Candidates are admitted to the examination room ten minutes before the start of the examination. On admission to the examination room, you are permitted to acquaint yourself with the instructions below and to read the question paper.

Do not write anything until the invigilator informs you that you may start the examination. You will be given five minutes at the end of the examination to complete the front of any answer books used.

April 2011

Answer Book
General Data Sheet
Lecture notes permitted
Any bilingual English language dictionary permitted
Only Casio-fx83 calculators are permitted

THE UNIVERSITY OF READING

MSc Examination for Courses in Sciences

Operational Data assimilation Techniques

MTMD02

2 hours

Answer **ANY TWO** questions

The marks for the individual components of each question are given in [ ] brackets.
The total mark for the paper is 100

1. Consider M direct observations of a tracer $\phi(x, y, z, t)$, where the ith observation, denoted ϕ_i, is made at position (x_i, y_i, z_i) and time t_i. The continuous space and time domain spans $[0, L_x]\times[0, L_y]\times[0, L_z]$ and $[0, T]$ respectively. A set of initial conditions of the state of the system is to be estimated, and an a-priori for these has been provided, $\phi_{\mathrm{b}}(x, y, z)$.

(a) What is meant by the three-dimensional variational data assimilation (3-D VAR) approximation? Write down the assumed model equation for the state $\phi(x, y, z, t)$ under this approximation. [4 marks]

(b) Explain the meanings of the terms *strong* and *weak* constraints applied to data assimilation. [4 marks]

(c) Let the error variance of each observation be W_{ob}^{-1}, the error variance of $\phi_{\mathrm{b}}(x, y, z)$ be the uniform value W_{ic}^{-1} and the error variance to which the model constraint in part (a) is satisfied at each position and time over the domain be W_{e}^{-1}. Write down a functional (with argument $\phi(x, y, z, t)$) that measures the total degree of misfit to the observations, to the initial conditions, and to the (*weakly constrained*) condition from part (a). [8 marks]

(d) The following Euler-Lagrange equations are satisfied

$$W_{\mathrm{ic}}\{\hat{\phi}(x, y, z, 0) - \phi_{\mathrm{b}}(x, y, z)\} = \hat{\mu}(x, y, z, 0),$$

$$W_{\mathrm{ob}}\sum_{i=1}^{M}\{\hat{\phi}(x_i, y_i, z_i, t_i) - \phi_i\}\delta(x - x_i)\delta(y - y_i)\delta(z - z_i)\delta(t - t_i) = \frac{\partial\hat{\mu}(x, y, z, t)}{\partial t},$$

when the following definitions are made

$$\frac{\partial\hat{\phi}(x, y, z, t)}{\partial t} = W_{\mathrm{e}}^{-1}\hat{\mu}(x, y, z, t),$$

$$\hat{\mu}(x, y, z, T) = 0.$$

Question 1 (d) continues overleaf

Question 1 (d) continues

What are $\hat{\phi}(x,y,z,t)$ and $\hat{\mu}(x,y,z,t)$ that appear in the Euler-Lagrange equations?

[4 marks]

(e) Given that the constraint imposed on $\phi(x,y,z,t)$ from part (a) (and hence on $\hat{\phi}(x,y,z,t)$) is made in the weak formulation:

(i) Explain whether the solution $\hat{\phi}(x,y,z,t)$ would vary or not vary in time.

[2 marks]

(ii) What simple change could be made to the above equations to make the constraint strong? [2 marks]

(iii) Which kind of constraint (weak, strong or both) would allow the solution $\hat{\phi}(x,y,z,t)$ to be exactly equivalent to the 3D-VAR solution?

[1 mark]

(iv) Under the strong constraint, would the solution for $\hat{\mu}(x,y,z,t)$ vary or not vary in time? [2 marks]

Question 1 continues overleaf

Question 1 continues

(f) The method of representers is to be adopted to solve the weak constraint Euler-Lagrange equations in part (d).

(i) How many forward representer functions, and how many backward representer functions would need to be computed using the method?

[2 marks]

(ii) Let the i th forward representer be called $r_i(x,y,z,t)$ and the i th backward representer be called $\alpha_i(x,y,z,t)$. Under the method of representers write down the equations that are satisfied by each (with any conditions imposed).

[4 marks]

(iii) Write down the standard form of the solution $\hat{\phi}(x,y,z,t)$ built of the representer functions and a set of (yet unknown) coefficients. [3 marks]

(iv) Use your answers to parts (i-iii) to derive a matrix equation whose solution gives the set of coefficients specified in part (iii). [14 marks]

2.

(a) A common way of assimilating meteorological data is by the method of three-dimensional variational data assimilation (3D-VAR), which is a means of estimating a state vector given some observations.

(i) With reference to the way that 3D-VAR. works, why is it called a variational method? [2 marks]

(ii) A background state is also needed to solve the data assimilation problem. What is the background state? [1 mark]

(iii) In an operational setting, how is the background state calculated? [1 mark]

(iv) Give two reasons why a background state is needed in variational data assimilation? [2 marks]

A 3D-VAR cost function, $J(\mathbf{x})$, has the following form in terms of the usual symbols

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

3D-VAR 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, $p(i, j, k)$, where (i, j, k) label the grid points. 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

$$\vec{x} = (x_1, \cdots, x_q, \cdots, x_r, \cdots x_n)^\mathrm{T} \quad \text{where}\,,$$

Question 2 (a) continues overleaf

Question 2 (a) continues

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

(v) The state vector has n components. How large is n in this example?

[1 mark]

(vi) What is the observation operator written in terms of model quantities?

[1 mark]

(vii) Is the observation operator as defined in (vi) linear or non-linear?

[1 mark]

(viii) Write down the Jacobian matrix for the observation operator defined in (vi).

[4 marks]

(ix) 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 any point in the domain including the observation point $(\tilde{i},\tilde{j},\tilde{k})$. To answer this question, you will assume that the matrices $\mathbf{B}$ and $\mathbf{R}$ have the following forms

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

[12 marks]

Question 2 continues overleaf

Question 2 continues

(b) Consider the discrete scalar model $x_{i+1} = 2x_i + 1$ with random initial condition x_0 that are normally distributed with mean $m_0 = 0$ and variance $\sigma_0^2 = 1$. Assume that the probability density distribution (pdf) of the initial state is approximated with a set of 5 ensemble members, given by: $\mathbf{x}_0 = (0.33, 0.01, -0.25, 0.41, -0.47)$. Now determine the ensemble mean at time t=2 (with unit time step). Compare the ensemble mean and the ensemble variance with the actual mean and variance of the pdf at t=2. Discuss your results in terms of the standard error of the ensemble mean and of the ensemble variance.

[10 marks]

(c) Compare some of the characteristics of ensemble-based and variational data assimilation.

[15 marks]

3. In the following you may use the following result: the inverse of a non-singular 2 x 2 matrix **A** is:

$$\mathbf{A}^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Consider a system described by linear first-order differential equation given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} \qquad (3.1)$$

and assume that the state vector **x** has two components.

(a) Find a temporal discretization of Eq. (3.1) in the case when the model **L** is time-invariant. [Hint: approximate the derivative forward in time using the Euler's forward method]

[5 marks]

(b) Let us now write the difference equation corresponding to Eq. (3.1) as $\mathbf{x}_{i+1} = \mathbf{M}\mathbf{x}_i$. Also assume $\mathbf{L} = \begin{pmatrix} 1 & 1 \\ 4 & -2 \end{pmatrix}$ and that initial condition vector $\mathbf{x}_0$ is a random vector that is normally distributed, with mean $E\{\mathbf{x}_0\} = \mathbf{m}_0 = (2, -3)^T$ and covariance $E\{(\mathbf{x}_0 - \mathbf{m}_0)(\mathbf{x}_0 - \mathbf{m}_0)^T\} = \mathbf{P}_0$, where the diagonal elements of $\mathbf{P}_0$ are 10% of the magnitude of the initial conditions and the correlation between the two components of the state is 0.5. By using the difference equation model, find the best estimate of the state at t=2, for a unit time step Δt.

[10 marks]

(c) Quantify the uncertainty of our estimate of the state at t=2, for a unit time step Δt.

[10 marks]

Question 3 continues overleaf

Question 3 continues

(d) Assume that at time t=2 (for a unit time step Δt) we make a direct measurement $\mathbf{z}$ of the state, that can be written as $\mathbf{z}_2 = \mathbf{H}\mathbf{x}_2 + \boldsymbol{\varepsilon}_2$. Also assume that the observation operator $\mathbf{H}$ is the unit matrix and that the observation error $\boldsymbol{\varepsilon}$ is normally distributed with zero mean and diagonal error covariance, with diagonal elements equal to 1% of the magnitude of $\mathbf{m}_2$.

(i) Find the best estimate of the state at t=2 when $\mathbf{z}_2 = (13.3, -7.2)$

[10 marks]

(ii) Quantify the uncertainty of the estimate. Compare the accuracy of the components of the estimate of the state before and after the measurement and discuss the result.

[5 marks]

(e) Now assume only the first component of the state is observed and find the best estimate of the state at t=2 and its uncertainty. Compare its uncertainty before and after the measurement and discuss the results.

[10 marks]

(End of Question Paper)