

# Assimilation of Observations in the Boundary Layer of the Met Office's NWP Data Assimilation System

William Grey

*Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, UK.*

*Email: william.grey@metoffice.gov.uk*

## Abstract

We explore the role of the physical parameterisations in the Met Office operational NWP 4DVAR linear model in spreading the information from observations in a more realistic manner. In particular we examine the impact of the boundary layer diffusion scheme. The diffusion scheme uses an empirical approach that relies on the discrimination between the different boundary layer conditions. We generate a set of empirical coefficients using several nonlinear model diagnostics as parameterisations for the boundary layer condition by minimising the error between the non-linear and linear increments. To test the scheme, we perform a set of single observation tests for two cases representing stable and unstable meteorological situations. In these experiments analysis increments are created from the insertion of a single observation of moisture and potential temperature. We show that the Unified Model (UM) turbulent mixing height diagnostic leads to the smallest linearisation error, but the spreading of information from the observation by the physics is still not influenced by the prevailing boundary layer conditions.

## 1 Introduction

In the current Met Office operational NWP 4DVAR data assimilation system many observations, including those from satellites, (e.g. cloud-top parameters and temperatures from SEVERI and IASI radiances) are not being correctly assimilated. Poor assimilation of observations typically occur for temperature inversions and stratocumulus cloud cases in the boundary layer. These problem cases have principally been attributed to inappropriate vertical covariances of the background covariance matrix that have a tendency to spread the background information too much (Lorenc, 2007). Here we examine the role of the Met Office 4DVAR linear model in spreading information from observations.

The linear model in the Met Office data assimilation system uses several physical parameterisations, but in this paper we focus on the boundary layer physics. The boundary layer scheme is composed of two sub-steps: a Buizza (1994) type approach that takes account of vertical diffusion and surface drag for momentum, and a second that parameterises the vertical diffusion using an empirical approach that seeks to minimise the difference between the non-linear and linear model increments. The vertical diffusion scheme uses nonlinear diagnostics to discriminate between different boundary layer conditions. This work is motivated by the current Unified Model (UM) NTML (number of turbulent mixing layers) diagnostic not sufficiently discriminating between the different boundary layer situations. Our purpose is to find a suitable alternative to NTML for characterising boundary layer conditions. We test a number of diagnostics that might allow discrimination between the different boundary layer conditions.

This paper is organised in the following way. In Section 2 we briefly outline the boundary layer scheme used in the linear model. Section 3 discusses the alternative diagnostics for parameterising

the diffusion component of the boundary layer scheme and describes the training and testing of the parameters. Section 4 presents the results of single observation experiments that allow us to gain a view of the relative spreading of information by the background covariance matrix and the linear model. Finally in Section 5 we present the key findings and highlight some future research priorities.

## 2 Boundary Layer Scheme

We briefly review the boundary layer scheme used in the Met Office 4DVAR linear model. For a more full account refer to Payne (In press). The linear boundary layer scheme comprises two sub-steps such that  $\psi'_{i-1} \rightarrow \psi'_{s,i} \rightarrow \psi'_i$  where  $\psi$  is a vector representing the geophysical parameters. The subscript  $i$  is the time-step,  $s$  is the sub-step, and the primes refers to the linear model but are dropped for notational convenience hereafter. The first sub-step is a Buizza-type approach that takes account of vertical diffusion and surface drag for momentum. In the Buizza scheme the vertical diffusion of momentum is:

$$\frac{\partial \psi_{i-1}}{\partial t} = \frac{1}{\rho} \frac{\partial F_{\psi_{i-1}}}{\partial z},$$

where  $\psi$  is either zonal  $U$  or meridional  $V$  wind speed,  $\rho$  is air density,  $F_{\psi}$  is the flux of momentum and  $z$  is height. Refer to Buizza (1994) and Payne (2006) for a full discussion of the scheme. Using Eulerian time-stepping we get:

$$\psi_{s,i} = \frac{\partial \psi_{i-1}}{\partial t} + \psi_{i-1}$$

The second sub-step parameterises the vertical diffusion using an empirical approach. The diffusion scheme relies on minimising the linearization error between a pair of non-linear runs, one of which is run from the background and the other from the background with the analysis increment and the linear model. That is we seek to minimise:

$$M' \delta x - (M(x) - M(x + \delta x))$$

where  $x$  is a vector representing the geophysical parameter (e.g. wind, potential temperature, moisture and so on) for every grid point.  $M$  is the non-linear model and the prime denotes the linear model. The diffusion scheme is implemented by finding the parameters that minimise the linearization area for a range of prevailing boundary layer conditions. These parameters are defined by a set of  $K$  matrices such that:

$$\frac{\partial \psi_{s,i}}{\partial t} = K_{\psi,j} \psi_{s,i}$$

The second sub-step is performed and we get:

$$\psi_i = (I + K_{\psi,j} \Delta t) \psi_{s,i}$$

where  $I$  is the identity matrix. The  $K$  matrices are a set of tridiagonal (or pentadiagonal) matrices that differ according to the boundary layer situation  $j$  and the parameter that we are interested in. The dimensions of  $K$  is  $n \times n$  where  $n$  is the number of layers in the boundary layers which in the case of the Met Office 50 Level global and 38 level NAE (North Atlantic and European) LAM (Local Area model) is 13. The  $K$  matrices are empirically derived by minimising the linearisation error. The Unified Model (UM) diagnostic NTML is currently used as the parameter for discriminating between the different BL situations, but here we seek alternative candidates.

In operational implementation, the boundary layer diffusion scheme leads to an improvement in forecast scores over the ocean but the impact is negative over land. Previously we have shown that diffusion contributed only a little to vertical spreading of observations, and that the extent of that spreading did not depend on the prevailing conditions. The spreading was similar under both stable and unstable conditions (Grey, 2009).

### 3 Training and Testing the K Matrices

#### 3.1 Alternative regression parameters

Our purpose is to investigate possible alternatives to NTML as a regression parameter. Previously we have shown that NTML does not discriminate sufficiently well enough between differing boundary layer situations (Grey, 2009). We are seeking alternative regression parameters that are more discriminating between the boundary layer conditions. Candidates that we investigate are:

1. Boundary layer type (Lock *et al.*, 2000)
2. Turbulent kinetic energy (5% of maximum)
3. Boundary layer depth
4. Mixing layer height
5. Wind stress (5% of maximum)
6. Combination of mixing layer height and boundary layer type
7. Land /ocean
8. Globally fixed  $K$

Some of these require further explanation. Boundary layer type refers to the seven boundary layer types defined in Lock *et al.* (2000) that represent a range of stable and unstable and cloudy boundary layers. Turbulent Kinetic Energy  $TKE$  is defined by:

$$TKE = \frac{1}{2}(u'^2 + v'^2 + w'^2)$$

where  $u'$  and  $v'$  refer to horizontal wind speed in the zonal and meridional direction and  $w'$  in the vertical direction. The primes denote deviation from the mean at a given timescale such that:

$$u' = U - u; v' = V - v, w' = W - w$$

where  $U$  is the mean wind speed and  $u$  is the wind speed and is analogous for the other wind components. We set boundary layer height to be 5% of maximum TKE where the maximum is usually in the lowest layer. We calculate the boundary layer height from the wind stress in the same way. The wind stress  $s$  is given by:

$$s = \sqrt{uw^2 + vw^2}$$

where  $uw$  and  $vw$  are the two horizontal wind stress components. We also use a combined set of diagnostics to see if we can find a smaller linearization error compared with using the diagnostics separately. For benchmarking we generate a single set of  $K$  matrices that are fixed globally and, fixed for ocean and land in another case.

### 3.2 Training the $K$ Matrices

In the training phase we seek a set of  $K$  matrices that minimise the linearisation error. The  $K$  matrices are applied to each grid point, but they are not different for every point, instead a set of  $K$  matrices are grouped according to the boundary layer conditions. Ideally we would like to discriminate clearly between the different boundary layer situations. We test several potential diagnostics, in order to seek the best at discriminating between the boundary layer types.

Nine sets of  $K$  matrices were derived using each of the regression parameters described in Section 3.1 and the NTML. The  $K$  matrices were trained over a global N216 with 70 levels case taken on 17 July 2009. We generated a set of non-linear increments generated in the usual way from a pair of non-linear UM runs to T+6. In the training phase we aim to find a set of coefficients for  $K$  that minimises the linearization error between the linear and non-linear increments.

For each regression parameters the number of  $K$  matrices are given in Table 1. In most cases there are nine for land and ocean. It is possible to use more than one case of training and the average the  $K$  matrices from several cases to give us more robust statistics, especially for cases where there are only a small number of profiles contributing to the  $K$  coefficients. Figure 1 shows the penalty here given as some energy norm that seeks to minimise the linearisation error. The norm uses total perturbation energy and moisture. Based on the training arrays we get a penalty lower than the existing NTML regression parameter with other diagnostics. The turbulent mixing height has the lowest penalty. As expected when we used a single fixed global  $K$  we have the highest penalty, although this is still an improvement over not using the diffusion scheme at all.

<b>Regression Parameter</b>	<b>Ocean</b>	<b>Land</b>
NTML	9	9
Boundary layer type	7	7
Turbulent kinetic energy	9	2
Boundary layer depth	9	9
Mixing layer height	9	9
Wind stress (5% max)	9	9
Combination of mixing layer height and boundary layer type	9	9
Land /ocean	1	1
Globally fixed K	0	1

Table 1: Number of  $K$  matrices for each candidate parameters.

### 3.3 Testing the $K$ Matrices

A more robust examination of the  $K$  matrices is to test them on an independent set of cases. Figure 2 shows the summation of linearization error penalty from 8 independent cases taken in July and August 2009. As expected every case leads to a reduction in penalty of approximately 5-7% compared with the using only the Buizza-type scheme without implementation of the vertical

diffusion scheme. Again there are several regression parameters that lead to lower penalties than the NTML. Turbulent mixing height provides the lowest penalty, although there are several other parameters that also reduce the penalty.

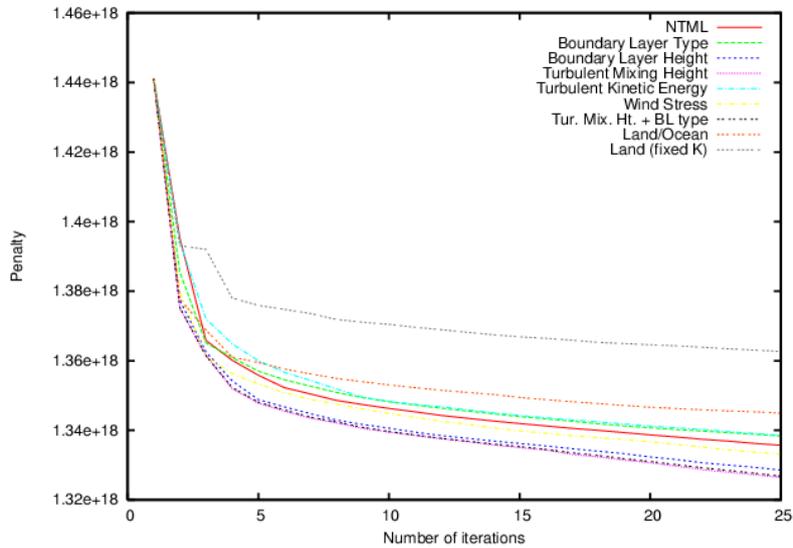


Figure 1: Penalty at each iteration for the regression parameters during the training stage of the K matrices on the 17 July 2009 case.

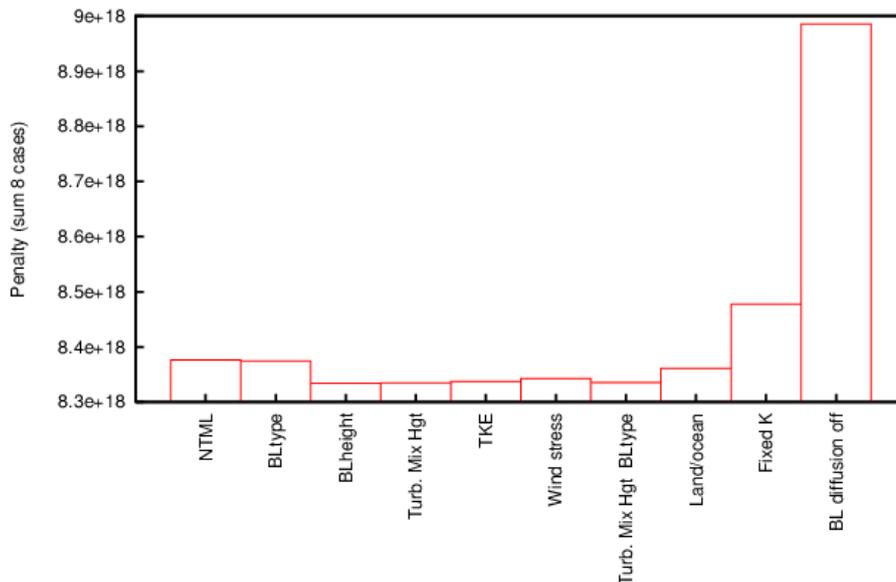


Figure 2: Penalty for the regression parameters K matrices during the testing stage on the global domain.

#### 4 Single Observation Experiments

While a reduction in penalty is certainly desirable, what we wish to see is improvement compared with reality. With this in mind we have performed a set of single observation experiments. We perform a set of single observation tests for cases representing two different meteorological situations so that we can examine the performance of 4DVAR under different scenarios. Two cases are presented here: a conditionally unstable case on 3/8/2009 for Watnall-Nottingham (52.8E, -1.1N) (Figure 3) and a stable case for Trappes near Paris (1.82E, 48.8N) on 16/8/2009 (Figure 4). The analyses were run for the NAE LAM even though the coefficients were generated from the

global model. Ideally we want to use the same model for training and testing, and we will seek to do this in a further update to this work. Analysis increments using 3DVAR and 4DVAR are created from the insertion of a single pseudo observation of moisture and potential temperature in the lowest level towards the end of the assimilation window at these locations. The increments we run for the 9 different sets of  $K$  matrices and without the boundary layer diffusion filtering sub-step. Selected regression parameter values taken from linearization states for Nottingham on 3/8/2009 and Trappes on 16/8/2009 are shown in Table 3.

The analysis increments for the single observation tests are displayed N-S cross-sectional plots at -4.3E, 51.9N. These are presented in Figures 5 and 6. For the 03/08/2009 case we have an unstable boundary layer therefore we would expect some diffusion and thus the information from observations to be spread out more. The diffusion does enhance the spreading of the observations in all cases compared with running without the diffusion scheme. For the stable 16/08/2009 case over Trappes, we would expect to see that there is less spreading. However this is not the case. Clearly even the alternate diagnostics does not allow us to control the spreading according the prevailing boundary layer situations.

For potential temperature we get greater spreading when the analysis increments are generated without the diffusion scheme as compared with the diffusion scheme 4DVAR analysis (see Table 3). Observations inserted toward the end of the time window have greater influence than those at the start because owing to the growth of the background covariances in 4DVAR. Thus, greater relative weighting is apportioned to the observations than to the forecast model at the end of the window than at the beginning of the window. As the physics spreads the information more throughout the window as the model evolves than without the physics, we need to start from a smaller increment in order to fit the observation with higher relative weighting at the end of the window.

	Nottingham: T+0	Nottingham: T+5	Trappes: T+0	Trappes: T+5
BL depth	4-6	4-7	<3	2
NTML	4	7	1	2-3
T mix hgt	4-6	4-7	<3	3
BL type	2/3	2/3	1	2

Table 2: Selected regression parameter values taken from linearization states for Nottingham on 3/8/2009 and Trappes on 16/8/2009.

	3/8/2009 ( $\times 10^{-5}$ )	16/8/2009 ( $\times 10^{-5}$ )
3DVAR	1.6626	-
NTML	1.4808	3.0611
No diffusion	1.3670	2.1606
Boundary layer height	1.6074	3.0342
Turbulent mixing height	1.5861	2.9411
Turbulent kinetic energy	1.5549	2.6360
Wind stress	1.6332	3.1456
Land/ocean	1.5606	-
Globally fixed K	1.7524	2.5971
Boundary layer type	-	3.0661
Combination of Boundary layer type and turbulent mixing height	-	2.8526

Table 3: Mean value of analysis increment for moisture for each of the regression parameters.

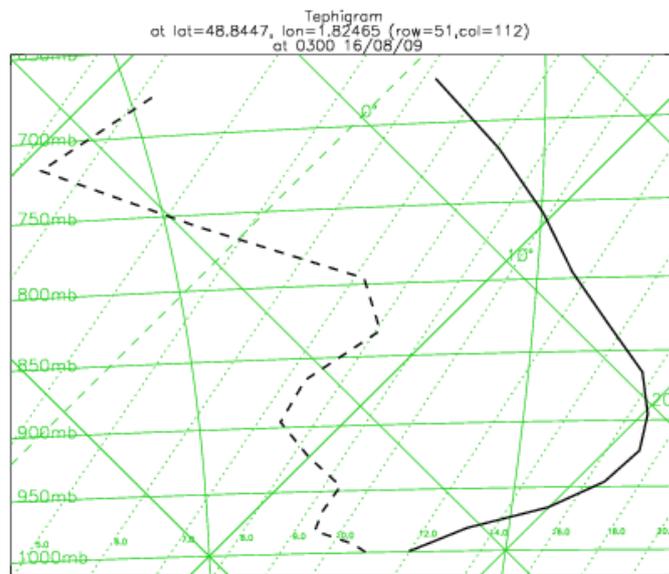


Figure 3: Tephigram from UM NAE LAM at 3Z on 16/08/2009 at 1.82E, 48.8N corresponding to Trappes near Paris for the lowest 13 levels.

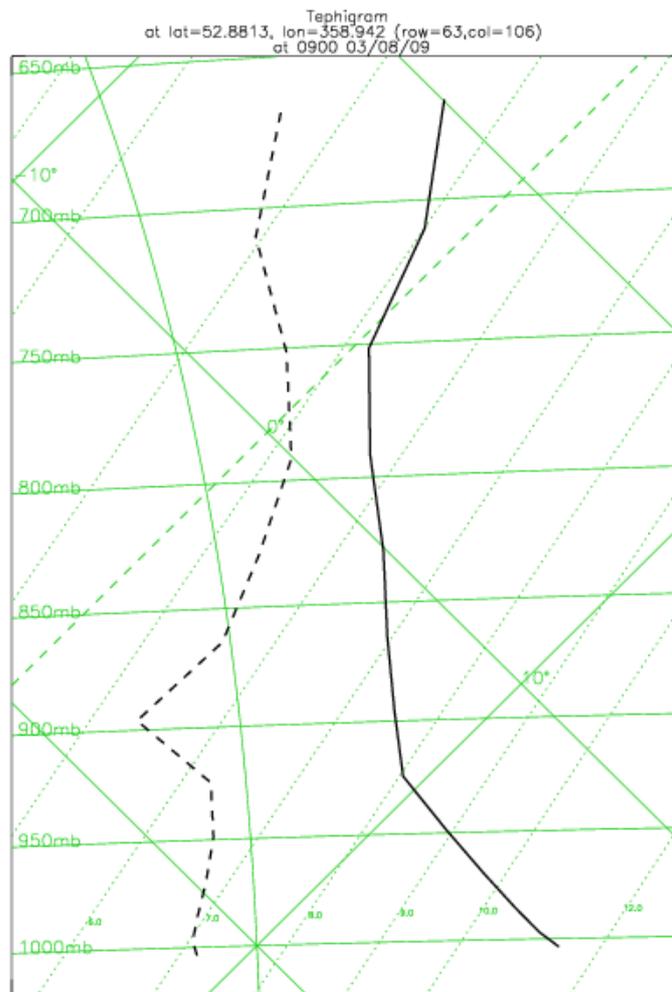


Figure 4: Tephigram from UM NAE LAM at 3Z on 3/08/2009 at 52.8E, -1.1N corresponding to Nottingham-Watnall for the lowest 13 levels.

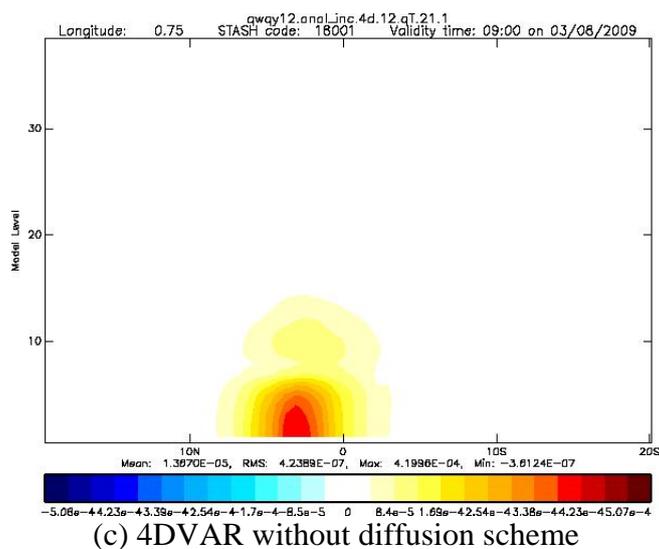
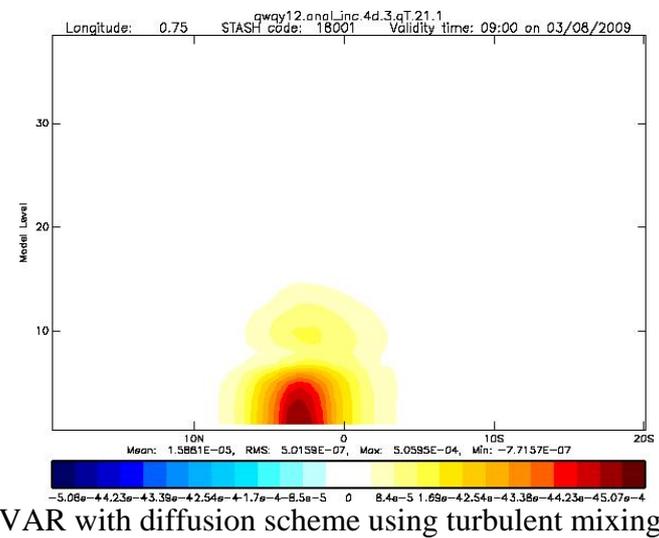
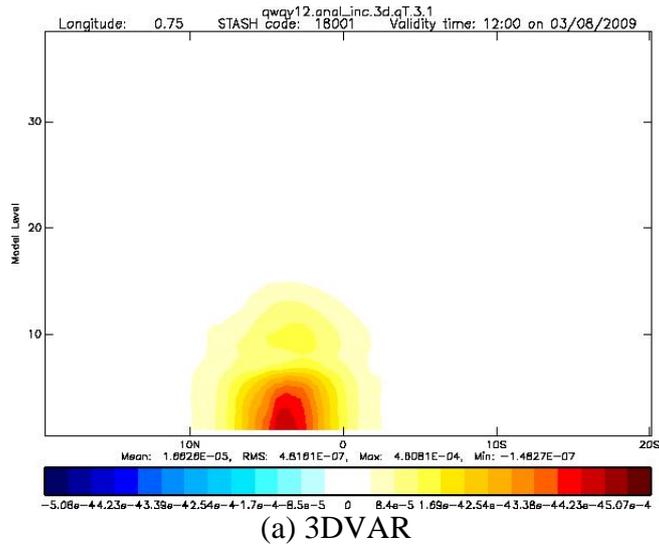


Figure 5: N-S cross-sectional plots of analysis increments for single observation tests of moisture showing moisture for the NAE analysis increments on 3/8/2009 at -4.3E, 51.9N.

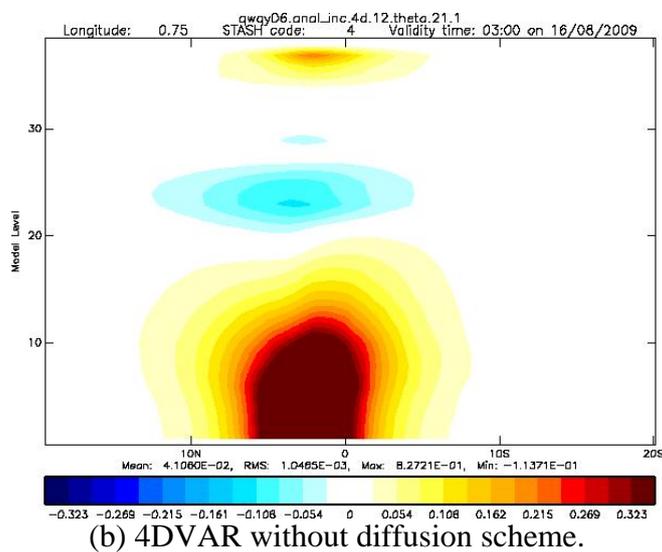
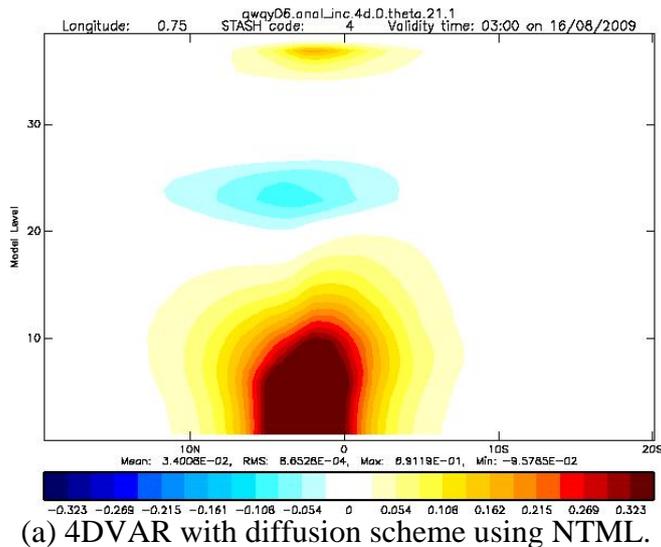


Figure 6: N-S cross-sectional plots of analysis increments for single observation tests of potential temperature showing potential temperature for the NAE analysis increments on 16/8/2009 at -4.3E, 51.9N.

## 5 Conclusions

There are several potential candidates for alternative regression parameters that discriminate between different boundary layer situations that are worth trialling. Perhaps the most suitable of these is turbulent mixing height. Diagnostics from the analysis increments for moisture show that the linear vertical diffusion scheme is spreading the information from the observation. Also we can get smaller analysis increments when diffusion is included for single observations inserted at the end of the window compared with 3DVAR and 4DVAR with diffusion.

Currently the nonlinear model diagnostic on which the operational suite is based is NTML. However, this diagnostic does not clearly distinguish between the different boundary layer conditions. As a consequence observations tend to be spread in the same manner regardless of the boundary layer conditions. However, other diagnostics that may be better at distinguishing

between the boundary layer conditions also do not make the distinction of spreading between stable and unstable boundary layer situations.

The priority of further work is to improve the covariance statistics before we can make further progress on the development of the physical parameterisations in the linear model. In addition, we need to get the linear model vertical diffusion scheme correctly characterising unstable and stable boundary layer conditions. We also need to better understand the reasons for the reduction in the analysis increment for 4DVAR over 3DVAR and whether or not this is desirable behaviour. We will get the diffusion scheme working in the operational 70 level Global, NAE and UKV models.

## **Acknowledgements**

The author would like to thank the following Met Office colleagues: Mike Cullen, Tim Payne, Marek Wlasak, Chiara Piccolo and Adrian Lock for their help.

## **References**

Buizza, R. (1994) Sensitivity of Optimal Unstable Structures, *Quarterly Journal of the Royal Meteorological Society*, 120, 429-451.

Grey, W. (2009) Diagnostics for the assimilation of observations in the boundary layer in 4DVAR, ECMWF Workshop on diagnostics of data assimilation system performance.

Lock, A.P., Brown, A.R., Bush, M.R., Martin, G.M. and Smith, R.N.B. (2000) A new boundary layer scheme for the Unified Model. Part I: Scheme description and single-column model tests, *Monthly Weather Reviews*, 128, 3187-3199.

Lorenc, A.C. (2007) A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers, *Met Office Forecasting Research Technical Report*, 504.

Payne, T. (In Press) Methodology for the Linear Model in 4D-Var, *Atmospheric Science letters*.

Payne, T. (2006) Surface exchange and vertical diffusion of momentum, heat and moisture in the PF and adjoint models, Var Scientific Documentation Paper, *Met Office*.