography: Theory and Practice*. Special issue, *J. Met. Soc. Japan*, **75**. Distributed by Universal Academic Press, (1997) 81–99.
- [35] A. Tarantola, *Inverse Problem Theory*. (Elsevier Science, 1987).
- [36] Q. Xu and J. Gong, Background error covariance functions for Doppler radial wind analysis, *Q.J. R. Meteorol. Soc.*, **129** (2003) 1703–1720.

TR1_DataAssim.tex