MSc in Data Assimilation, Exam Question for 3d-Var (2007)

Ross Bannister, Dept. of Meteorology

A common way of assimilating meteorological data is by the method of 'three-dimensional variational data assimilation' (or 3d-Var for short), which is a means of estimating the state vector given some observations.

(a) With reference to the way that 3d-Var. works, why is it called a variational method? [1 mark]

(b) A background state is also needed to solve the data assimilation problem.
(i) What is the background state? [1 mark]
(ii) In an operational setting, how is the background state calculated? [1 mark]
(iii) Give a reason why is the background state needed in variational data assimilation? [1 mark]

(c) A 3d-Var. cost function, J, has the following form in terms of the usual symbols

$$J = \tfrac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \tfrac{1}{2}(\vec{y} - \vec{h}[\vec{x}])^T \mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}]).$$

This method is to be used to assimilate one observation. The domain is a Cartesian grid with $\nu = N_x \times N_y \times N_z$ points and each point carries two variables - temperature, $T(i, j, k)$, and pressure, $p(i, j, k)$. The observation is in the form of a quantity called *potential temperature* (θ), which is related to pressure and temperature by the following formula

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

where p_0 and κ are positive constants. The observation is labelled y and is coincident with the grid point at position $(\tilde{i}, \tilde{j}, \tilde{k})$. The state vector has the following structure

$$\vec{x} = (x_1, \cdots, x_q, \cdots, x_r, \cdots x_n),$$

where q is the index labelling the position in the state vector representing $T(\tilde{i}, \tilde{j}, \tilde{k})$ and r is the index representing $p(\tilde{i}, \tilde{j}, \tilde{k})$.

(i) The state vector has n components. How large is n in this example? [1 mark]
(ii) What is the observation operator written in terms of model quantities? [1 mark]
(iii) Is this operator linear or non-linear? [1 mark]
(iv) Write down the Jacobian matrix of this observation operator. [3 marks]

(d) The Best Linear Unbiased Estimator (BLUE) formula gives an estimate for the analysis increment that 3d-Var. produces. From the general BLUE formula, write down an expression that gives the analysis increment of T and of p at the observation point $(\tilde{i}, \tilde{j}, \tilde{k})$. Let the $\mathbf{B}$ and $\mathbf{R}$ matrices have the following forms

$$\mathbf{B}=\begin{pmatrix} B_{11} & \cdots & B_{1q} & \cdots & B_{1r} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \iddots & \vdots \\ B_{q1} & \cdots & B_{qq} & \cdots & B_{qr} & \cdots & B_{qn} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ B_{r1} & \cdots & B_{rq} & \cdots & B_{rr} & \cdots & B_{rn} \\ \vdots & \iddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{nq} & \cdots & B_{nr} & \cdots & B_{nn} \end{pmatrix}, \quad \mathbf{R}=R_{11}.$$

[10 marks]

(e) Given that $\mathbf{B}$ is a non-sparse matrix, explain with the aid of your workings to (d) how the analysis increments at locations away from point $(\tilde{i},\tilde{j},\tilde{k})$ would be affected by the observation. [2 marks]

(f) Based on your answer to (c), would you expect the BLUE result to give exactly the same result as the 3d-Var? [2 marks]

(g) By running a 3d-Var. system to convergence with a single observation whose observation operator is linearly related to the state vector, and by assuming that $\mathbf{B}$ and $\mathbf{R}$ are accurate, what is the expected value of J at the analysis? [1 mark]