A simple column model to explore anticipated problems in variational assimilation of satellite observations

A. C. Rudd[1], I. Roulstone[2], and J. R. Eyre[3]

[1]Present address: Agriculture Building, Earley Gate, University of Reading, Reading UK.
[2]Department of Mathematics, University of Surrey, Guildford, UK
[3]Met Office, Exeter, UK

Abstract

We investigate a simplified form of variational data assimilation in a fully nonlinear framework with the aim of extracting dynamical development information from a sequence of observations over time. Information on the vertical wind profile, $w(z)$, and profiles of temperature, $T(z,t)$, and total water content, $q_t(z,t)$, as functions of height, z, and time, t, are converted to brightness temperatures at a single horizontal location by defining a two-dimensional (vertical and time) variational assimilation testbed. The profiles of T and q_t are updated using a vertical advection scheme. A basic cloud scheme is used to obtain the fractional cloud amount and, when combined with the temperature field, this information is converted into a brightness temperature, using a simple radiative transfer scheme.

It is shown that our model exhibits realistic behaviour with regard to the prediction of cloud, but the effects of nonlinearity become non-negligible in the variational data assimilation algorithm. A careful analysis of the application of the data assimilation scheme to this nonlinear problem is presented, the salient difficulties are highlighted, and suggestions for further developments are discussed.

1 Introduction

Data assimilation is used to integrate data with models in a wide range of different fields, such as hydrology (Clark et al., 2006; Neal et al., 2009; Seo et al., 2009), crop science (Naud et al., 2007), oceanography (van Leeuwen, 2003), morphodynamic (Smith et al., 2011) and air quality modelling (van Velzen and Segers, 2010).

Satellite data have been available for over 40 years, from the launch of TIROS-1 in 1960, which was the first satellite able to provide images of the Earth, and hence weather (Eyre, 2007). However, cloud imagery is not currently used in numerical weather prediction (NWP) to extract the type of dynamical information that experienced forecasters have extracted subjectively for many years. For example, rapidly developing mid-latitude cyclones have characteristic signatures in the cloud imagery that are most fully appreciated from a sequence of images rather than from a single image.

There have been many improvements to the quality and frequency of the observations over this time. To date it has not been possible to make use of much of the information effectively because data assimilation methods have taken a rather 'static' view of the observations. There has been no regard for the temporal information contained in a sequence of satellite images; a set of observations around a given time has just been used to estimate the state of the atmosphere at that time. With the advent of 4D-Var (four-dimensional variational data assimilation (VAR)) this restriction is being lifted and we can think of extracting information from sequences of observations. 4D-Var brings with it the opportunity to use sequences of satellite images in an objective and quantitative manner.

It is generally accepted that by improving the representation of the initial cloud profile, a more realistic description of the three-dimensional structure of the diabatic heating produced by condensation, known to interact with the dynamical fields, such as velocity and pressure, will result (Errico et al., 2007). The potential for improving forecasts in cloudy regions is therefore greatest when the assimilation is continuous in time so that the temporal changes observed in the imagery can be used to make improvements to the forecast. Also, if dynamical information can be extracted from sequences of observed infra-red brightness temperature[1], it

[1]The brightness temperature is the apparent temperature of an object assuming that the object is radiating as a black body, i.e. that its brightness may be related to its temperature (Dunlop, 2001).

could help avoid significant error in the initial conditions and improve forecast accuracy. The study presented in this paper represents a "toy" 2D problem, to extract information from a sequence of observations over time. It explores the mathematical problems expected in a full 4D implementation when attempting to extract information from a satellite image sequence over time. However, it does not, in itself, address the extraction of information from images, as addressed, for example, by Titaud et al. (2010).

The main reason that the direct assimilation of satellite data is challenging is because satellite-borne instruments do not directly measure values of wind, temperature or humidity which would then be used to update the atmospheric model directly. Therefore nonlinear observation operators and/or forward models are needed to convert from model space to observation space (Simmons, 2000). This step can be highly nonlinear especially due to cloud generation being nonlinear and discontinuous.

The work presented here is one of the first attempts to assimilate observations of this nature in a 'simple' testbed. It is desirable to understand the problem within an idealised framework that retains the critical nonlinearities in order to gain insight into the assimilation within the full operational 4D-Var system. Our principal goal is to determine whether it is possible to recover a simple profile of vertical motion from observations of cloud-affected brightness temperatures, via a standard VAR scheme. Our experiment is a stringent test of VAR, because such schemes rely on nonlinear processes being well approximated by linear models, at least locally in space and time.

We define a 2D Var system (in height, z, and time, t) in an idealised setting to facilitate an investigation into the effect of the nonlinearities, which could not be easily done in the complex operational systems. This paper summarises our findings; for a more detailed description of the work the reader is referred to Rudd (2009).

Our simple model demonstrates that the standard tests raise some concerns about the validity of the tangent linear hypothesis of VAR (see §5) for highly nonlinear systems. This will become more important in future operational systems, as attempts to assimilate more complex and nonlinear processes are made.

This paper addresses the issue of assimilating sequences of satellite data using a simple column model. Section 2 describes the general principles of four-dimensional data assimilation, §3 defines the nonlinear model, and §4 describes basic tests. The tangent linear model tests are explained in §5. Section 6 outlines the variational assimilation testbed and §7 describes the adjoint testing. Section 8 contains some 2D-Var assimilation results and the conclusions are contained in §9.

2 Four dimensional data assimilation

There are many techniques for estimating an optimal analysis. These include variational techniques and optimal interpolation. Lorenc (1986) has shown how these and many other techniques are underpinned by Bayesian statistics.

The aim of data assimilation is to combine the background state $\mathbf{x}_b$ and the observations $\mathbf{y}$ to produce an analysis $\mathbf{x}_a$ which is optimally 'close' to the true state.

2.1 The 4D-Var algorithm

4D-Var is a non-sequential intermittent approach to data assimilation (Daley, 1991; Kalnay, 2003). It is the process by which a sequence of observations, distributed in time, are assimilated over a time interval called the *assimilation window* ($t_0 \rightarrow t_N$). A typical assimilation window in operational data assimilation for global NWP is 6 hours. 4D-Var incorporates observations which span not only three-dimensional space, but also a time domain.

The 4D-Var analysis state $\mathbf{x}_a$ is given by the initial state $\mathbf{x}_0$ which minimises a cost function subject to the constraint that $\mathbf{x}_i$ also satisfies the model equations

$$\mathbf{x}_{i+1} = \mathcal{M}(\mathbf{x}_i), \tag{1}$$

where $\mathbf{x}_i$ is the state vector at time t_i and $\mathcal{M}$ is the forward model used to integrate the state vector forward in time. This formulation is known as a strong constraint, which means that the model is assumed to be perfect.

In general, the 4D-Var cost function $J(\mathbf{x}_0)$ is a dimensionless scalar with two components;

$$J(\mathbf{x}_0) = J^b + J^o. \tag{2}$$

J^b is the background term and it measures the departure of the model state $\mathbf{x}_0$ at time t_0 (the control variable) from the background state $\mathbf{x}_b$ (at the initial time). The observation term, denoted J^o, measures the departure

of the observations $\mathbf{y}_i$ from the model-predicted observations $h[\mathbf{x}_i]$, where the *observation operator*, h, maps the control vector from model space to observation space.

The components of the cost function are defined by

$$J^b = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b), \tag{3}$$

and

$$J^o = \frac{1}{2}\sum_{i=0}^{N} (\mathbf{y}_i - h[\mathbf{x}_i])^T \mathbf{R}^{-1} (\mathbf{y}_i - h[\mathbf{x}_i]), \tag{4}$$

where $\mathbf{B}$ and $\mathbf{R}$ are the *background* and *observation error covariance matrices* respectively. Note that J^o involves a summation over time.

The non-uniform spatial distribution of the observations is also a problem that the data assimilation algorithm must address. It must be able to interpolate between the observations in a meteorologically realistic way, whilst filtering out any bad observations. The algorithm also needs to be able to solve large-dimensional problems (operationally there are about 10^7 unknowns) quickly in order to generate the next forecast in a timely manner.

The VAR problem is solved in an iterative fashion using a minimisation algorithm which uses information on the cost function J and the gradient with respect to the control vector, $\nabla_{\mathbf{x}_0} J$, to refine an estimate of $\mathbf{x}_0$ at each iteration. To compute J the forward model is integrated from time t_0 to t_N using the nonlinear forward model (1). To obtain the gradient we integrate backward from t_N to t_0 using an adjoint model (Lewis and Derber, 1985). The adjoint model is the adjoint of the tangent linear version of the nonlinear forward model; the version linearised about the forward trajectory $\mathbf{x}_0, \ldots, \mathbf{x}_N$ (Simmons, 2000).

As we mentioned earlier, because the satellite instruments do not measure directly the variables used in the models (w, T, q_t, p) an observation operator is required to convert from model variables in model space to an observation in observation space. The relationship between the satellite observation and the model variables is nonlinear and could lead to a multi-modal cost function, and the assimilation could find a local minimum rather than the global minimum and optimal state.

2.2 Error covariance matrices

The background error covariance matrix $\mathbf{B}$, is a very important part of the data assimilation algorithm. It is responsible for spreading out the data to surrounding grid points in a smooth and consistent way, to ensure a smooth analysis field. It is also used to specify correlations between variables and is defined by

$$\mathbf{B} = \overline{(\mathbf{x}_b - \mathbf{x}^t)(\mathbf{x}_b - \mathbf{x}^t)^T} = \overline{(\boldsymbol{\epsilon_b})(\boldsymbol{\epsilon_b})^T}, \tag{5}$$

where $\boldsymbol{\epsilon_b} = \mathbf{x}_b - \mathbf{x}^t$, i.e. the difference between the background state value and its 'true' state value (the state of the atmosphere if it was known exactly), and the overbar denotes an average. The diagonal elements of $\mathbf{B}$ are the error variances. Section 6.5 describes the $\mathbf{B}$ matrix for our 2D-Var system.

The observation error covariance matrix $\mathbf{R}$ is a statistical description of the random errors in $\mathbf{y}$. It is defined by

$$\mathbf{R} = \overline{(\mathbf{y} - h[\mathbf{x}^t])(\mathbf{y} - h[\mathbf{x}^t])^T} = \overline{(\boldsymbol{\epsilon_o})(\boldsymbol{\epsilon_o})^T}. \tag{6}$$

It is usually assumed diagonal, indicating that errors between each observation are *uncorrelated*. The diagonal elements of $\mathbf{R}$ are the error variances. Observation errors represented in the $\mathbf{R}$ matrix include instrument errors, representativeness errors from the discretisation, and errors in the observation operator. Section 6.6 describes the $\mathbf{R}$ matrix for our 2D-Var system.

3 Nonlinear model

The nonlinear forward model is a single column in the vertical for one grid-point location in the horizontal. We investigate the evolution of the properties of this column with respect to time, t, and the vertical coordinate, z.

Our *state vector* $\mathbf{x}$ is a column vector defined as

$$\mathbf{x}(z,t) = [w(z), T(z,t), q_t(z,t)]^T, \tag{7}$$

where w, T and q_t are vertical velocity, temperature and total water content respectively. The vertical velocity is only a function of height, whereas temperature and total water content are functions of both height and time. Vertical motion is prescribed to be constant in time because an initial goal is to investigate whether the 2D-Var system can recover a simple steady profile.

The advection scheme is outlined in §3.1 and cloud scheme in §3.5. The observation operator is the radiative transfer scheme (§3.6), which converts the state vector information on w, T and q_t into an upwelling brightness temperature observation for use in our 2D-Var assimilation system. Reference to the nonlinear model (NLM) means both the nonlinear forward model and the observation operator.

3.1 Advection

The model is based on two Lagrangian conservation laws for potential temperature[2], $\theta(z,t)$, and total water content, $q_t(z,t)$:

$$\frac{D\theta}{Dt} = 0, \qquad \frac{Dq_t}{Dt} = 0, \tag{8}$$

where $D/Dt = \partial/\partial t + w\partial/\partial z$ is the Lagrangian derivative, and vertical advection is prescribed by a function of height alone, $w(z)$. A semi-Lagrangian (SL) approach is used to integrate the model variables forward in time. The SL scheme is used for q_t and also T with an adiabatic adjustment (see (12)). This is easy to calculate and avoids the need to make repeated conversions between T and θ.

An SL approximation to the advection equation is written in the form;

$$\frac{\phi(z^j, t_{i+1}) - \phi(\tilde{z}_i^j, t_i)}{\Delta t} = 0, \tag{9}$$

or

$$\phi(z^j, t_{i+1}) = \phi(\tilde{z}_i^j, t_i), \tag{10}$$

where ϕ is a conserved quantity, the superscript j denotes the model level and the *departure point*, $\tilde{z}_i^j$, is the location at time t_i for the parcel that is located at z^j at time $t_i + \Delta t$. This position does not, in general, lie on a grid point, so that evaluation of the right hand side of (10) requires interpolation from grid point values at time t. For $w > 0$ the position $\tilde{z}_i^j$ lies between the grid points z^{j-p} and z^{j-p-1} where p is the integer part of the expression $w\Delta t/\Delta z$ (a measure of the number of model levels (Δz) traversed in a timestep (Δt)).

Approximating $\phi(\tilde{z}_i^j, t_i)$ by linear interpolation, (10) becomes

$$\phi_{i+1}^j = (1-\alpha)\phi_i^{j-p} + \alpha\phi_i^{j-p-1}, \tag{11}$$

where $\phi_i^j = \phi(z^j, t_i)$ and $\alpha = \frac{z^{j-p} - \tilde{z}_i^j}{\Delta z}$.

3.2 Updating the temperature profile

The temperature is updated using the temperature obtained from the SL method and an update, ΔT^j, which is the change in the temperature profile due to adiabatic heating/cooling:

$$\Delta T^j = -\Psi^j w^j \Delta t, \tag{12}$$

where w^j is the vertical velocity profile, which is constant in time, Δt is the time step of the model and Ψ^j is the rate of change of temperature of an ascending parcel.

3.3 Calculation of pressure

Pressure is needed for the calculation of the lapse rate (the rate at which the air temperature cools with height) required in the calculation of Ψ. Pressure is obtained assuming hydrostatic balance with the assumption that T varies in a linear fashion within an individual layer,

$$p_t = p_b \exp\left(-\frac{g}{R}\left(\frac{z_t - z_b}{T(z_t) - T(z_b)}\ ln\left[\frac{T(z_t)}{T(z_b)}\right]\right)\right), \tag{13}$$

[2]The potential temperature of a parcel of air is $T\left(\frac{p_0}{p}\right)^{\frac{R}{c_p}}$, where T is temperature (K), R is the specific gas constant for air (J kg^{-1} K^{-1}), c_p is the specific heat capacity at constant pressure (J Kg^{-1} K^{-1}), p is pressure (hPa) and p_0 is some reference pressure, usually taken to be 1000hPa. The potential temperature is therefore the temperature that the parcel would have if it was adiabatically brought to a standard reference pressure.

where p_t is the pressure at the top of a layer and p_b is the pressure at the bottom of a layer. R is the specific gas constant for dry air (287.05 J kg^{-1} K^{-1}), g is the acceleration due to gravity (9.81 ms^{-2}), z_t and z_b are the height values at the top and bottom of a layer respectively, and $T(z_t)$ and $T(z_b)$ are the temperatures at the top and bottom of the layer respectively. The pressure on each model level is then obtained by integrating from the surface to the top model level. The surface pressure is prescribed.

3.4 Thermodynamics

The lapse rate for dry air is given by

$$\frac{dT}{dz} = -\frac{g}{c_p} = -\Gamma_d, \tag{14}$$

where Γ_d is known as the *dry adiabatic lapse rate* (DALR) (Houghton, 2002, p. 5). For the Earth's atmosphere, $c_p = 1005$ J Kg^{-1} K^{-1} and $g = 9.81$ ms^{-2}. However, if the atmosphere is *saturated*, a parcel will ascend with lapse rate Γ_s, where Γ_s is the *saturated adiabatic lapse rate* (SALR). The SALR is calculated using the formula from Houghton (2002)

$$\Gamma_s = \Gamma_d \, \frac{\left(1 + \frac{L_v \; e_s \; M_{rv}}{p \; R \; T}\right)}{\left(1 + \frac{L_v \; e_s \; M_{rv}}{p \; R \; T} \frac{\epsilon \; L_v}{c_p \; T}\right)}, \tag{15}$$

where L_v is the latent heat of vapourisation (2.5×10^6 J kg^{-1} at 273 K), M_{rv} is the molecular weight of water vapour (18.015), R is the gas constant per mole (8.3143 kJ K^{-1} kmol^{-1}), ϵ is the ratio of molecular weights (0.622), c_p is the specific heat at constant pressure (1005 J kg^{-1} K^{-1}) and e_s is the saturation vapour pressure (hPa).

We compute the effective lapse rate Ψ for a mixed column containing both saturated and unsaturated air by weighting the DALR and SALR by the cloud fraction, f. The DALR applies to the unsaturated part of the "column" and the SALR to the saturated part, we use an appropriate average of the two. Ψ^j is computed from

$$\Psi_i^j = (1 - f_i^j) \; \Gamma_d + f_i^j \; \Gamma_{si}^j, \tag{16}$$

where i is a time level and j is a model level.

3.5 Calculation of cloud fraction

We define a simple cloud scheme (17) which is a continuous and differentiable approximation to the Smith (1990) scheme (used operationally by the UK Met Office). Equation (17) asymptotes to 0 and 1, therefore the gradient is never exactly zero, which is highly desirable for variational data assimilation.

The cloud scheme calculates a fractional cloud cover, f, from inputs of total water content, q_t, saturation specific humidity, $q_{sat}(T, p)$ and critical relative humidity, RH_{crit} (value at which cloud starts to form),

$$f = \frac{1}{2}\left[1 + tanh\left(\frac{2\,(q_t - q_{sat}(T,p))}{q_{sat}(T,p)(1 - RH_{crit})}\right)\right]. \tag{17}$$

The nonlinearity of the function can be adjusted using the RH_{crit} parameter, which varies the slope of the tanh curve. We choose (17) because it is smooth, continuous and similar to the functional fit of Wood and Field (2000) for aircraft measurements.

3.6 Observation operator

This section outlines the radiative transfer scheme, which calculates the upwelling brightness temperature, TB, using the cloud fraction, f, and temperature, T. The value of TB at time level i and model level $j = N$ (top level in z) is equivalent to what would be seen by a satellite, i.e. the observation.

$$TB_i^j = (1 - f_i^j) TB_i^{j-1} + f_i^j T_i^j \; ; \tag{18}$$

with the following boundary condition;

$$TB_i^1 = T_i^1. \tag{19}$$

Starting at the surface ($j = 1$) the contribution to the upwelling brightness temperature is calculated up to the top model level ($j = N$).

The formulation of our radiative transfer scheme incorporates certain assumptions, and consequently, simplifications. The first assumption is that we assume that the vertical overlap of clouds is random. This assumption gives rise to the simple form of (18). The second assumption is that the top model level is the top of the atmosphere (TOA) for radiative transfer purposes because we have no cloud above level N, i.e. there is no radiative contribution from above model level N and our atmosphere is transparent. We also assume that radiance and temperature are linearly related. This is not the case in the real world but makes for a good simplification for our study. An additional simplification is that only cloud emission/absorption are considered, no gaseous absorption/emission is included, i.e. a perfect window channel is assumed. This model is simplistic, but is sufficient for our needs within this simulation study. These assumptions would need to be re-examined when applied to real data.

Figure 1 illustrates the vertical structure of the nonlinear model. The vertical velocity is prescribed to be zero at the surface and top model levels. All the variables, except TB are defined at each model level, j. The way that the brightness temperature is calculated means that it is only the brightness temperature at the top model level, at the other model levels the values of TB^j are just contributions to the value of the brightness temperature observation. A contains an overview of the algorithm.

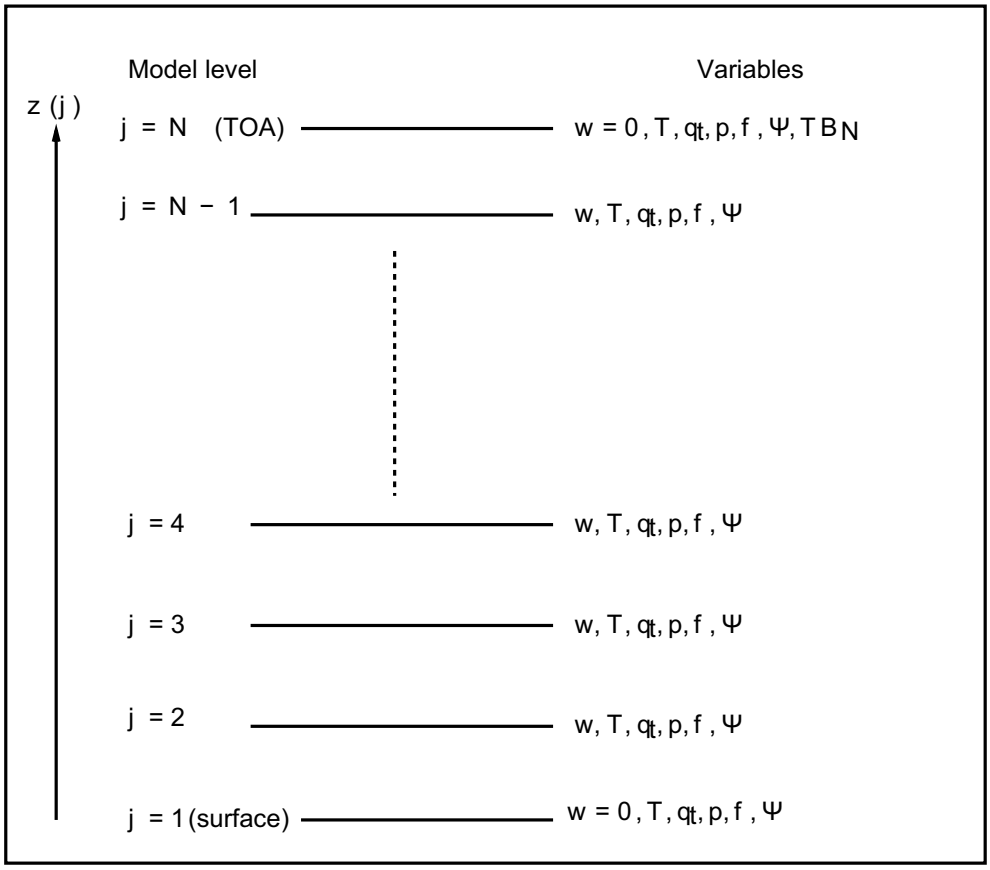


Figure 1: *A schematic of the vertical structure of the model.*

4 Basic testing

Changing the vertical resolution and time step

We investigated the effect of changing the vertical resolution (26, 51 and 101 model levels) and time step (10, 5, 2.5 minutes) of the model (not shown). Neither had a significant effect on the model solutions over a period of 6 hours (typical assimilation window).

We also studied the effect of perturbing the initial conditions by the size of typical background errors. For each experiment only one variable (T, q_t or w) was perturbed in order to study the effects in isolation.

Perturbing the initial vertical velocity profile

Table 1 shows the effect on the observations of brightness temperature from perturbing the initial vertical velocity profile. Increasing w by 50% has the effect of generating more cloud over time due to uplift and cooling, this gives rise to a cooler observation of brightness temperature compared to the observation from the unperturbed original initial profile. The opposite is true when we decrease w by 50%. Table 1 illustrates that the effect on the observations of brightness temperature is greater when the w profile is negatively perturbed.

Time	Unperturbed TB (K)	$w + 50\%$ TB (K)	Difference (K)
Initial	261.01	261.01	0.0
After 2 hours	230.59	227.54	3.05
After 6 hours	224.02	222.86	1.16
	Unperturbed TB (K)	$w - 50\%$ TB (K)	Difference (K)
Initial	261.01	261.01	0.0
After 2 hours	230.59	238.56	-7.97
After 6 hours	224.02	227.54	-3.52

Table 1: *Effect on the observations of brightness temperature from perturbations to the initial vertical velocity profile.*

Perturbing the initial temperature profile

When the temperature profile at the initial time is perturbed the profiles of p, Ψ, f, and the observation are changed, the w profile remains unchanged. The initial q_t profile is still derived from an initial (constant) RH_T profile of 65%.

Perturbing the temperature profile changes the value of the observation of brightness temperature because it is calculated directly from the cloud fraction and temperature profiles. Increasing the temperature by 1 K has the effect of decreasing the cloud fraction, and this gives rise to the warmer brightness temperature observations, as seen in Table 2. The opposite is true when the temperature profile is perturbed by minus 1 K.

Time	Unperturbed TB (K)	$T + 1$ K TB (K)	Difference (K)
Initial	261.01	262.01	-1.0
After 2 hours	230.56	231.67	-1.11
After 6 hours	224.02	225.05	-1.03
	Unperturbed TB (K)	$T - 1$ K TB (K)	Difference (K)
Initial	261.01	260.01	1.0
After 2 hours	230.56	229.81	0.75
After 6 hours	224.02	222.99	1.03

Table 2: *Effect on the observations of brightness temperature from perturbations to the initial temperature profile.*

Perturbing the initial total water content profile.

Perturbing the initial total water content profile by 15% over height has the effect of changing the initial cloud fraction profile, w and T remain unchanged.

Increasing the q_t profile by 15% over all model levels increases the initial cloud fraction profile and consequently decreases the value of the brightness temperature observation. The opposite is true when the total water content profile is perturbed by minus 15%.

We can see that there is a strong sensitivity to humidity and cloud at the beginning of the time window, but much smaller sensitivity at the end (Table 3).

Time	Unperturbed TB (K)	$q_t + 15\%$ TB (K)	Difference (K)
Initial	261.01	239.56	21.45
After 2 hours	230.56	226.60	3.96
After 6 hours	224.02	223.14	0.88
	Unperturbed TB (K)	$q_t - 15\%$ TB (K)	Difference (K)
Initial	261.01	280.74	-19.73
After 2 hours	230.56	237.27	-6.71
After 6 hours	224.02	225.07	-1.05

Table 3: *Effect on the observations of brightness temperature from perturbations to the initial total water content profile.*

The next section examines the effect of changing the critical relative humidity parameter, RH_{crit}, the relative humidity at which cloud starts to form. This parameter changes the slope of the tanh curve for calculating the cloud fraction f, i.e. one of the nonlinearities of the nonlinear forward model.

Changing the critical relative humidity

Figure 2 illustrates the effect of changing RH_{crit}, the value at which cloud will form. Increasing RH_{crit} steepens the slope of the total relative humidity

$$RH_T = \frac{q_t}{q_{sat}(T,p)} \tag{20}$$

versus cloud fraction (f) curve. Increasing RH_{crit} also decreases the initial cloud fraction profile. A typical value of RH_{crit} in NWP models is 85%.

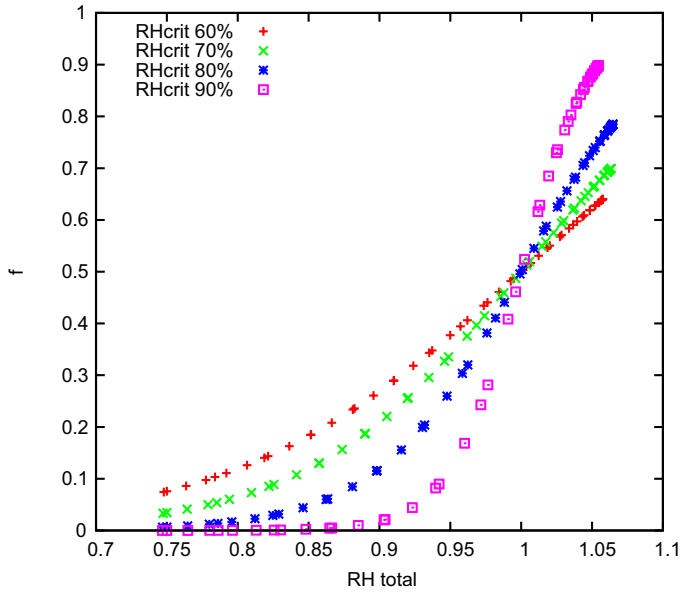


Figure 2: *The total relative humidity versus cloud fraction for critical relative humidity values of* 60% *to* 90%. *These are values after a* 2 *hour integration of the nonlinear forward model with* q_t *derived from an initial* RH_T *value of* 75%.

5 Tangent linear model (TLM)

A key requirement of many data assimilation systems is the linearisation of the nonlinear forward model and nonlinear observation operators. The usual method to construct the linear model is to first discretise the nonlinear equations, and then linearise the discrete numerical scheme in order to obtain a discrete linear model. This is known as the *discrete* method and the discrete linear model formed in this way is called the *tangent linear model* (TLM) (Lawless et al., 2003). We obtain the TLM by linearising each line of the nonlinear source code by hand.

The next sections outline two standard tests. The correctness test (§5.2), used to check that the TLM is coded correctly and the validity test (§5.3) used to examine the evolution of perturbations (in both the NLM and TLM) over time. Use is also made of the *relative error* (ratio between the norms of the nonlinear higher-order terms (HOTs) and the linear first-order term of the Taylor expansion) to determine the validity of the TLM (§5.4). The magnitude of the relative error is a good indicator of whether the tangent linear (TL) hypothesis is valid. Low values indicate that the evolution in the nonlinear model is reasonably linear and that the TL hypothesis is valid. However, high values indicate significant influences from nonlinear effects on the evolution and therefore validity of the TL hypothesis is questionable.

Even with our simple model we find that the standard tests raise some concerns about the validity of the TL hypothesis for more nonlinear systems. This will become more important in the future as attempts to assimilate more complex, and nonlinear, processes are made. For example, as the resolution of NWP models increases there will be a need to explicitly represent the moist physical processes in a less parameterised fashion.

5.1 Tangent linear hypothesis

The Taylor series expansion at first order of a nonlinear model, $\mathcal{M}$, around the state vector, $\mathbf{x}$ is

$$\mathcal{M}(\mathbf{x}+\delta\mathbf{x}) = \mathcal{M}(\mathbf{x}) + \mathcal{M}'(\mathbf{x})\ \delta\mathbf{x} + \frac{1}{2}\ \mathcal{M}''(\mathbf{x})\ \delta\mathbf{x}^2 + \text{HOTs}. \tag{21}$$

The linear model is the first order term of the Taylor series

$$\mathcal{M}(\mathbf{x}+\delta\mathbf{x}) - \mathcal{M}(\mathbf{x}) = \mathbf{M}(\mathbf{x})\ \delta\mathbf{x} + \text{HOTs}, \tag{22}$$

where $\mathbf{M}(\mathbf{x})$ is the linear model, which evolves the perturbation $\delta\mathbf{x}$.

The *tangent linear hypothesis* states that the nonlinear model should exhibit similar behaviour to the linear model for (sufficiently small) perturbations used in the VAR scheme. The TLM describes the linear evolution of perturbations along the solutions of the nonlinear model. A thorough discussion of the validity of TLMs in the context of data assimilation and ensemble forecasting is given by Park and Droegemeier (1997). In VAR it is important to calculate the gradient of the cost function as accurately as possible, especially if the components of ∇J behave very differently. The gradient with respect to the initial conditions is calculated using the adjoint technique, as described in Section 2.1, and the adjoint code is obtained directly from the TLM. Therefore, when the problem is very nonlinear, it is important to assess the extent to which the TLM characterises the nonlinear evolution of the perturbations.

5.2 The correctness test

We use a standard method to measure the accuracy of the TLM (Yang et al., 1998) which involves comparing the solution of the evolution of a perturbation in the linear model with the perturbation defined by the difference between two runs of the nonlinear model (see also Lawless et al., 2003). The purpose of this test is to check whether the linear model is the correct linearisation of the original nonlinear problem in the vicinity of a given trajectory.

First define the model state vector at the initial time t_0 to be $\mathbf{x}_0$ and $\gamma\delta\mathbf{x}_0$ to be a small perturbation to this model state, where γ is a scalar parameter. The perturbations, $\delta\mathbf{x}_0$, were those used for the background errors viz: $w \pm 50\%$, $T \pm 1K$ and $q_t \pm 15\%$.

Let $\mathcal{M}$ represent the nonlinear forward model, such that at time t_n the model state $\mathbf{x}_n$ is given by

$$\mathbf{x}_n = \mathcal{M}(t_n, t_0, \mathbf{x}_0). \tag{23}$$

Define the perturbation evolved by the nonlinear model (hereafter NLP) at time t_n as

$$N_n[\gamma\delta\mathbf{x}_0] = \mathcal{M}(t_n, t_0, \mathbf{x}_0 + \gamma\delta\mathbf{x}_0) - \mathcal{M}(t_n, t_0, \mathbf{x}_0), \tag{24}$$

and compare this with the perturbation evolved by the TLM, which we denote $\mathbf{M}(t_n, t_0)\gamma\delta\mathbf{x}_0$.

To quantify the error we define the *linearisation error*, E_n, of the TLM at time level n by

$$E_n \equiv N_n[\gamma\delta\mathbf{x}_0] - \mathbf{M}(t_n, t_0)\ \gamma\delta\mathbf{x}_0, \tag{25}$$

and compare it with the size of the linear perturbation ($\mathbf{M}(t_n, t_0)\ \gamma\delta\mathbf{x}_0$). This gives the *relative error*, E_R, of the TLM, defined by

$$E_R = 100\frac{||\mathcal{M}(\mathbf{x}_0 + \gamma\delta\mathbf{x}_0) - \mathcal{M}(\mathbf{x}_0) - \mathbf{M}\gamma\delta\mathbf{x}_0||}{||\mathbf{M}\gamma\delta\mathbf{x}_0||}, \tag{26}$$

where $||\ ||$ represents the L_2 norm.

The standard method of proving that a TLM is coded correctly is to show that E_R tends linearly to zero as the scalar parameter γ is reduced (Lawless et al., 2003).

The TLM passed the correctness test with the relative error tending linearly towards zero (not shown).

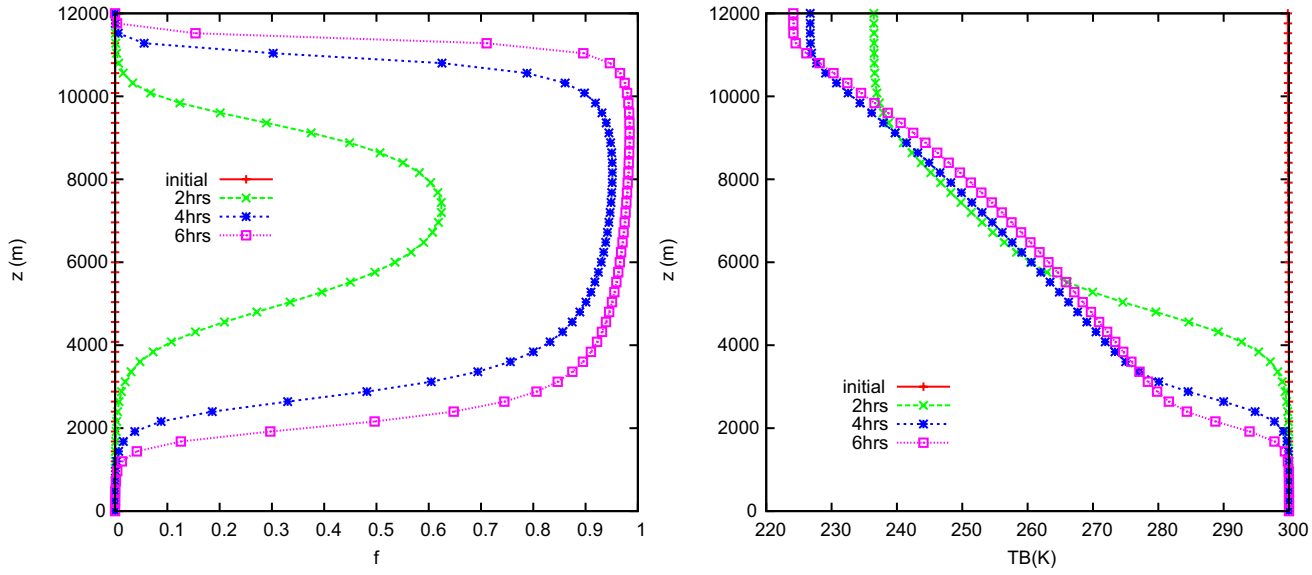
For very small perturbations ($\gamma < 1 \times 10^{-5}$) we saw that the relative error starts to increase. This behaviour was also observed by Yang et al. (1998) and was due to machine precision/round off errors.

5.3 The validity test

For the TLM to be valid it must exhibit similar behaviour to the nonlinear system to a reasonable degree of accuracy. The standard way to test the *validity* is to compare the evolution of perturbations in both the nonlinear and linear models, i.e. compare $N_n[\gamma\delta\mathbf{x}_0]$ with $\mathbf{M}(t_n, t_0)\gamma\delta\mathbf{x}_0$. As γ is reduced we expect the difference between the nonlinear model runs to decrease and therefore the TLM to become 'more valid', this is because for smaller perturbations the amount of nonlinear behaviour neglected is reduced.

We conducted experiments to test the validity of the TL hypothesis using four different initial conditions over a period of 6 hours. The different initial conditions allowed us to test the TLM at different stages of the forward integration. We test the validity for RH_{crit} values of 85% and 60%.

The perturbations to vertical velocity, temperature and total water content, for all the experiments in this section, are the same as those used in §5.2.



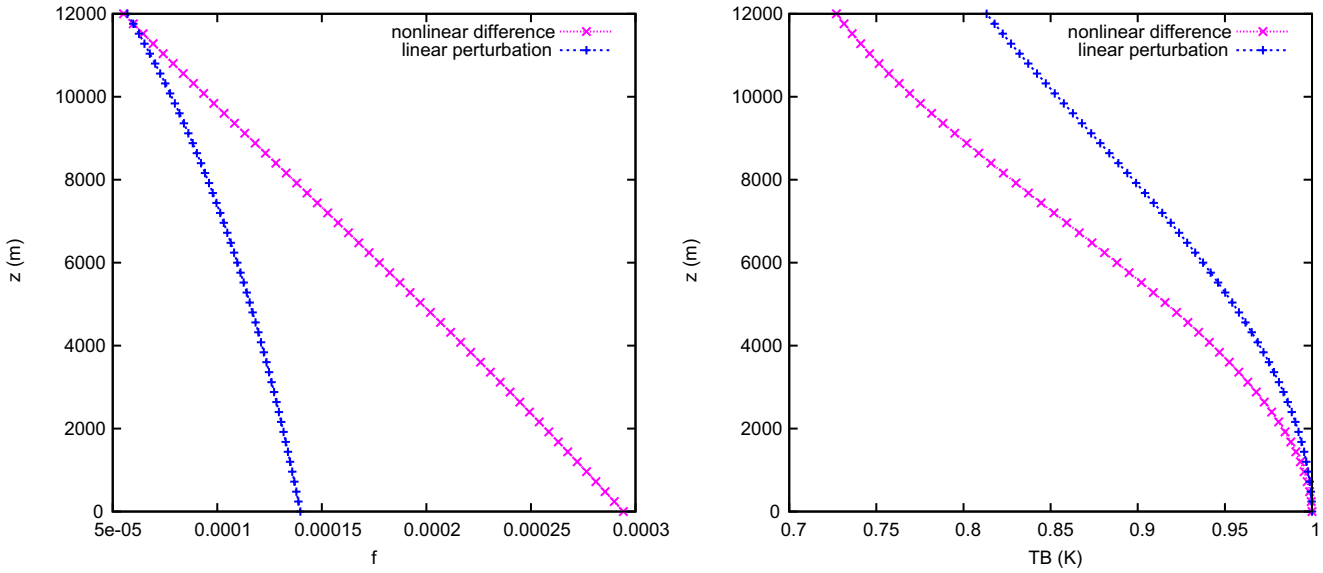
(a) Cloud fraction

(b) Brightness temperature

Figure 3: *The cloud fraction (a) and brightness temperature (b) profiles generated from the unperturbed initial state.*

Figure 3 illustrates the highly nonlinear nature of this regime. An RH_{crit} value of 85% quickly leads to cloud saturating in the mid-troposphere after just a few hours for this particular w profile. Because of the multiple cloud layers and the cloud overlap assumption (random), in Fig. 3(b) the brightness temperature saturates even faster. This will lead to a highly nonlinear variational problem.

Putting these two points together, we expect problems with observations close to the initial time, because the sensitivity of brightness temperature to changes in state variables is very small, and also at later times, because we have highly nonlinear saturation effects.



(a) Cloud fraction

(b) Brightness temperature

Figure 4: *The linear perturbation and the nonlinear difference for the cloud fraction (a) and the brightness temperature (b), at the initial time.* RH_{crit} *is* 85%.

Figures 4 to 6 show the initial, after 2 hours and after 6 hours perturbations to cloud fraction and brightness temperature ($RH_{crit} = 85\%$). The perturbations to cloud fraction and brightness temperature are very small, and it would therefore not be surprising to find that it is difficult to use the variational approach to recover the correct solution using observations in these conditions. Figure 4 shows that the linear perturbation underestimates the nonlinear difference for cloud fraction, which in turn leads to an over-estimation of the perturbation to the top model level brightness temperature.

In Fig. 5 we can see that the linear perturbation follows the shape of the nonlinear difference but it can overestimate the size of the perturbations. This is because it is in the nonlinear model that the perturbations of cloud fraction are constrained to lie in the range 0 to 1, because the values of cloud fraction themselves must lie in this range. The TLM perturbations are not constrained at all in this way and can therefore be outside the range, 0 to 1.

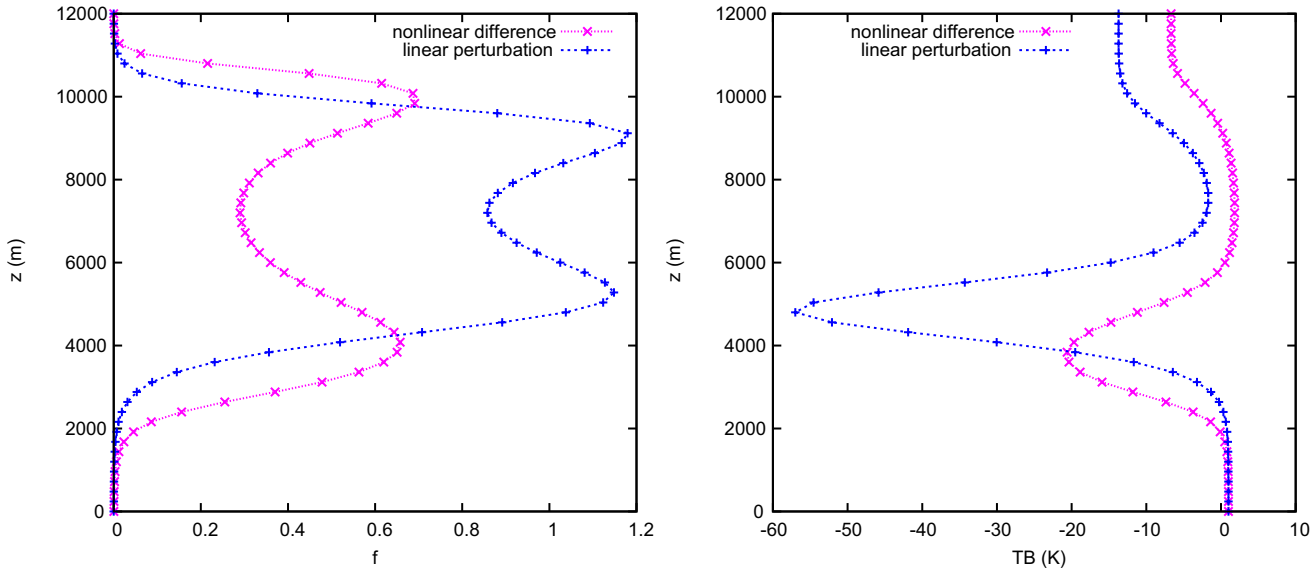
Increasing the RH_{crit} value to 85% has the effect of increasing the nonlinearity of the problem, leading to larger cloud fraction perturbations in the TLM, however the overall behaviour of the perturbations did not change. These results suggest that for the more nonlinear regimes the use of a TL approximation is questionable.

The nonlinear perturbation (NLP) and TLM perturbations got closer as γ was reduced (not shown), this suggests that the TL hypothesis is 'more valid' as the perturbation sizes are reduced.

Another way to investigate the validity of the TL hypothesis is to look at normalised values of the differences between the nonlinear and linear model solutions, i.e. the evolution of the relative error over time.

5.4 Relative error

We now investigate these issues further by looking at the relative error. In the calculations presented below we have plotted $E_R/100$, where E_R is defined as in the correctness test (26). Following Park and Droegemeier (1997) we note that for the TL hypothesis to be valid the relative error should be $<< 1$. When E_R is $>> 1$ the nonlinear effects dominate in the perturbation solution and the TL hypothesis is 'not valid'. Similarly, when E_R is $<< 1$ the perturbation is principally under the control of linear dynamics and the TL hypothesis is 'valid' (Park and Droegemeier, 1997). However, for the cases when the relative error lies in the range $0.1 < E_R < 1$ we might consider the TL hypothesis to be 'marginally valid'.



(a) Cloud fraction

(b) Brightness temperature

Figure 5: *The linear perturbation and the nonlinear difference for the cloud fraction (a) and the brightness temperature (b), after 2 hours.* RH_{crit} *is* 85%.

Using the relative error as a diagnostic we found that the TL hypothesis becomes 'more valid' as you reduce the perturbation sizes. In Fig. 7 ($RH_{crit} = 60\%$) and Fig. 8 ($RH_{crit} = 85\%$) we plot the evolution of the relative error, for each variable, over 6 hours (typical assimilation window length) using an *idealised state* (IS) (see Fig. 9) as the initial conditions. When calculating the relative error separately for each variable the behaviour is as would be expected for the vertical velocity ($E_R << 1$, not shown) and total water content (Fig. 7(b)) variables. However, the validity of the TL hypothesis becomes questionable as we move to more nonlinear regimes (Fig. 8), due to the behaviour of the perturbations for temperature, cloud fraction and brightness temperature.

6 Variational assimilation testbed

The main focus of this work was to investigate the effect of nonlinearity on the variational assimilation of satellite observations. This is addressed using 2D-Var identical twin experiments with the simple column model described in Section 3. These experiments facilitate a careful investigation into the behaviour of the data assimilation algorithm which is not possible when using real data and operational models. There are two main stages to the experiment. First, the numerical model is integrated forward in time to generate the 'true' atmospheric state. Then, observations of the 'true' state are used by the data assimilation algorithm to find the optimal state, known as the analysis. Identical twin experiments are equivalent to assuming that the model is perfect.

6.1 Observations

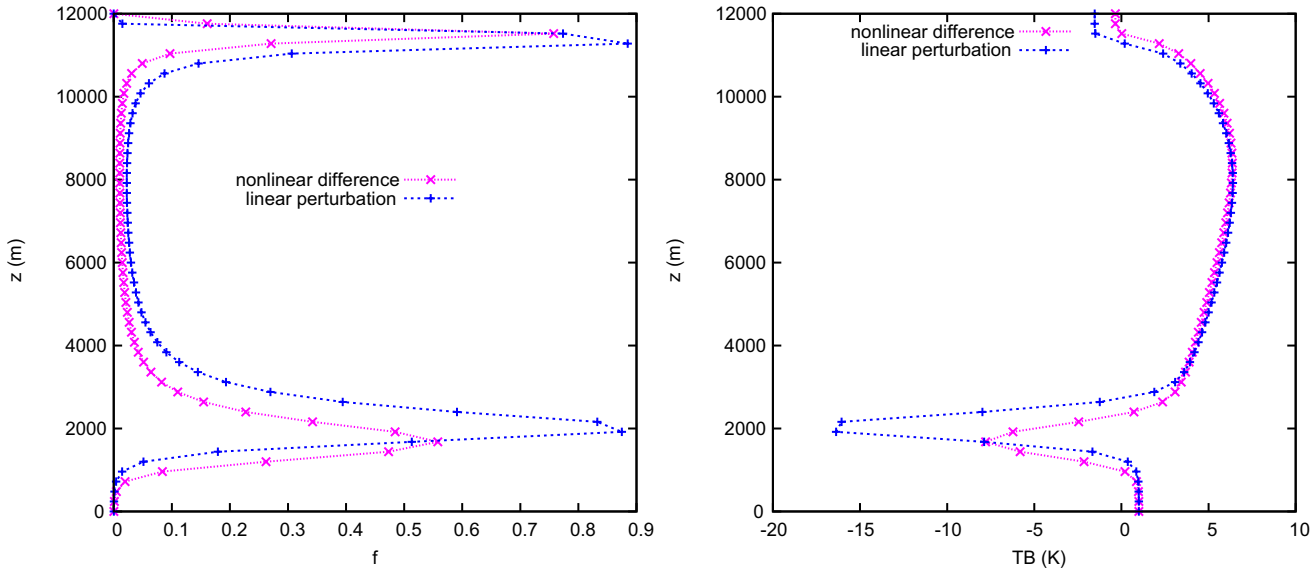
Identical twin experiments allow us to assimilate observations by using the nonlinear model to create both the 'true' and simulated observations.

Suppose that the *true state* of the atmosphere can be represented by a vector denoted $\mathbf{x}^t$ of dimension n and that the *background state* is given by $\mathbf{x}_b$. Now suppose that k observations contained in a vector $\mathbf{y}$ are related to the true state variables through an *observation operator*, h. The errors on the background state and observations are denoted $\boldsymbol{\epsilon_b}$ and $\boldsymbol{\epsilon_o}$, such that

$$\mathbf{x}_b = \mathbf{x}^t + \boldsymbol{\epsilon_b}, \tag{27}$$

and

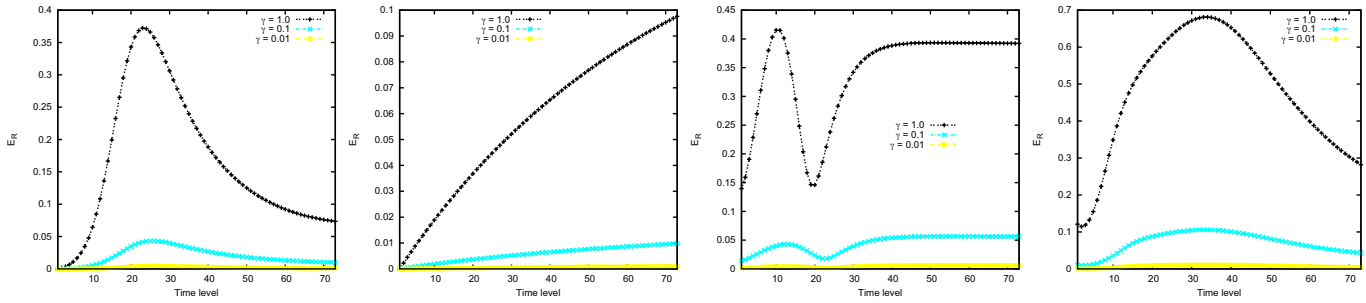
$$\mathbf{y} = h[\mathbf{x}^t] + \boldsymbol{\epsilon_o}, \tag{28}$$



(a) Cloud fraction

(b) Brightness temperature

Figure 6: *The linear perturbation and the nonlinear difference for the cloud fraction (a) and the brightness temperature (b), after 6 hours.* RH_{crit} *is* 85%.



(a) T

(b) q_t

(c) f

(d) TB

Figure 7: *Relative error for temperature,* $RH_{crit} = 60\%$. *IS.*

and in our experiments the errors are controlled to be unbiased ($E(\boldsymbol{\epsilon_b}) = E(\boldsymbol{\epsilon_o}) = 0$), where $E(x)$ denotes the expectation of x. The errors have covariances; $\mathbf{B} = E(\boldsymbol{\epsilon_b}\boldsymbol{\epsilon_b}^T)$ and $\mathbf{R} = E(\boldsymbol{\epsilon_o}\boldsymbol{\epsilon_o}^T)$, where $\mathbf{B}$ is the *background error covariance matrix* and $\mathbf{R}$ is the *observation error covariance matrix*. It is important to note that $\mathbf{R}$ not only represents the effects of measurement errors, but also the effect of errors in the observation operator (or forward model) (Simmons, 2000).

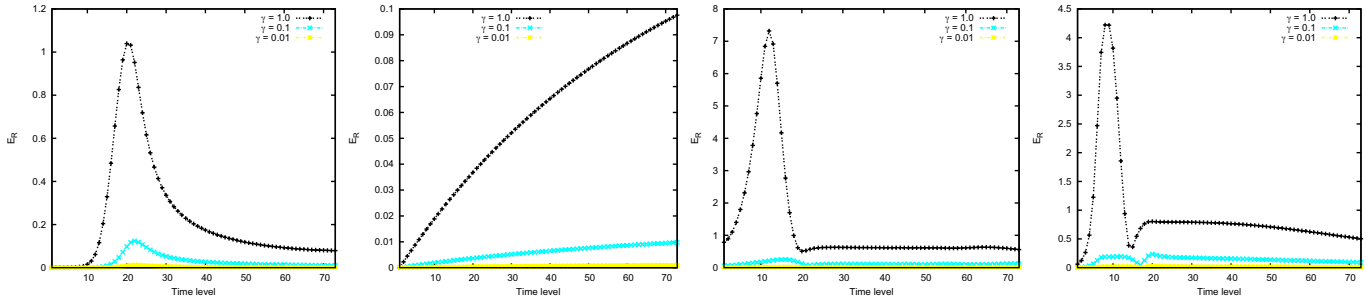
There are two types of observations used in the identical twin experiments. The 'true' observations (§6.2) which are generated from the true state as the initial conditions, and the simulated observations given the current estimate of the 'true' state (§6.3).

6.2 'True' observations

We define a true state $\mathbf{x}^t$ at the initial time, such that

$$\mathbf{x}^t = [w^t(z), T^t(z), q_t^t(z)]^T, \tag{29}$$

where the superscript t denotes the 'truth'. This true state is used as an initial condition for the nonlinear model, which is integrated forward in time in steps of Δt. This model run generates values of brightness



(a) T (b) q_t (c) f (d) TB

Figure 8: *Relative error for temperature,* $RH_{crit} = 85\%$*. IS*

temperature, TB_i^t, equivalent in our case to the TOA brightness temperature, at each observation time. The true observation, y_i, is then generated by adding error to TB_i^t as follows:

$$y_i = TB_i^t + \epsilon_o, \tag{30}$$

where ϵ_o is a random number drawn from a Gaussian distribution of zero mean and unit variance. In practice, most observation types have an error distribution close to Gaussian after quality control.

6.3 Simulated observations

We also need the simulated observations, these are the observations given the current values of the control variable in the minimisation procedure,

$$y_i^{model} = h[\mathbf{x}_i] = TB_i. \tag{31}$$

The current state will differ from the true state and therefore there will be a difference between the true and simulated observations. This difference is called the *innovation* when the current state is the background state.

6.4 Background state

We define a background state $\mathbf{x}_b$ at the initial time, such that

$$\mathbf{x}_b = [w_b(z), T_b(z), q_{tb}(z)]^T, \tag{32}$$

where the subscript b denotes the background. In operational NWP the background state would be provided by a forecast from an earlier model run. We do not have this information and therefore prescribe the background state. In NWP the background state is the state from which the minimisation is started, i.e. the first guess.

6.5 Background error covariance matrix

Strictly, we would calculate $\mathbf{B}$ from a population of forecast errors. This is not possible in our case, so instead we formulate a model of $\mathbf{B}$, with prescribed standard deviations and correlations.

Our background error covariance matrix is a real symmetric 151×151 matrix. We assume $\mathbf{B}$ is block-diagonal with no correlations between the background errors in w, T and q_t. The structure of $\mathbf{B}$ is shown below

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_w & 0 & 0 \\ 0 & \mathbf{B}_T & 0 \\ 0 & 0 & \mathbf{B}_{q_t} \end{pmatrix}, \tag{33}$$

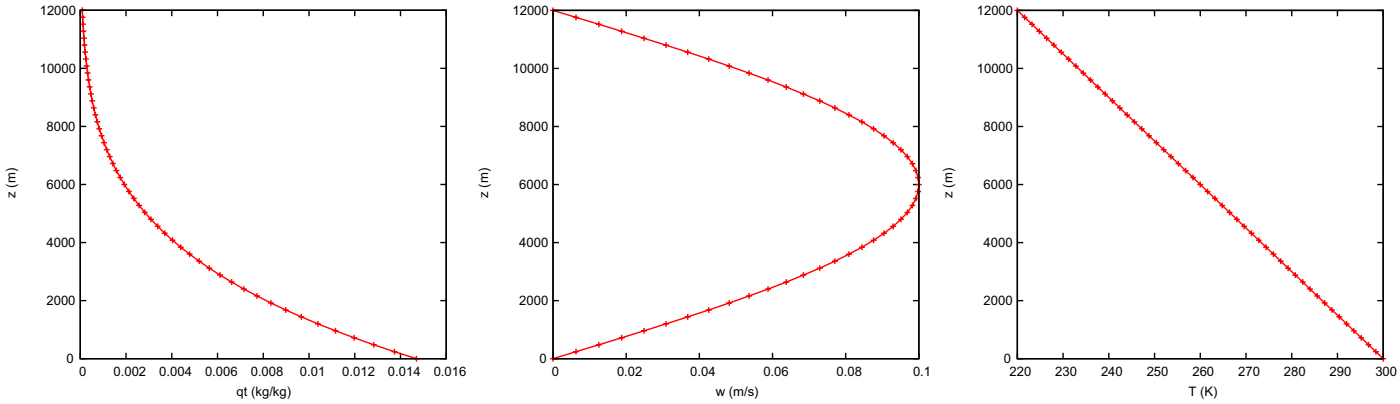
where $\mathbf{B}_w$, $\mathbf{B}_T$ and $\mathbf{B}_{q_t}$ are matrices containing the covariances of w, T and q_t respectively.

We define the $\mathbf{B}$ matrix to be of the following form;

$$\mathbf{B} = \mathbf{\Sigma}\ \mathbf{C}\ \mathbf{\Sigma}, \tag{34}$$

where $\mathbf{\Sigma}$ is a square diagonal matrix containing the standard deviations (σ_i) on the diagonals and $\mathbf{C}$ is a square symmetric matrix containing the correlations.

To construct the $\mathbf{B}$ matrix we prescribe the standard deviations σ_i and σ_j. We prescribe a 50% error over height for an Idealised State (IS) profile for vertical velocity (see Fig. 9): we do not know the error in w so we are allowing for a large uncertainty. The error on temperature and total water content are prescribed to be the typical values used in NWP, i.e. 1 K for T and 15% of the IS for q_t, these are summarised in Table 4.



(a) Vertical velocity (b) Temperature (c) Total water content

Figure 9: *Profiles of w, T and q_t defined as the Idealised State and used to calculate the standard deviations in the construction of the background error covariance matrix.*

Variable	σ
Vertical velocity	50% of w IS (ms^{-1})
Temperature	1 (K)
Total water content	15% of q_t IS (kgkg^{-1})

Table 4: *Standard deviations used to construct the background error covariance matrix. The Idealised State is shown in Fig. 9.*

The correlations $\mathbf{C}_{ij}$ are calculated using a correlation function (§6.5.1). Due to the way that $\mathbf{C}$ and $\mathbf{B}$ are constructed, the diagonal elements of $\mathbf{B}$ are the variances, σ^2.

6.5.1 Correlation matrix

For simplicity we assume there are no correlations between w, T and q_t. This means that the correlation matrix $\mathbf{C}$ is also block-diagonal and of size (151×151);

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_w & 0 & 0 \\ 0 & \mathbf{C}_T & 0 \\ 0 & 0 & \mathbf{C}_{q_t} \end{pmatrix}. \tag{35}$$

$\mathbf{C}_w$ is of size (49×49), and $\mathbf{C}_T$ and $\mathbf{C}_{qt}$ are of size (51×51).

We use a correlation function to calculate the non-diagonal elements of $\mathbf{C}$,

$$C_{ij} = \exp\left(-\left(\frac{z_i - z_j}{z_0}\right)^2\right), \tag{36}$$

where z_0 is a length scale. We specify reasonable length scales for each parameter, see Table 5. For vertical velocity we use 5 km so that it represents only large-scale motion that is coherent throughout the troposphere.

Variable	z_0
Vertical velocity	5 km
Temperature	2 km
Total water content	1 km

Table 5: *Length scale, z_0, values used to construct the $\boldsymbol{C}$ matrix. The total height in z is 12 km and the vertical spacing between levels is 240 m.*

6.6 Observation error covariance matrix

For simplicity we specify that the observation error covariance matrix is diagonal, i.e. there are no correlations between observations at different times. This is a reasonably good approximation for most observation types, and it is used in operational practice. The diagonal elements are the error variances, σ_o^2, which we prescribe to be 1 K^2. This is a typical value of an error of an observation of brightness temperature from a satellite, if we include both the error in the observation and the error in the forward model. Imposing a diagonal matrix in a 2D setting removes the need for matrix inversion of $\mathbf{R}$. For calculations of the cost function, $\frac{1}{\sigma_o^2}$ is used.

7 Adjoint model (ADJM)

The adjoint model is required in the 2D-Var assimilation and was obtained by transposing the TLM source code. We tested the ADJM using two standard tests.

7.1 The adjoint test

The adjoint code should satisfy

$$< \mathbf{M}\mathbf{x}_0, \mathbf{M}\mathbf{x}_0 >=< \mathbf{x}_0, \mathbf{M}^T\mathbf{M}\mathbf{x}_0 > \tag{37}$$

where the brackets $< \ldots, \cdots >$ denote an inner product and $\mathbf{M}$ is the TLM, which is integrated from t_0 to t_N, and $\mathbf{M}^T$ is the adjoint model, which is integrated from t_N to t_0. Evaluating the left hand side only involves the tangent linear code, while the right hand side uses the adjoint code (Li et al., 1994). We found that (37) was satisfied to the accuracy of machine precision when applied to our code, and therefore conclude that the correct adjoint model had been constructed.

7.2 The gradient test

The second test (*gradient test*) is used to check the cost function and its gradient are coded correctly. The gradient test is based on a comparison of a finite difference representation of the gradient of the cost function and the gradient from the adjoint method. Rearranging a Taylor series expansion of the cost function we can define

$$\Phi(\alpha) = \frac{J(\mathbf{x}_0 + \alpha\delta\mathbf{x}_0) - J(\mathbf{x}_0)}{\alpha\delta\mathbf{x}_0^T\nabla J(\mathbf{x}_0)} = 1 + \mathcal{O}(\alpha), \tag{38}$$

where α is a small scalar and $\nabla J(\mathbf{x}_0)$ is the gradient obtained by the adjoint. If the gradient is correct we expect $\Phi(\alpha) \to 1$ as $\alpha \to 0$. Thus $\Phi(\alpha) - 1$ should be close to zero for values of α which are small but not too close to machine zero (Navon et al., 1992). We then plot the variation of $\Phi(\alpha)$, and the variation of the residual, log $(|\Phi(\alpha) - 1|)$, with respect to α.

Usually $\delta\mathbf{x}_0$ is taken to be

$$\delta\mathbf{x}_0 = \frac{\nabla J(\mathbf{x}_0)}{||\nabla J(\mathbf{x}_0)||}, \tag{39}$$

where $||\ ||$ is the L_2 norm. This is so that $\delta\mathbf{x}_0$ is a vector in the gradient direction and thus the variation of the variables gives consistent scaling (Li et al., 1993).

We used this test to separately test the gradient of the background and observation terms of the cost function. We often found that $\Phi(\alpha)$ tended towards one, but not exactly one, for the observation term and tended to exactly one for the background term (not shown). The offset from one increased as the value of RH_{crit} was increased (Fig. 10). The next section investigates whether the nonlinearities in the model give rise to the offset from one and highlight the importance of full investigation when using these standard tests.

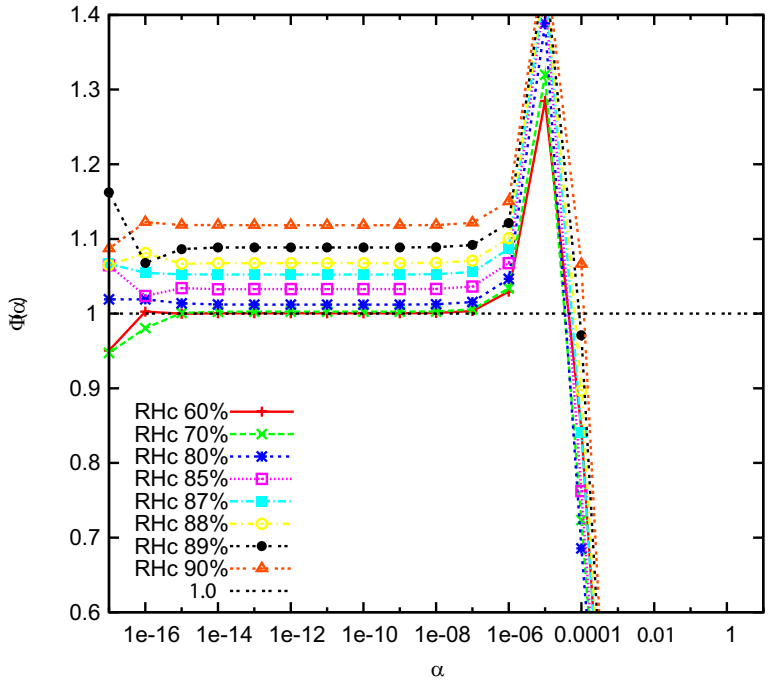


Figure 10: *Verification of the gradient calculation for RH_{crit} in the range 60% to 90% for a 2 hour integration of the nonlinear model.*

7.3 Higher-order terms

An adaptation of a method used by Lawless et al. (2003) is used to estimate the higher-order terms that are neglected in the construction of the TLM and ADJM. This calculation enables us to investigate whether the higher-order terms (HOTs) that are neglected in the construction of the TLM and ADJM are significant.

Following Lawless et al. (2003), we define a function $\mathcal{E}_n$;

$$\mathcal{E}_n = \frac{\mathcal{N}_n[\beta\delta\mathbf{x}_0] - \beta\mathcal{N}_n[\delta\mathbf{x}_0]}{\beta^2 - \beta}, \tag{40}$$

where β is a small scalar parameter, and $\mathcal{N}_n[\beta\delta\mathbf{x}_0]$ and $\mathcal{N}_n[\delta\mathbf{x}_0]$ are two nonlinear perturbations defined in terms of the cost function J:

$$\mathcal{N}_n[\beta\delta\mathbf{x}_0] = J(\mathbf{x}_0 + \beta\delta\mathbf{x}_0) - J(\mathbf{x}_0) \tag{41}$$

and

$$\mathcal{N}_n[\delta\mathbf{x}_0] = J(\mathbf{x}_0 + \delta\mathbf{x}_0) - J(\mathbf{x}_0). \tag{42}$$

We compare (40), as $\beta \to 0$, with the linearisation error (HOTs from the adjoint)

$$E_n = J(\mathbf{x}_0 + \delta\mathbf{x}_0) - J(\mathbf{x}_0) - \nabla J(\mathbf{x}_0).\ \delta\mathbf{x}_0, \tag{43}$$

where $\nabla J(\mathbf{x}_0)$ is the gradient calculated using the adjoint.

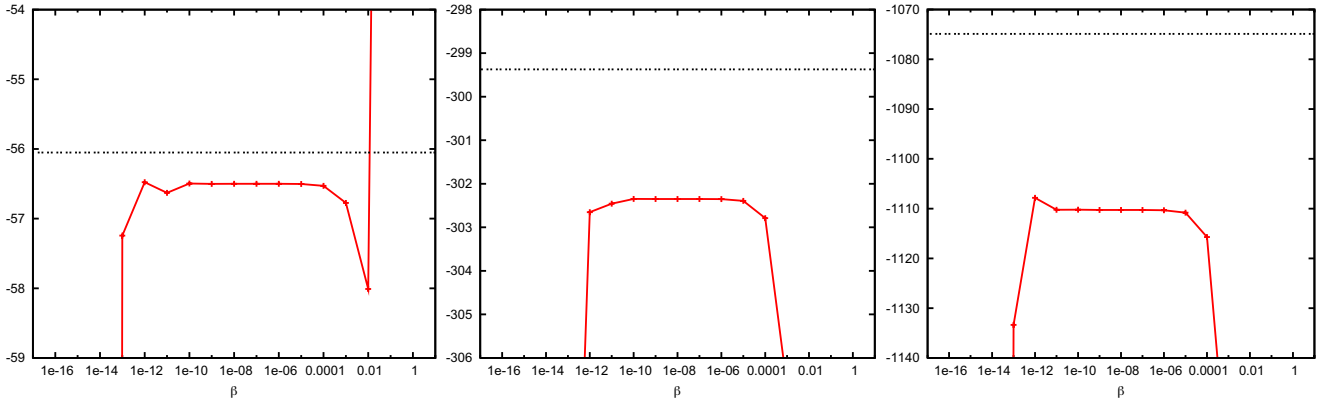
Using a Taylor series expansion it can be shown that

$$\begin{aligned}\mathcal{E}_n = &\frac{1}{2!}\sum_{i=1}^{p}\sum_{j=1}^{p}\frac{\partial^2 J}{\partial x^i \partial x^j}(\mathbf{x}_0)\delta x_0^i \delta x_0^j \\ &+ (1+\beta)\frac{1}{3!}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\frac{\partial^3 J}{\partial x^i \partial x^j \partial x^k}(\mathbf{x}_0)\delta x_0^i \delta x_0^j \delta x_0^k \\ &+ \text{higher-order terms.}\end{aligned} \tag{44}$$

A comparison of (44) with (43) shows that for small values of β and small perturbations $\delta\mathbf{x}_0$ we have $\mathcal{E}_n \approx E_n$. We calculate (40) from the nonlinear model alone and this provides us with the information we require — an estimate of the higher-order terms that are ignored in the gradient test, calculated without using the adjoint.

We apply these calculations to the observation term of the cost function, as this is where the nonlinearity is prevalent. The initial conditions are the same as those used in the gradient test (§7.2). We calculate the two linearisation errors for a 30 minute and 2 hour integration of the nonlinear forward model.

We find that the estimates of the HOTs from the two methods are a similar magnitude, suggesting that the gradient from the adjoint is correct. We conclude that the offset in the gradient test is due to the nonlinearity and that the gradient from the adjoint model is correct.



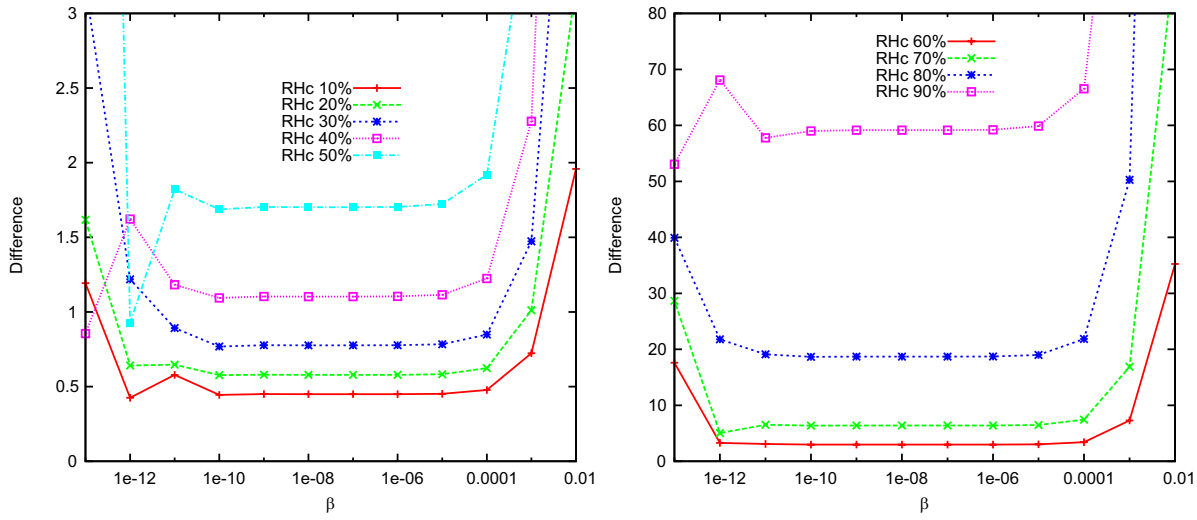
(a) RH_{crit} 10% (b) RH_{crit} 60% (c) RH_{crit} 85%

Figure 11: *HOTs calculated using the adjoint, E_n (dashed) and cost function, $\mathcal{E}_n$ (red). We assimilate four observations (every 30 mins).*

The calculation of $\mathcal{E}_n$ versus E_n is given in Fig. 11 for different RH_{crit} values. The difference between the two calculations, i.e. the difference,

$$\text{Difference} = \mathcal{E}_n - E_n, \tag{45}$$

is shown in Fig. 12. Figure 12(a) shows the difference between the HOTs when $10\% \leq RH_{crit} \leq 50\%$, and Fig. 12(b) shows the difference when $60\% \leq RH_{crit} \leq 90\%$. Note the magnitude of the scales, the difference is much greater in Fig. 12(b) for the larger RH_{crit} values. These results support the conclusion that the offsets in the gradient test are due to HOTs.



(a) (b)

Figure 12: *Difference between the estimated HOTs and the HOTs from the adjoint for* $10\% \leq RH_{crit} \leq 90\%$.

Changing RH_{crit} changes the threshold at which cloud starts to form, and it increases the nonlinearity of the system. We conclude that our gradient is coded correctly but are aware that because of the nonlinear nature of our system the gradient, from the adjoint integration, may not be as accurate as it would be for a linear system.

The gradient test appears to be very sensitive to the nonlinearities in the model and it is questionable how appropriate this test will be in the future as we try to make more use of nonlinear models. The testing of the TLM and ADJM suggested that we might encounter problems within the minimisation.

8 2D-Var experiments

The preliminary 2D-Var experiments were designed to answer three main questions:

1. Can the algorithm converge when the simulated observations are colder than the true observations?
2. Can the algorithm converge when the true observations are colder than the simulated observations?
3. Can the vertical velocity profile that would have given rise to the difference in the observations be recovered?

The assimilation window for these experiments was 3 hours, with a time step of 5 minutes and an observation every 30 minutes, i.e. 6 observations. We investigate the behaviour of the 2D-Var system for RH_{crit} values of 60% and 85%. The results presented here are from identical twin experiments, as outlined in §6. For the experiments to address questions 1 and 2 the initial vertical velocity and temperature profiles are identical for the true and background states. The difference in the observation comes only from differences in the initial total water content profiles, and the error on the observations. We specify the initial RH_T value (constant with height), from which the total water content profile is derived. The larger the initial RH_T, the wetter the profile and the more cloud there will be initially.

For the experiments where we prescribe that there is more cloud, at the initial time, in the background state than in the true state, this gives rise to colder simulated observations and ($y_i > y_i^{model}$). For $RH_{crit} = 60\%$, with a true total water content profile derived from an initial RH_T value of 10% and the background from $RH_T = 30\%$, the minimisation algorithm (nonlinear conjugate gradient) stops after 120 iterations when the fractional change in the cost function is no greater than 0.1%. Figure 13(a) shows how the cost function varies with iteration number. We can see that the value of the cost function has decreased considerably, a feature of convergence. Figure 13(c) shows the variation of the norm of the gradient of J. This norm would be zero if the algorithm had converged on the global (or a local) minimum. We can see that although the value of the norm of the gradient of J has also decreased (a sign of convergence), it is still large in magnitude.

To investigate why the value of the norm of the gradient of J remains high we plot the contributions to the norm of the gradient from each of the state variables. Figure 13(d) shows the squared components of the norm of the gradient for vertical velocity, temperature and total water content separately.

The norm of ∇J is

$$\begin{aligned} |\nabla J| &= \sqrt{\sum_{j=1}^{151} \left(\frac{\partial J}{\partial x_j}\right)^2} \\ &= \sqrt{\sum_{k=1}^{49} \left(\frac{\partial J}{\partial w_k}\right)^2 + \sum_{l=1}^{51} \left(\frac{\partial J}{\partial T_l}\right)^2 + \sum_{l=1}^{51} \left(\frac{\partial J}{\partial q_{tl}}\right)^2}, \end{aligned} \tag{46}$$

and the square of the norm is denoted GSQ:

$$GSQ \equiv |\nabla J|^2. \tag{47}$$

In Fig. 13(d) we can see that the largest contribution to the gradient is from the total water content variable. The other components are considerably smaller. It is the gradients with respect to the total water content variable that are making the value of the norm of the gradient remain high in Fig. 13(c).

The algorithm removes cloud to fit the observations, and it does this by drying out the total water content profile, mainly at model levels $z > 2$ km. It has not attempted to fit the observations by significantly changing

the temperature and vertical velocity profiles. The difference between the background and analysis profiles are very insignificant compared to the changes to the total water content profile (not shown). We see similar behaviour of the algorithm for $RH_{crit} = 85\%$.

For the experiments where we prescribe that there is less cloud, at the initial time, in the background state than in the true state, this gives rise to colder true observations and ($y_i < y_i^{model}$). In these cases the algorithm tries to fit the observations by moistening the profile and creates cloud in the range $5km < z < 12km$. Figure 14 shows the value of the cost function and the norm of the gradient with iteration number when $RH_{crit} = 85\%$.

When RH_{crit} is 85% we often found the minimisation algorithm failed to stop on our convergence criteria. This is due to the highly nonlinear nature of these regimes leading to large differences between the true and simulated observations over time. The cost function often oscillated and could not converge on a particular estimate of $\mathbf{x}_0$ (Fig. 14).

For the cases where the true observations were colder than the simulated (from the background state) observations, and constant over time, the minimisation algorithm was able to fit the observations by making the profile wetter and creating cloud at the top model levels. This is a result of the hidden layer problem. The observations cannot "see" the atmosphere below the uppermost levels.

When RH_{crit} is 85%, and the true observations are colder than the simulated observations, we could not find a case where the algorithm managed to improve the fit to the true observations.

We also tested the algorithm by setting the background and true state temperature and total water content profiles to be identical, and only changed the vertical velocity profiles. This was to test if we could recover the correct vertical velocity profile that gave rise to the observations, however, once again the algorithm attempted to fit the observations by only adjusting the total water content profile and this was not adequate to fit to the true observations over time.

9 Discussion and conclusions

This paper has presented results from an idealised experiment to test a simplified 2D-Var system for directly assimilating sequences of satellite observations. Every stage has been thoroughly tested using standard tests and it has become apparent that as grid resolutions become finer and models become more nonlinear, the appropriateness of these tests becomes questionable.

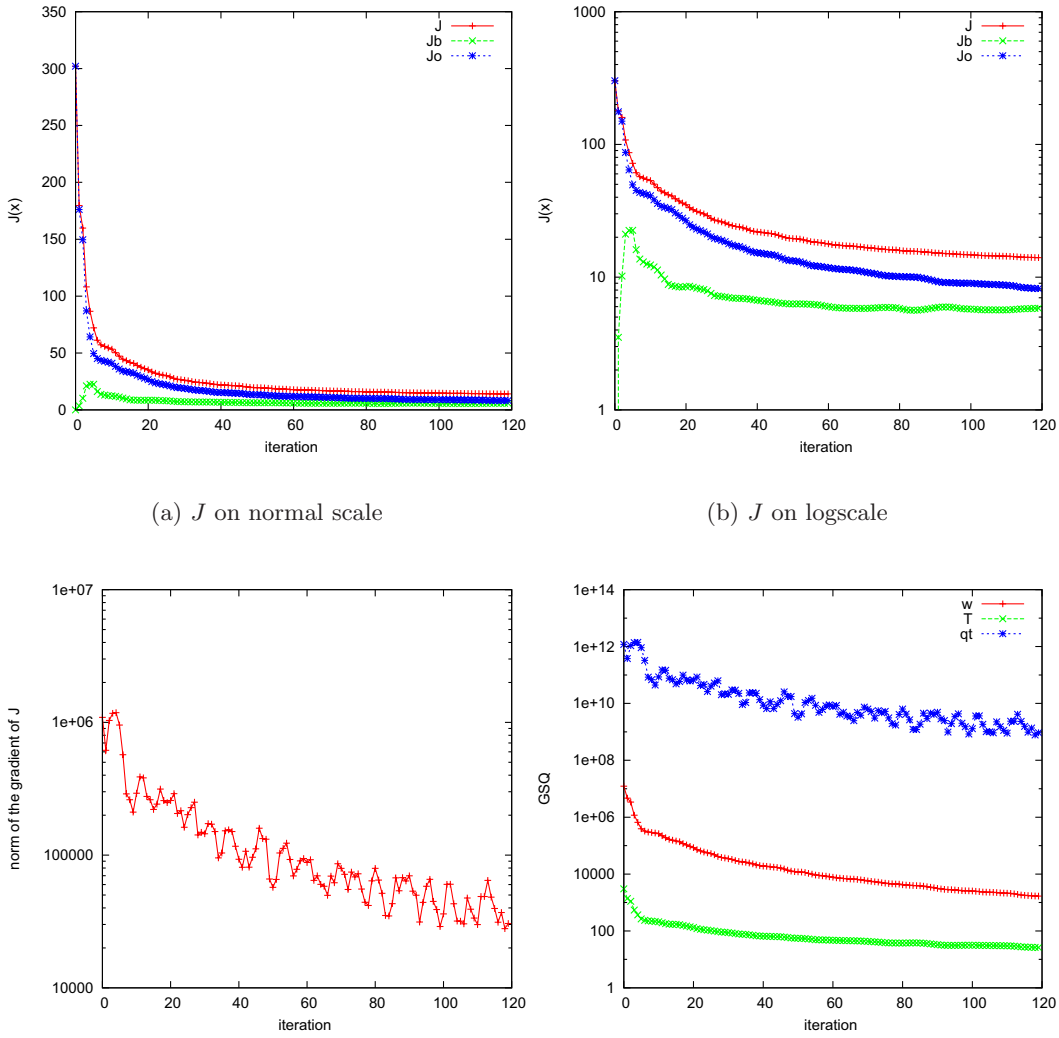
We have shown that the TL hypothesis becomes questionable as we move to more nonlinear regimes, i.e. when RH_{crit} is 85%. We note that the gradient from the adjoint might not be as accurate for these regimes. The term 'accuracy' here is used in conjunction with the recognition that the cost function may possess multiple minima, at least locally. Therefore small changes in the descent direction could lead to spurious results.

It has been shown that there are significant higher-order terms in the cost function for more nonlinear regimes, i.e. those when RH_{crit} is greater than 60%. An adaptation of the Lawless et al. (2003) method for estimating the linearisation error was used to estimate the higher-order terms neglected in the construction of the TLM and ADJM.

During the 2D-Var experiments the minimisation algorithm tried to fit the observations by changing the total water content profile even when the changes in the observations were only due to the effect of the vertical velocity profile. The minimisation algorithm was unable to recover the vertical velocity profile.

We also conducted an observability study (see Rudd (2009) §10.2.5) to investigate if it is reasonable to expect our 2D-Var assimilation system to be able to recover the vertical profiles of w, T, and q_t from an observation of brightness temperature. We examined how the value of the observation term of the scalar cost function J^o varies with perturbations (size of background errors and $RH_{crit} = 85\%$) to the initial conditions. This provides an insight into how sensitive the cost function and its gradient are to changes in the state variables. We found that when the initial profile is wet ($80\% \geq RH_T$) J^o is most sensitive to changes in the initial total water content profile. However, when the profile is drier J^o becomes comparably sensitive to the vertical velocity. The sensitivity of J^o to perturbations of T are comparable with the sensitivity to q_t for the wettest profile examined ($RH_T = 80\%$). However, negative perturbations to the initial T profile have the opposite effect on J^o to negative perturbations to q_t. For the drier profile ($RH_T = 60\%$) the sensitivity of J^o to w was comparable with the sensitivity to q_t. This study did suggest that the size of the background errors for w were sufficient for producing sensitivities in J^o.

We conclude that in certain regimes (wetter initial conditions, before perturbing) the observation term of the cost function seems to be more sensitive to changes in the q_t profile than to changes in T or w. Suggesting that in certain regimes making changes to the total water content profile, instead of the vertical velocity profile or



(a) J on normal scale

(b) J on logscale

(c) Norm of ∇J

(d) Components of GSQ

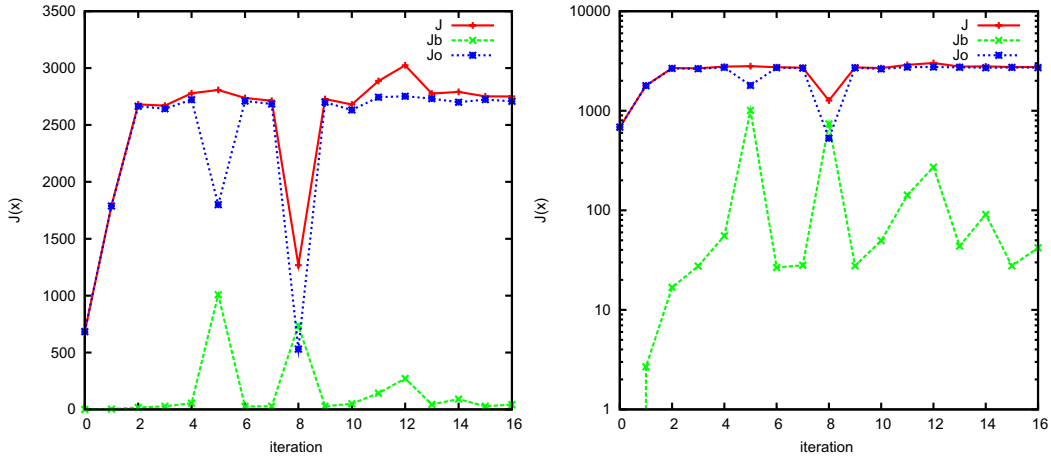
Figure 13: *The variation of the cost function with iteration number in the minimisation (13(a) and 13(b)). The variation of the norm of the gradient of J, and the components of GSQ, on a logscale (13(c) and 13(d)). $RH_{crit} = 60\%$.*

temperature profile, may give a more significant reduction in the cost function, which is what the minimisation algorithm is trying to achieve.

Our goal was to design and build a simplified system which allows explicit investigation of the nonlinearities inherent in the problem. It is shown that our model exhibits realistic behaviour with regard to the prediction of cloud, but the effects of nonlinearity become non-negligible in the variational data assimilation algorithm.

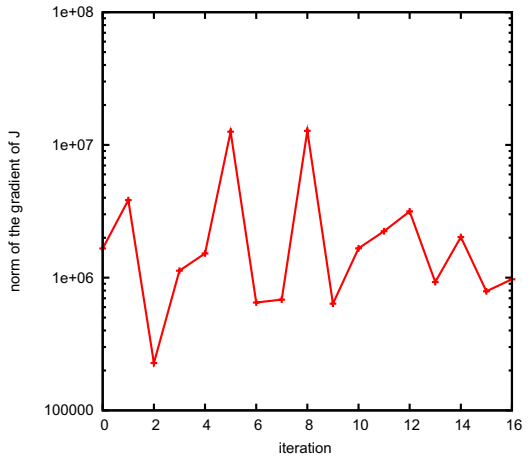
We have highlighted several issues that need further investigation. In particular our study indicates that care needs to be taken in using 4D-Var in situations in which nonlinearity becomes important, for example, as model resolution increases. Although we have investigated nonlinearities arising from the cloud scheme we note that there is also a source of nonlinearity that stems from the overlap assumption within the radiative transfer scheme. An investigation into this overlap assumption is something that can be addressed in subsequent work.

The model presented here has been used for further studies (see Vetra-Carvalho et al. (2010)), within the context of an ensemble data assimilation method, and as part of on-going work at the Met Office and the University of Surrey. The code is available from the Corresponding Author.

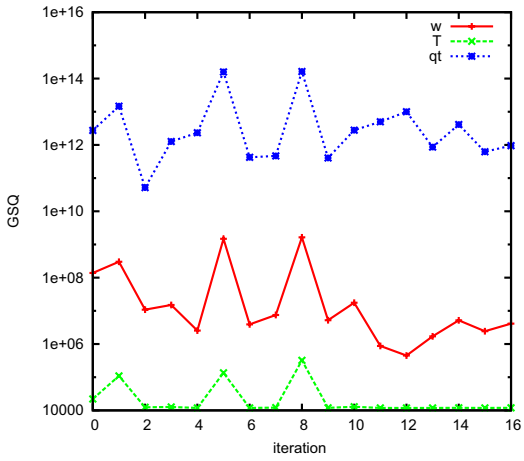


(a) J on normal scale

(b) J on logscale



(c) Norm of ∇J



(d) Components of GSQ

Figure 14: *The variation of the cost function with iteration number in the minimisation (14(a) and 14(b)). The variation of the norm of the gradient of J, and the components of GSQ, on a logscale (14(c) and 14(d)). $RH_{crit} = 85\%$.*

Acknowledgements

The authors would like to thank Dr. Amos Lawless, Dr. Ross Bannister, Prof. Mike Cullen, Prof. Nancy Nichols, Dr. Ed Pavelin and Martin Sharpe for many useful discussions on this work. The authors would also like to thank three anonymous reviewers for their helpful comments. The Corresponding Author would like to thank the EPSRC and the UK Met Office for financial support through a CASE studentship.

References

Clark, M., Slater, A., Barrett, A., Hay, L., McCabe, G., Rajagopalan, B., Leavesley, G., 2006. Assimilation of snow covered area information into hydrologic and land-surface models. Adv. Water Resour. 29, 1209–1221.

Daley, R., 1991. Atmospheric Data Analysis. Cambridge University Press.

Dunlop, S., 2001. A Dictionary of Weather. Oxford University Press.

Errico, R. M., Bauer, P., Mahfouf, J.-F., 2007. Issues regarding the assimilation of cloud and precipitation data. J. Atmos. Sci 64, 3785–3798.

Eyre, J., 2007. Progress achieved on assimilation of satellite data in numerical weather prediction over the last 30 years. ECMWF Seminar on recent development in the use of satellite observations in NWP, 3-7 September 2007, 1–28.

Houghton, J., 2002. The Physics of Atmospheres. Cambridge University Press.

Kalnay, E., 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge, CB2 2RU, UK.

Lawless, A. S., Nichols, N. K., Ballard, S., 2003. A comparison of two methods for developing the linearization of a shallow-water model. Q. J. R. Meteorol. Soc 129, 1237–1254.

Lewis, J., Derber, J., 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus 37A 4, 309–322.

Li, Y., Navon, I. M., Courtier, P., Gauthier, P., 1993. Variational data assimilation with a semi-Lagrangian semi-implicit global shallow-water equation model and its adjoint. Mon. Weather Rev 121, 1759–1769.

Li, Y., Navon, I. M., Yang, W., Zou, X., Bates, J, R., Moorthi, S., Higgins, R., 1994. Four-dimensional variational data assimilation experiments with a multilevel semi-Lagrangian semi-implicit global circulation model. Mon. Weather Rev 122, 966–983.

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

Naud, C., Makowski, D., Jeuffroy, M.-H., 2007. Application of an interacting particle filter to improve nitrogen nutrition index predictions for winter wheat. Ecol. Model. 207, 251–263.

Navon, I., Zou, X., Derber, J., Sela, J., 1992. Variational data assimilation with an adiabatic version of the NMC spectral model. Mon. Weather Rev. 120, 1433–1446.

Neal, J., Atkinson, P., Hutton, C., 2009. Evaluating the utility of the ensemble transform Kalman filter for adaptive sampling when updating a hydrodynamic model. J. Hydro. 375, 589–600.

Park, S., Droegemeier, K. K., 1997. Validity of the tangent linear approximation in a moist convective cloud model. Mon. Weather Rev. 125, 3320–3340.

Rudd, A. C., 2009. The effect of nonlinearity on the variational assimilation of satellite observations using a simple column model. Ph.D. thesis, Department of Mathematics, University of Surrey.

Seo, D.-J., Cajina, L., Corby, R., Howieson, T., 2009. Automatic state updating for operational streamflow forecasting via variational data assimilation. J. Hydrol. 367, 255–275.

Simmons, A., 2000. Assimilation of satellite data for numerical weather prediction: Basic importance, concepts and issues. ECMWF Seminar: Exploitation of the New Generation of Satellite Instruments for Numerical Weather Prediction, 4-8 September 2000, 21–46.

Smith, P., Dance, S., Nichols, N., 2011. A hybrid data assimilation scheme for model parameter estimation: Application to morphodynamic modelling. Comput. fluids 46, 436–441.

Smith, R. N. B., 1990. A scheme for predicting layer clouds and their water contents in a general circulation model. Q. J. R. Meteorol. Soc. 116, 435–460.

Titaud, O., Vidard, A., Souopgui, I., Le Dimet, F.-X., 2010. Assimilation of image sequences in numerical models. Tellus 62A, 30–47.

van Leeuwen, P., 2003. A variance-minimizing filter for large-scale applications. Mon. Weather Rev. 131, 2071–2084.

van Velzen, N., Segers, A., 2010. A problem-solving environment for data assimilation in air quality modelling. Env. Modell. Softw. 25, 277–288.

Vetra-Carvalho, S., Migliorini, S., Nichols, N. K., 2010. http://www.reading.ac.uk/nmsruntime/saveasdialog.aspx?lID=52575\&sID=90309.

Wood, R., Field, P., 2000. Relationships between total water, condensed water, and cloud fraction in stratiform clouds examined using aircraft data. J. Atmos. Sci. 57, 1888 – 1905.

Yang, Y., Navon, I., Todling, R., 1998. Documentation of the multitasked tangent linear and adjoint models of the adiabatic version of the NASA GEOS-2 GCM (version 6.5). www.math.fsu.edu/~aluffi/archive/paper93.ps.gz.

A The forward model

At the initial time, on all model levels, j:

- prescribe w^j, T^j and q_t^j,
- calculate p^j,
- call the cloud scheme to calculate e_s^j and f^j,
- call the compute Psi subroutine to calculate Γ_s^j and Ψ^j,
- call the ratiative transfer scheme to calculate TB,
- call model obs to save TB as y^{model}.

We now have values for all the variables at the initial time. Next the advection scheme is used to update temperature and total water content in time and then the observation of brightness temperature is calculated using the cloud and radiative transfer schemes:

- calculate, on all model levels:

$$q_{t_{(i+1)}}^j(SL) = (1-\alpha)q_{t_i}^{j-p} + \alpha q_{t_i}^{j-p-1},$$
$$\Psi_{i+1}^j(SL) = (1-\alpha)\Psi_i^{j-p} + \alpha\Psi_i^{j-p-1},$$
$$T_{i+1}^j(SL) = (1-\alpha)T_i^{j-p} + \alpha T_i^{j-p-1},$$

  include the adjustment to temperature (adiabatic cooling/warming)

$$\Delta T_{i+1}^j = -\Psi_{t+1}^j(SL)w_1^j\Delta t,$$
$$T_{i+1}^j = T_{i+1}^j(SL) + \Delta T_{i+1}^j,$$

- calculate p_{i+1}^j,
- call the cloud scheme to compute $e_{s_{(i+1)}}^j$ and f_{i+1}^j,
- call the compute Psi subroutine to calculate Ψ_{i+1}^j using the new cloud profile,
- call the radiative transfer scheme to compute the upwelling brightness temperature TB_{i+1},
- call model obs to save TB_{i+1} as y_{i+1}^{model}.