Assimilation of Research Satellite Data into the Met Office 3d-Var. System
Part I: Forward Models and Adjoints for Layer Averaging
DARC Technical Report No. 3

Ross N. Bannister, Data Assimilation Research Centre, University of Reading, UK
Last revised 27th January 2002

Contents

1. Introduction

The starting point of atmospheric variational data assimilation [1], is an estimate of the atmospheric state in model representation. The aim is to refine this estimate using the best available information, consisting of the latest observations and a forecast model. Adjustments to the model state are made by minimizing a cost function, J, which is a measure of the misfit between the current estimate and the incoming information.

Let the current estimate of the state of the atmosphere, $\vec{x}$, be specified as a perturbation, $\vec{x}'$ from a guess state, $\vec{x}_G$, such that,

$$\vec{x} = \vec{x}_G + \vec{x}', \tag{1.1}$$

and the background, $\vec{x}_B$, is,

$$\vec{x}_B = \vec{x}_G + \vec{x}'_B. \tag{1.2}$$

The cost function in perturbation variables is,

$$J(\vec{x}') = \frac{1}{2}(\vec{x}'_B - \vec{x}')^T \mathbf{B}^{-1}(\vec{x}'_B - \vec{x}') + \frac{1}{2}(\vec{y} - \vec{H}[\vec{x}_G + \vec{x}'])^T (\mathbf{E} + \mathbf{F})^{-1}(\vec{y} - \vec{H}[\vec{x}_G + \vec{x}']). \tag{1.3}$$

In Eq. (1.3), $\mathbf{B}$ is the background error covariance matrix, $\vec{y}$ is the vector of observations and $\vec{H}$ is the forward model operator (giving model equivalent of the observations). The error correlation matrix of the observations, $\mathbf{E} + \mathbf{F}$, is a combination of the error of representativeness and the instrumental error [2].

As it stands, the cost function in model space (ie $\vec{x}'$-space) is badly conditioned, and involves directly the use of $\mathbf{B}^{-1}$, which has a rank too large to deal with practically. Instead of $\vec{x}'$, minimization is done with respect to a

control vector, $\vec{v}'$, which has none of these problems. If $\vec{v}'$ is related to $\vec{x}'$ via [2],

$$\vec{x}' = \mathbf{U}\vec{v}', \qquad \vec{v}' = \mathbf{T}\vec{x}', \qquad (1.4)\ (1.5)$$

($\mathbf{T} = \mathbf{U}^{-1}$) then the preconditioned cost function in control variable space can be written,

$$J(\vec{v}') = \frac{1}{2}(\vec{v}'_B - \vec{v}')^T(\vec{v}'_B - \vec{v}') + \frac{1}{2}(\vec{y} - \vec{H}[\vec{x}_G + \mathbf{U}\vec{v}'])^T(\mathbf{E} + \mathbf{F})^{-1}(\vec{y} - \vec{H}[\vec{x}_G + \mathbf{U}\vec{v}']). \quad (1.6)$$

Here the background perturbation in $\vec{v}'$-space is $\vec{v}'_B = \mathbf{T}\vec{x}'_B$. We have chosen the transformation such that the background error co-variance matrix is absent in the background term (first term of Eq. (1.6)). Thus in $\vec{v}'$-space, the background error correlation is the unit matrix [3]. This is an objective of preconditioning [2], and $\mathbf{B}$ is implicit in the transformation, $\mathbf{B} = \mathbf{U}\mathbf{U}^T$.

In order to simplify the notation, it is convenient in the discussion below to write, $\vec{y}_{mo} = \vec{H}[\vec{x}_G + \mathbf{U}\vec{v}']$.

How is J minimized? A descent algorithm makes the desired adjustments, but this requires the gradient of J with respect to $\vec{v}'$. The state $\vec{v}'$ is a field that we think of as a vector. The gradient of J thus is also a vector, each component being the partial derivative with respect to its component in $\vec{v}'$,

$$\text{gradient of } J \text{ wrt } \vec{v}' = \nabla_{v'}J = \left(\frac{\mathrm{d}J}{\mathrm{d}\vec{v}'}\right)^T = \begin{pmatrix} \partial J/\partial v'_1 \\ \partial J/\partial v'_2 \\ \cdots \\ \partial J/\partial v'_n \end{pmatrix}. \qquad (1.7)$$

The gradient of J is expressed fully as,

$$\nabla_{v'}J = -(\vec{v}'_B - \vec{v}') - \mathbf{U}^T\mathbf{H}^T(\mathbf{E} + \mathbf{F})^{-1}(\vec{y} - \vec{y}_{mo}) \qquad (1.8)$$

$$= \nabla_{v'}J_B + \nabla_{v'}J_o, \qquad (1.9)$$

which follows from differentiating Eq. (1.6) with respect to each element of $\vec{v}'$ individually, or, by first differentiating with respect to $\vec{x}'$, and then using the chain rule [4] to give the gradient in $\vec{v}'$-space. The treatment of the background contribution is trivial in $\vec{v}'$-space as the gradient of J_B is just a difference of vectors. The way that the gradient of J_o is found is a three stage process: (i) by applying the forward model (to calculate $\vec{y}_{mo}$), (ii) differencing with the actual observations and operating with $(\mathbf{E} + \mathbf{F})^{-1}$ as in Eq. (1.8), and (iii) applying the adjoint model to find the gradient. In Eq. (1.8), $\mathbf{H}$ (a matrix) is the linearization of $\vec{H}$ (a vector operator). Since in this report, we are dealing with observations, we will focus entirely on J_o and its gradient.

(i) Forward model state

The sequence of actions that predict the observations from the model state is the forward model. In the Met Office 3d-Var., the forward model is constructed along the following lines.

---

1. Convert the variational model state, $\vec{v}'$ (a perturbation state in a transformed space) to the model state, $\vec{x}'$ (the perturbation state in 'normal' model space) via the **U** operator (Eq. (1.4)).
2. Use interpolation to derive columns of these perturbation quantities at the horizontal positions of the observations, $\vec{C}'_x$.
3. Add to $\vec{C}'_x$ the columns, $\vec{C}_x$, found beforehand from the guess state (in a similar way to $\vec{C}'_x$). The result is $\vec{C}^+_x$.
4. Operate on the $\vec{C}^+_x$ columns to give the model version of the observations, $\vec{y}_{mo}$. These can then be compared with the actual observations, $\vec{y}$.

Steps 2 to 4 are implicit in the $\vec{H}$-operator.

(ii) The residual

When comparing $\vec{y}_m$ with $\vec{y}$, the following is computed (as part of Eq. (1.8)),

$$-(\mathbf{E} + \mathbf{F})^{-1}(\boldsymbol{y} - \boldsymbol{y}_{mo}). \qquad (1.10)$$

This is the gradient of J_o with respect to $\vec{y}_{mo}$, $(\mathrm{d}J_o/\,\mathrm{d}\vec{y}_{mo})^T \equiv \nabla_{y_{mo}} J_o$, which is an adjoint vector.

(iii) Adjoint model state

The gradient of J_o with respect to $\vec{v}'$ is found by acting with a string of adjoint operators, which propagate the adjoint state in reverse order to the forward models (the forward model part corresponding to each adjoint are given in brackets below each operator),

$$\begin{aligned}
\nabla_{v'} J_o &= \left(\frac{\mathrm{d}J_o}{\mathrm{d}\vec{v}'}\right)^T \\
&= \left(\frac{\mathrm{d}\vec{x}'}{\mathrm{d}\vec{v}'}\right)^T \left(\frac{\mathrm{d}\vec{C}'_x}{\mathrm{d}\vec{x}'}\right)^T \left(\frac{\mathrm{d}\vec{C}^+_x}{\mathrm{d}\vec{C}'_x}\right)^T \left(\frac{\mathrm{d}\vec{y}_{mo}}{\mathrm{d}\vec{C}^+_x}\right)^T \left\{-(\mathbf{E} + \mathbf{F})^{-1}(\vec{y} - \vec{y}_{mo})\right\} \qquad (1.11) \\
&= \quad (A) \quad (B) \quad (C) \quad (D) \\
&= \mathbf{U}^T \qquad \mathbf{H}^T \qquad \nabla_{y_{mo}} J_o. \qquad (1.12)
\end{aligned}$$

In the last line we have highlighted the fact that the operator (A) is the adjoint of **U**, and the combined operator (B) - (D) is the adjoint of **H**. (Note that in practice, the model field state $\vec{x}'$ is replaced in the above by a low resolution field -called $\vec{w}'$ in [2]. We will not be concerned with this here.)

In the Met Office Var. scheme, the first three operators, (A) - (C), are performed by the core scheme. When adding new observation operators, we need be concerned only with the gradient of J_o with respect to $\vec{C}_x^+$,

$$\text{Gradient of } J_o \text{ wrt } \vec{C}_x^+ = \nabla_{C_x^+} J_o = \left(\frac{\mathrm{d}\vec{y}_{mo}}{\mathrm{d}\vec{C}_x^+}\right)^T \nabla_{\vec{y}_{mo}} J_o \qquad (1.13)$$

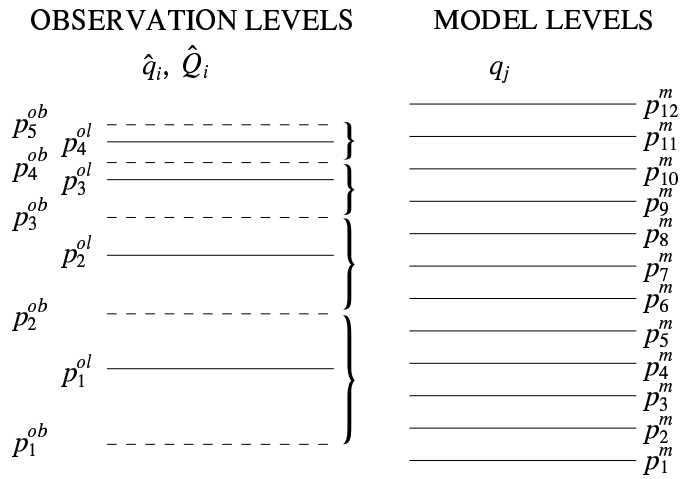
$$= \left(\frac{\mathrm{d}\vec{y}_{mo}}{\mathrm{d}\vec{C}_x^+}\right)^T \left\{-(\mathbf{E} + \mathbf{F})^{-1}(\vec{y} - \vec{y}_{mo})\right\}. \qquad (1.14)$$

The nature of the observations that we are to assimilate

We report here on how we evaluate the forward and adjoint models (up to the stage where the gradient is with respect to $\vec{C}_x^+$, as above) for vertical-profile satellite retrievals of ozone, relative humidity and temperature (sections 2, 3 and 4 respectively). We will also describe the treatment of total column ozone (section 5). In the case of profiles, the observed values are specified on a set of vertical pressure levels. We take into account that values of the retrieved profile on these levels do not represent point values, but instead are more like layer averages.

2. Ozone retrieval assimilation

Ozone is specified as a mass mixing ratio, and so the methodology outlined in this section can be applied for any quantity given in this way. Let the pressure of the ith observation level (on which retrieved ozone, $\hat{Q}_i$, is specified) be p_i^{ol}. These levels are bounded by a set of level boundaries with pressure p_i^{ob} ($p_i^{ob} < p_i^{ol} < p_{i+1}^{ob}$). These define the layers that the specified mixing ratios are layer-averaged within). The model level pressures are p_j^m on which there are ozone mixing ratios q_j. All of these levels and boundaries are shown schematically in the Fig. below.



The forward model for ozone should mass-weight (within each observation layer) the model's mixing ratio as given on model levels. The result is $\hat{q}_i$, the model's version of the observations. We assume that the weight of each

model level j contributing to the measured signal within the layer i is the same (ie that a 'top-hat' weighting function, which is unity within the observation layer and zero elsewhere, is appropriate). This evaluation will involve some interpolation as the observation layer boundaries will not, in general, lie on the model levels.

Given that the mass mixing ratio of ozone between two model pressure levels (found from the arithmetic mean of the mixing ratio at the bounding levels) is $\frac{1}{2}(q_j + q_{j+1})$, then the total amount of ozone in this model layer per unit horizontal area is,

$$\hat{q} = \tfrac{1}{2}\rho(q_j + q_{j+1})\,\delta z, \tag{2.1}$$

where ρ is the density, and δz is the height thickness of the layer. We assume that the shallow atmosphere approximation is valid as we will take the horizontal area of the column to be constant with height. The height thickness, δz, can be found from the pressure thickness, δp, under the hydrostatic approximation,

$$\rho\delta z = -\frac{\delta p}{g}. \tag{2.2}$$

This information allows us to write a 'layer averaging' operator $\mathbf{X}^{om}$ (see below). As this operator is linear, its adjoint, $\mathbf{X}^{omT}$ follows in a straightforward manner (also see below).

The forward model

The information required by the layer averaging operator includes: N_m (the number of model levels), N_{ob} (the number of observation layers), p_i^{ob} (the observation layer boundary pressures - there are $N_{ob} + 1$ of these), p_j^m (the model level pressures) and q_j (the model's mass mixing ratio of the trace substance that we are interested in - in this case ozone). The output is $\hat{q}_i$ (the observation layer averaged mixing ratio). $\hat{q}_i$ is the model version of $\hat{Q}_i$ (the observed value).

It is convenient to describe also an auxiliary structure I_i. This is an index pertaining to observation layer boundary i and references the first model level immediately below it. To use the schematic in the above Fig. as an example, it can be seen that $I_2 = 5$ indicating that model level 5 is immediately below observation layer boundary 2.

The algorithm for $\mathbf{X}^{om}$ is now described.

For each observation layer i (the layer is bounded by pressures p_i^{ob} and p_{i+1}^{ob}):

If $I_{i+1} > I_i$ then

(there is at least one model level in this observation layer)

First sum over any complete model layers

If $I_{i+1} > I_i + 1$ then

(there is at least one model layer within this observation layer),

$\hat{q}_i^A = \sum_{j=I_i+1}^{I_{i+1}-1} (q_j + q_{j+1})(p_{j+1}^m - p_j^m)/2.$

end if-block

Include the contribution from the 'ends' (incomplete model layers)

'Lower' contribution: $\hat{q}_i^B = (q_{I_i} + (\mathrm{d}q/\mathrm{d}p)_{I_i}(p_i^{ob} - p_{I_i}^m) + q_{I_i+1}) \times (p_{I_i+1}^m - p_i^{ob})/2$

'Upper' contribution: $\hat{q}_i^C = (2q_{I_{i+1}} + (\mathrm{d}q/\mathrm{d}p)_{I_{i+1}}(p_{i+1}^{ob} - p_{I_{i+1}}^m)) \times (p_{i+1}^{ob} - p_{I_{i+1}}^m)/2$

Layer average: $\hat{q}_i = (\hat{q}_i^A + \hat{q}_i^B + \hat{q}_i^C)/(p_{i+1}^{ob} - p_i^{ob})$

else

(there are no model levels in this observation layer)

$\hat{q}_i = (2q_{I_i} + (\mathrm{d}q/\mathrm{d}p)_{I_i}(p_i^{ob} - 2p_{I_i}^m + p_{i+1}^{ob}))/2$

end if-block

Importantly, in the above, the gradient $(\mathrm{d}q/\mathrm{d}p)_j$ is calculated on model levels as a forward difference, $(\mathrm{d}q/\mathrm{d}p)_j = (q_{j+1} - q_j)/(p_{j+1}^m - p_j^m)$.

The adjoint model

Since the forward model is linear, it is straightforward (albeit tedious and time consuming) to give the algorithm for the adjoint. In the case of the layer average operator, the adjoint takes us from the gradient (of J_o) with respect to the layer averages, to the gradient with respect to the model columns of mixing ratio (see section 1),

$$\left(\frac{\partial}{\partial \vec{q}}\right)^T J_o = \left(\frac{\mathrm{d}\vec{\hat{q}}}{\mathrm{d}\vec{q}}\right)^T \left(\frac{\partial}{\partial \vec{\hat{q}}}\right)^T J_o. \qquad (2.3)$$

Note that since J_o depends on model variables in other ways, we have partial derivatives. Equation (2.3) is nothing more than the generalized chain rule [4]. In this Eq., the values of q_j have been assembled into the model vector $\vec{q}$ (akin to a subset of $\vec{C}_x^+$ introduced in section 1 of this report - as in Eq. (1.13)), and the layer average quantities $\hat{q}_i$ into the vector $\vec{\hat{q}}$ (akin to a subset of $\vec{y}_{mo}$). In order to construct the adjoint, it is useful to first imagine the forward operator for $\hat{q}_i$ in differential form written as a linear combination of the model values, q_j,

$$\hat{q}_i = \sum_j X_{ij}^{om} q_j \qquad \text{where} \qquad X_{ij}^{om} = \frac{\partial \hat{q}_i}{\partial q_j}. \qquad (2.4), (2.5)$$

The coefficients X_{ij}^{om}, which can be derived by considering the forward algorithm, form an $N_{ob} \times N_m$ matrix and the above left-hand Eq. can be written

in matrix form, $\vec{\hat{q}} = \mathbf{X}^{om}\vec{q}$. With the coefficients known, the adjoint can be constructed. From first principles, this is done via the chain rule component-by-component,

$$\frac{\partial}{\partial q_j} = \sum_i \frac{\partial \hat{q}_i}{\partial q_j}\frac{\partial}{\partial \hat{q}_i} = \sum_i X_{ij}^{om}\frac{\partial}{\partial \hat{q}_i} = \sum_i (X^{om})_{ji}^{T}\frac{\partial}{\partial \hat{q}_i}. \qquad (2.6)$$

In matrix form this is (as in Eq. (2.3)),

$$\left(\frac{\partial}{\partial \vec{q}}\right)^T = \mathbf{X}^{omT}\left(\frac{\partial}{\partial \vec{\hat{q}}}\right)^T. \qquad (2.7)$$

A simple strategy to construct the adjoint is the following. First note that a matrix element, X_{ij}^{om}, appearing in the forward model transfers information from model level j to observation layer i, as $\delta\hat{q}_i = X_{ij}^{om}q_j$. For each such element, the adjoint spreads information the other way such that $\delta(\partial/\partial q_j) = X_{ij}^{om}(\partial/\partial\hat{q}_i)$ (of course the transpose instruction is absent in the last Eq. as it is not in matrix form). Repeating this last action over all non-zero matrix elements achieves the adjoint operation.

The adjoint algorithm follows from this procedure. Note first that when writing the forward model as a linear combination, one should expand the derivatives $\mathrm{d}q/\mathrm{d}p$ that are present in the forward model algorithm. This yields more terms than is first apparent. In the following, each adjoint contribution is accompanied (in red and within curly brackets) by the forward contribution on which it is based.

For each observation layer i (the layer is bounded by pressures p_i^{ob} and p_{i+1}^{ob}):

If $I_{i+1} > I_i$ then

(there is at least one model level in this observation layer)

First deal with any complete model layers

If $I_{i+1} > I_i + 1$ then

For each integer (j) between $I_i + 1$ and $I_{i+1} - 1$

Let $f_1 = (p_{j+1}^m - p_j^m)/2(p_{i+1}^{ob} - p_i^{ob})$

$\delta\nabla_{q,j}J_o = f_1\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = f_1 q_j\}$

$\delta\nabla_{q,j+1}J_o = f_1\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = f_1 q_{j+1}\}$

end loop

end if-block

Include the contribution from the 'ends' (incomplete model layers)

The 'lower' contribution:

Let $f_1 = (p_{I_i+1}^m - p_i^{ob})/2(p_{i+1}^{ob} - p_i^{ob})$, $f_2 = (p_i^{ob} - p_{I_i}^m)/(p_{I_i+1}^m - p_{I_i}^m)$,

$f_3 = f_1 f_2$

$\delta\nabla_{q,I_i}J_o = (f_1 - f_3)\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = (f_1 - f_3)q_{I_i}\}$

$\delta\nabla_{q,I_i+1}J_o = (f_1 + f_3)\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = (f_1 + f_3)q_{I_i+1}\}$

The 'upper' contribution:

Let $f_1 = (p^{ob}_{i+1} - p^m_{I_{i+1}})/2\,(p^{ob}_{i+1} - p^{ob}_i)$, $f_2 = (p^{ob}_{i+1} - p^m_{I_{i+1}})\,/\,(p^m_{I_{i+1}+1} - p^m_{I_{i+1}})$,

$f_3 = f_1 f_2$

$\delta\nabla_{q,I_i+1}J_o = (2f_1 - f_3)\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = (2f_1 - f_3)q_{I_i+1}\}$

$\delta\nabla_{q,I_i+1+1}J_o = f_3\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = f_3 q_{I_i+1+1}\}$

else

(there are no model levels in this observation layer)

Let $f_1 = (p^{ob}_i + p^{ob}_{i+1} - 2p^m_{I_i})\,/\,2\,(p^m_{I_i+1} - p^m_{I_i})$

$\delta\nabla_{q,I_i}J_o = (1 - f_1)\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = (1 - f_1)q_{I_i}\}$

$\delta\nabla_{q,I_i+1}J_o = f_1\nabla_{\hat{q},i}J_o \quad \{\delta\hat{q}_i = f_1 q_{I_i+1}\}$

end if-block

In this algorithm, $\nabla_{q_j}J_o$ is the derivative of J_o with respect to the jth component of model ozone. Similarly, $\nabla_{\hat{q},i}J_o$ is the derivative of J_o with respect to the ith component of the model-observation (layer averaged) ozone.

3. Relative humidity retrieval assimilation

Relative humidity is not a mixing ratio, and so it is meaningless to layer average this quantity. Instead, we must convert the model values to something that does resemble a mixing ratio.

The forward model

The forward model is comprised of the algorithm outlined in the following list. Note that the model variables available are θ^m_j and RH^m_j (potential temperature and relative humidity respectively, each on model level j). These are actually contained within the vector $\vec{C}^+_x$. The model version of the relative humidity observations are, for observation layer i, RH^o_i, and form part of the vector $\vec{y}_{mo}$. Here is the forward model:

1. Convert θ^m_j to temperature, T^m_j by multiplying by the exner pressure,

$$T^m_j = \left(\frac{p^m_j}{p_0}\right)^{\kappa}\theta^m_j, \qquad (3.1)$$

(κ is the thermodynamic constant of 0.286, and p_0 is the reference pressure of 100 hPa).

2. Calculate the saturation water vapour partial pressure, $e^m_{sat,j}$. The partial pressure is a function of temperature only and in the Met Office, values are found by reference to look-up tables (the derivative, $\mathrm{d}\,e^m_{sat,j}\,/\,\mathrm{d}\,T^m_j$, which is required for the adjoint calculation later, can also be returned at this stage).

3. From $e^m_{sat,j}$, calculate the saturated specific humidity, $q^m_{sat,j}$,

$$q^m_{sat,j} = \frac{\varepsilon e^m_{sat,j}}{p^m_j}, \tag{3.2}$$

(ε is the ratio of molecular weights of water vapour to dry air, $\varepsilon = 0.622$).

4. Compute the actual specific humidity,

$$q^m_j = RH^m_j \, q^m_{sat,j}, \tag{3.3}$$

(the relative humidity in the Met Office scheme is between 0 and 1).

5. The specific humidities from step 4 resemble mixing ratios and so we can apply the layer averaging as for ozone (the $\mathbf{X}^{om}$ operator in section 2). The results are on observation layers, $q^{mo}_{sat,i}$ and q^{mo}_i.

6. Convert the layer-averaged specific humidities back to partial pressures,

$$e^{mo}_{sat,i} = \frac{q^{mo}_{sat,i} p^o_i}{q^{mo}_{sat,i}(1 - \varepsilon) + \varepsilon}, \tag{3.4}$$

$$e^{mo}_i = \frac{q^{mo}_i p^o_i}{q^{mo}_i (1 - \varepsilon) + \varepsilon}. \tag{3.5}$$

7. Calculate the model-observation relative humidity on observation layers,

$$RH^{mo}_i = \frac{e^{mo}_i}{e^{mo}_{sat,i}}. \tag{3.6}$$

Note that there are differences in (i) the definition of relative humidity used in the calculation on model levels (step 4) and in observation layers (step 7) and (ii) the relationship between specific humidity and water vapour partial pressure in the same context (steps 3 and 6 respectively). These reflect the different definitions/approximations used within the Met Office Var. scheme and within the observation processing stages of the assimilation.

The adjoint model

Since the calculation of layer average relative humidity is a complicated multi-stage process, we show here the principle of the adjoint calculation. The forward process is non-linear and so it is helpful to first show the corresponding linearized model (or perturbation forecast model). This is most simply done via the chain rule, noting that an increment in RH^{mo}_i depends ultimately on increments of RH^m_j and θ^m_j,

$$RH^{mo}_i = \frac{\partial RH^{mo}_i}{\partial e^{mo}_i} \frac{\partial e^{mo}_i}{\partial qm^o_i} \sum_j \frac{\partial q^{mo}_i}{\partial q^m_j} \left(\frac{\partial q^m_j}{\partial RH^m_j} \delta RH^m_j + \frac{\partial q^m_j}{\partial q^{\,m}_{sat,j}} \frac{\partial q^{\,m}_{sat,j}}{\partial e^m_{sat,j}} \frac{\partial e^m_{sat,j}}{\partial T^m_j} \frac{\partial T^m_j}{\partial \theta^m_j} \delta \theta^m_j \right)$$

$$+\frac{\partial RH_i^{mo}}{\partial e_{sat,i}^{mo}}\frac{\partial e_{sat,i}^{mo}}{\partial q_{sat,i}^{mo}}\sum_j \frac{\partial q_{sat,i}^{mo}}{\partial q_{sat,j}^{m}}\frac{\partial q_{sat,j}^{m}}{\partial e_{sat,j}^{m}}\frac{\partial e_{sat,j}^{m}}{\partial T_j^m}\frac{\partial T_j^m}{\partial \theta_j^m}\delta\theta_j^m. \qquad (3.7)$$

This expression has been derived systematically from the forward model, starting at step 7 and working backwards and the sum is over model levels. We prefer to refactorize the expression with the θ dependencies together,

$$\delta RH_i^{mo} = \frac{\partial RH_i^{mo}}{\partial e_i^{mo}}\frac{\partial e_i^{mo}}{\partial q_i^{mo}}\sum_j \frac{\partial q_i^{mo}}{\partial q_j^m}\frac{\partial q_j^m}{\partial RH_j^m}\delta RH_j^m + \sum_j\left(\frac{\partial RH_i^{mo}}{\partial e_i^{mo}}\frac{\partial e_i^{mo}}{\partial q_i^{mo}}\frac{\partial q_i^{mo}}{\partial q_j^m}\frac{\partial q_j^m}{\partial q_{sat,j}^m} + \frac{\partial RH_i^{mo}}{\partial e_{sat,i}^{mo}}\frac{\partial e_{sat,i}^{mo}}{\partial q_{sat,i}^{mo}}\frac{\partial q_{sat,i}^{mo}}{\partial q_{sat,j}^m}\right) \times \frac{\partial q_{sat,j}^m}{\partial e_{sat,j}^m}\frac{\partial e_{sat,j}^m}{\partial T_j^m}\frac{\partial T_j^m}{\partial \theta_j^m}\delta\theta_j^m, \qquad (3.8)$$

which will make the adjoint more straightforward. We note the following formulae for some of the partial derivatives,

$$\frac{\partial RH_i^{mo}}{\partial e_i^{mo}} = \frac{1}{e_{sat,i}^{mo}}, \qquad (3.9)$$

$$\frac{\partial e_i^{mo}}{\partial q_i^{mo}} = \frac{\varepsilon p_i^o}{(q_i^{mo}(1-\varepsilon)+\varepsilon)^2}, \qquad (3.10)$$

$$\frac{\partial q_j^m}{\partial RH_j^m} = q_{sat,j}^m, \qquad (3.11)$$

$$\frac{\partial q_j^m}{\partial q_{sat,j}^m} = RH_j^m, \qquad (3.12)$$

$$\frac{\partial RH_i^{mo}}{\partial e_{sat,i}^{mo}} = -\frac{e_i^{mo}}{e_{sat,i}^{mo\,2}}, \qquad (3.13)$$

$$\frac{\partial e_{sat,i}^{mo}}{\partial q_{sat,i}^{mo}} = \frac{\varepsilon p_i^o}{(q_{sat,i}^{mo}(1-\varepsilon)+\varepsilon)^2}, \qquad (3.14)$$

$$\frac{\partial q_{sat,j}^m}{\partial e_{sat,j}^m} = \frac{\varepsilon}{p_j^m}, \qquad (3.15)$$

$$\frac{\partial e_{sat,j}^m}{\partial T_j^m} = \text{(from look-up tables)}, \qquad (3.16)$$

$$\frac{\partial T_j^m}{\partial \theta_j^m} = \left(\frac{p_j^m}{p_0}\right)^\kappa. \qquad (3.17)$$

Putting these terms into the series yields,

$$\delta RH_i^{mo} = \frac{1}{e_{sat,i}^{mo}}\frac{\varepsilon p_i^o}{(q_i^{mo}(1-\varepsilon)+\varepsilon)^2}\sum_j \frac{\partial q_i^{mo}}{\partial q_j^m}q_{sat,j}^m\delta RH_j^m + \sum_j\left(\frac{1}{e_{sat,i}^{mo}}\frac{\varepsilon p_i^o}{(q_i^{mo}(1-\varepsilon)+\varepsilon)^2}\frac{\partial q_i^{mo}}{\partial q_j^m}RH_j^m - \frac{e_i^{mo}}{e_{sat,i}^{mo\,2}}\frac{\varepsilon p_i^o}{(q_{sat,i}^{mo}(1-\varepsilon)+\varepsilon)^2}\frac{\partial q_{sat,i}^{mo}}{\partial q_{sat,j}^m}\right)\frac{\varepsilon}{p_j^m}\frac{\partial e_{sat,j}^m}{\partial T_j^m}\left(\frac{p_j^m}{p_0}\right)^\kappa\delta\theta_j^m, \qquad (3.18)$$

which we rewrite once more exactly as (using Eqs. (3.4) to (3.6)),

$$RH_i^{mo} = \frac{\varepsilon e_i^{mo}}{p_i^o}\left[\frac{RH_i^{mo}}{q_i^{mo2}}\sum_j \frac{\partial q_i^{mo}}{\partial q_j^m}q_{sat,j}^m\delta RH_j^m + \sum_j\left(\frac{RH_i^{mo}}{q_i^{mo2}}\frac{\partial q_i^{mo}}{\partial q_j^m}RH_j^m - \frac{1}{q_{sat,i}^{mo\,2}}\frac{\partial q_{sat,i}^{mo}}{\partial q_{sat,j}^m}\right)\times\frac{\varepsilon}{p_j^m}\frac{\partial e_{sat,j}^m}{\partial T_j^m}\left(\frac{p_j^m}{p_0}\right)^\kappa\delta\theta_j^m\right]. \qquad (3.19)$$

In vector and matrix notation, the above expression is written as ($\forall i$),

$$\delta\vec{RH}^{mo} = \mathbf{A}^o\left[\mathbf{B}^o\mathbf{X}^{om}\mathbf{C}^m\delta\vec{RH}^m + \left(\mathbf{B}^o\mathbf{X}^{om}\mathbf{D}^m - \mathbf{E}^o\mathbf{X}^{om}\right)\mathbf{F}^m\delta\vec{\theta}^m\right],$$

where $\mathbf{A}^o$, $\mathbf{B}^o$ and $\mathbf{E}^o$ are diagonal operators in observation layer space, $\mathbf{C}^m$, $\mathbf{D}^m$ and $\mathbf{F}^m$ are diagonal operators in model level space, and $\mathbf{X}^{om}$ is the non-diagonal layer averaging operator. This transforms from model level to observational layer space, and implies the sum over levels. The diagonal operators are defined as the following,

$$\mathbf{A}^o = \frac{\varepsilon e_i^{mo}}{p_i^o}, \qquad (3.20)$$

$$\mathbf{B}^o = \frac{RH_i^{mo}}{q_i^{mo2}}, \qquad (3.21)$$

$$\mathbf{C}^m = q^m_{sat,j}, \qquad (3.22)$$

$$\mathbf{D}^m = RH_j^m, \qquad (3.23)$$

$$\mathbf{E}^o = \frac{1}{q^{mo}_{sat,i}{}^2}, \qquad (3.24)$$

$$\mathbf{F}^m = \frac{\varepsilon}{p_j^m}\frac{\partial e^m_{sat,j}}{\partial T_j^m}\left(\frac{p_j^m}{p_0}\right)^{\kappa}. \qquad (3.25)$$

Once written in this way, it is straightforward to write the adjoint expression, which propagates gradient information from observation to model variables,

$$\left(\frac{\partial}{\partial \vec{RH}^m}\right)^T = \left(\frac{\partial \vec{RH}^{mo}}{\partial \vec{RH}^m}\right)^T \left(\frac{\partial}{\partial \vec{RH}^{mo}}\right)^T = \mathbf{C}^m \mathbf{X}^{om^T} \mathbf{B}^o \mathbf{A}^o \left(\frac{\partial}{\partial \vec{RH}^{mo}}\right)^T, \qquad (3.26)$$

$$\left(\frac{\partial}{\partial \vec{\theta}^m}\right)^T = \left(\frac{\partial \vec{RH}^{mo}}{\partial \vec{\theta}^m}\right)^T \left(\frac{\partial}{\partial \vec{RH}^{mo}}\right)^T = \mathbf{F}^m \left(\mathbf{D}^m \mathbf{X}^{om^T} \mathbf{B}^o - \mathbf{X}^{om^T} \mathbf{E}^o\right) \mathbf{A}^o \left(\frac{\partial}{\partial \vec{RH}^{mo}}\right)^T. \qquad (3.27)$$

Note that diagonal operators are self adjoint, and so the transpose superscript has been omitted for these.

4. Temperature retrieval assimilation

Special relations apply to the assimilation of temperature data. We have at our disposal the hydrostatic approximation, $\partial p / \partial z = -\rho g$, and the Eq. of state, $p = \rho RT$. These can be combined to give an expression for the layer-mean temperature,

$$\hat{T} = \frac{g\delta z / c_p}{\kappa \delta \ln p}, \qquad (4.1)$$

where δz is the layer thickness in height, $\delta \ln p$ is the layer thickness in log-pressure, c_p is the specific heat capacity of air at constant pressure, and κ is the thermodynamic constant of 0.286. This is the "hypsometric" relation.

Routines to perform the layer average in temperature pre-exist in the Met Office Var. scheme and so the forward model routine and its adjoint will not be documented further here.

5. Total column ozone assimilation

The forward and adjoint models for total column ozone measurements are a subset of the layer averaging operators (section 2) in the limit of a single layer spanning the entire thickness of the model atmosphere. The implementation described here however is separate.

The forward model

The calculation of total column ozone has two contributions. The first is the bulk contribution from all model levels and the second is from the surface layer. Surface ozone is not a model variable and so we extrapolate linearly (in log-pressure) to the surface.

$$TCO_3^{bulk} = \sum_{j=1}^{N_m-1} (q_j + q_{j+1})(p_j^m - p_{j+1}^m)/2g$$

$$TCO_3^{surf} = \left\{(q_2 - q_1)\ln(p_*^m/p_1^m)/\ln(p_2^m/p_1^m) + 2q_1\right\}(p_*^m - p_1^m)/2g$$

$$TCO_3 = TCO_3^{bulk} + TCO_3^{surf}$$

In this algorithm, p_*^m is the surface pressure (this is known as it is a model variable). For total column ozone, the surface contribution to the total column will probably be negligible, but it is left in the forward model algorithm to make the code extendable to other species that may be present here at higher concentrations.

The adjoint model

The forward model is linear and is, in effect, a $1 \times N_m$ matrix. The adjoint is a $N_m \times 1$ matrix that propagates the gradient information from that with respect to a single model measurement to that with respect to a model ozone profile. As always it is helpful to derive the adjoint model from the perturbation forecast model found beforehand. For this reason each line of the the adjoint algorithm (below) is accompanied by its corresponding line in the perturbation forecast model (in red and inside curly brackets).

For each integer (j) between 3 and $N_m - 1$

$$\nabla_{q,j}J_o = (p_{j-1}^m - p_{j+1}^m)/2g\nabla_{TCO_3}J_o \quad \{\delta TCO_3 = q_j(p_{j-1}^m - p_{j+1}^m)/2g\}$$

end loop

Contribution of the top ozone level

$$\nabla_{q,N_m}J_o = (p_{N_m-1}^m - p_{N_m}^m)/2g\nabla_{TCO_3}J_o \quad \{\delta TCO_3 = q_{N_m}(p_{N_m-1}^m - p_{N_m}^m)/2g\}$$

Contribution of ozone level 2

$$\nabla_{q,2}J_o = \{p_1^m - p_3^m + (p_*^m - p_1^m)(\ln p_*^m/p_1^m)/(\ln p_2^m/p_1^m)\}/2g\nabla_{TCO_3}J_o$$

$$\{\delta TCO_3 = q_2\{p_1^m - p_3^m + (p_*^m - p_1^m)(\ln p_*^m/p_1^m)/(\ln p_2^m/p_1^m)\}/2g\}$$

Contribution of ozone level 1

$$\nabla_{q,1}J_o = \{(p_*^m - p_1^m)(1 - (\ln p_*^m/p_1^m)/(\ln p_2^m/p_1^m)) + p_*^m - p_2^m\}/2g\nabla_{TCO_3}J_o$$

$$\{\delta TCO_3 = q_1\{(p_*^m - p_1^m)(1 - (\ln p_*^m/p_1^m)/(\ln p_2^m/p_1^m)) + p_*^m - p_2^m\}/2g$$

The derivative notation is as the adjoint algorithm for layer mean quantities, additionally with $\nabla_{TCO_3}J_o = \partial J_o/\partial(TCO_3)$.

We make a special note about total column quantities. Total column 'observations' are not direct measurements, but are highly processed from satellite measurements. Despite their name, they are not exactly the total column

amounts of the substance in question (e.g. ozone), but are a *weighted* total column (weighted with some averaging kernel) [5]. Although this is strictly the case, we assume in the above algorithm that the averaging kernel is unity.

6. Routine names

Name	Section	Module directory	Author
Var_LayerAv	2	VarMod_MLS	DARC (RNB)
Var_LayerAv_Adj	2	VarMod_MLS	DARC (RNB)
Var_RHLayerAv	3	VarMod_MLS	DARC (RNB)
Var_RHLayerAv_Adj	3	VarMod_MLS	DARC (RNB)
Gen_LayerTemperature	4	GenMod_Utilities	Met Office
Gen_LayerTemperature_Adj	4	GenMod_Utilities	Met Office
Var_CalcModelTCO3	5	VarMod_GOME	DARC (RNB)
Var_CalcModelTCO3_Adj	5	VarMod_GOME	DARC (RNB)

7. References

[1] R. Daley, Atmospheric data analysis, Cambridge University Press (1996).

[2] A.C. Lorenc, S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne, F.W. Saunders, The Met Office 3-dimensional variational data assimilation scheme, Q.J.R. Meteorol. Soc. 126, pp. 2991-3012 (2000).

[3] D.F. Parrish, J.C. Derber, The National Meteorological Center's spectral statistical interpolation analysis system, Mon. Wea. Rev. 120, pp. 1747-1763 (1992).

[4] R.N. Bannister, Applications of the chain rule, http://www.met.rdg.ac.uk/~ross/Documents/Chain.html (2001).

[5] C.D. Rodgers, Inverse methods for atmospheric sounding: theory and practice, Series on atmospheric, oceanic and planetary physics, Vol. 2, World Scientific (2000).