School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

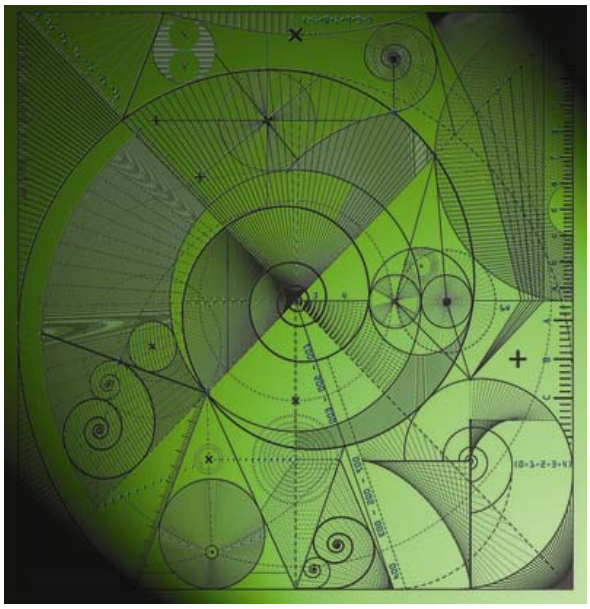
Preprint MPS-2012-19

12 September 2012

Representativity error for temperature and humidity using the Met Office high resolution model

by

J. A. Pocock, S. L. Dance, A. S. Lawless,
N. K. Nichols and J. R. Eyre



Representativity error for temperature and humidity using the Met Office high resolution model

J. A. Pocock[1], S. L. Dance[1], A. S. Lawless[1], N. K. Nichols[1], and J. R. Eyre[2]

[1]School of Mathematical and Physical Sciences, University of Reading, Reading, Berkshire, United Kingdom
[2]Met Office, Fitzroy Road, Exeter, Devon, United Kingdom

September 12, 2012

Abstract

The observation error covariance matrix used in data assimilation contains contributions from instrument errors, representativity errors and errors introduced by the approximated observation operator. Forward model errors arise when the observation operator does not correctly model the observations or when observations can resolve spatial scales that the model cannot. Previous work to estimate the observation error matrix for particular observing instruments has shown that it contains significant correlations. In particular correlations for humidity data are more significant than those for temperature. However it is not known what proportion of these correlations can be attributed to the representativity errors. In this paper we calculate errors of representativity for temperature and humidity using data from the Met Office high resolution UK variable resolution model. Our results show that errors of representativity are correlated and more significant for specific humidity than temperature. We also find that representativity error varies with height. This suggests that they may improve the analysis if explicitly included in a data assimilation scheme.

1 Introduction

In data assimilation model states are combined with observations, making use of their associated error statistics. These are included in the assimilation scheme in the background and observation error covariance matrices. The observation error covariance matrix can be split into three components. One contains information on the instrument error, one describes the error in the observation operator, the other contains information on the representativity error, which has been linked to the forward interpolation error [Lorenc, 1986]. The instrument error is determined for specific instruments under a set of test conditions by the instrument manufacturer or from in-orbit calibration data. It may also contain errors introduced by the pre-processing of the data, which can often be much greater than pure instrument noise. Alternatively they can be included as part of the forward model error. The representativity error is the error that arises when the observations resolve spatial scales that the model cannot [Daley, 1993].

Previous work has shown that observation error statistics are correlated for certain observation types [Stewart et al., 2009, 2012b, Bormann et al., 2002, Bormann and Bauer, 2010, Bormann et al.,

2010] and it has been suggested that part of the correlation comes from representativity error rather than the instrument error or errors in the observation operator [Stewart, 2010, Weston, 2011]. Until recently it has been assumed that it is too expensive to include correlated observation error matrices in assimilation schemes and that it is only feasible to use a diagonal observation error covariance matrix. The effect of correlated error is reduced by using techniques such as observation thinning [Lahoz et al., 2010] or superobbing [Daley, 1991], and variance inflation [Hilton et al., 2009, Whitaker et al., 2008]. Calculations are also simplified by assuming that the observations errors are the same at each model level [Dee and Da Silva, 1999]. Efforts are being made to find methods of reducing the cost of using correlated observation error matrices [Stewart et al., 2012a, Stewart, 2010, Healy and White, 2005, Fisher, 2005]. Once these methods are in place it will be important to have accurate estimates of the covariance matrices, as these are required to obtain the optimal estimate from any data assimilation system [Houtekamer and Mitchell, 2005, Stewart et al., 2008]. It is therefore important to understand how to estimate representativity error.

Despite the difficulties in calculating correlated error, there have been some attempts. The Hollingsworth and Lönnberg method [Hollingsworth and Lönnberg, 1986] has been used to calculate the statistics of the innovations. A method proposed by Desroziers et al. [2005] makes use of information from the first guess and analysis departures and yields an approximation to the observation error covariance matrix. Once the innovation statistics or the observation error covariances have been calculated, the background and/or instrument error terms can be subtracted to leave an approximation of forward model error for specific observing instruments. Other methods [Daley, 1993, Liu and Rabier, 2002] assume that observations can be written as a projection of a high resolution model state on to observation space with the representativity error being the difference between this high resolution projection and the model representation of the observation. Many of these approaches yield a static approximation of representativity error, but Janjic and Cohn [2006] show that it is state dependent and correlated in time.

Work has been carried out by Stewart [2010], Stewart et al. [2009, 2012b] and Bormann and Bauer [2010] to calculate estimates of the full observation error covariance matrix. They show that the observation error covariance matrices for observing instruments such as IASI, AMSU-A, HIRS and MHS contain significant correlations. In particular the correlations for the humidity channels are more significant than those for temperature. The calculated matrices contain contributions from both the representativity error and the instrument error. Due to the complex nature of observation error statistics it is not known what portion of the error is representativity error. As humidity fields contain smaller scale features than temperature fields, it is possible that it is the representativity error that contributes to the more significant error correlations.

In this paper we investigate whether the significant correlations are representativity error. We calculate the representativity error for temperature and humidity data over the UK using a method described by Daley [1993] and Liu and Rabier [2002]. We investigate whether representativity error is more significant for humidity than temperature, and whether one approximation of representativity error is suitable for all pressure levels.

In section 2 we describe the method used for calculating representativity error. We then describe the model and available data in section 3. Our experimental design is given in section 4 and we present our results in section 5. Finally we give conclusions in section 6.

2 Representativity Error

2.1 Representativity Error

Forward model error,

$$\boldsymbol{\epsilon}^{\mathbf{R}} = \mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathbf{m}}), \tag{1}$$

is the difference between the noise free observation vector, $\mathbf{y}$, of length p and the mapping of the exact model state vector, $\mathbf{x}^{\mathbf{m}}$, of length N into observation space using the possibly non-linear observation operator $\mathcal{H}$. The noise free observation vector is a theoretical construct that represents an observation measured by a perfect observing instrument, i.e. with no instrument error. It is related to the actual measurement via the equation

$$\mathbf{y}^{\mathbf{o}} = \mathbf{y} + \boldsymbol{\epsilon}^{\mathbf{I}}, \tag{2}$$

where $\mathbf{y}^{\mathbf{o}}$ is the observation vector and $\boldsymbol{\epsilon}^{\mathbf{I}}$ is the instrument error.

The covariance of the forward model error $E[\boldsymbol{\epsilon}^R \boldsymbol{\epsilon}^{R^T}] = \mathbf{R}^{\mathbf{H}}$ is included in the observation error covariance matrix $\mathbf{R} = \mathbf{R}^{\mathbf{H}} + \mathbf{R}^{\mathbf{I}}$, where $\mathbf{R}^{\mathbf{I}} = E[\boldsymbol{\epsilon}^I \boldsymbol{\epsilon}^{I^T}]$ is the instrument error covariance matrix. To calculate the representativity errors in this paper we use a method defined by Daley [1993] and Liu and Rabier [2002]. In this method it is assumed that the observations can be written as the mapping of a high resolution state into observation space, and that the model state $\mathbf{x}^{\mathbf{m}}$ is a truncation of this high resolution state. This method also allows us to correctly specify the observation operator so our forward model errors consist only of errors of representativity.

2.2 Representativity error on a 1D domain

We restrict our calculations to the 1D domain of length $L = 2a\pi$, where a is a constant that determines the length of the domain, and assume that the observation operator $\mathbf{H}$ is linear. It is assumed that the high resolution state $x(r)$ at position r can be expressed as a Fourier series truncated at wave number K. At N points on the physical domain, $-a\pi \leq r \leq a\pi$, the function values $x(r_j)$, $j = 1 \ldots N$, can be expressed in matrix form as

$$\mathbf{x} = \mathbf{F}\hat{\mathbf{x}} \tag{3}$$

where $\hat{\mathbf{x}}$ is a vector of length $M = 2K+1$ of spectral coefficients and $\mathbf{F}$ is a Fourier transform matrix of dimension $N \times M$. In this work a number of Fourier matrices are used to calculate representativity error. A Fourier matrix $\mathbf{F}$ of size $m \times n$ has elements

$$\mathbf{F}_{j,k} = \exp(\tfrac{2ikj\pi}{m}), \tag{4}$$

where $j = 1 \ldots m$ and $k = 1 \ldots n$.

The model representation of the actual state is a wave number limited filter of the high resolution state, $\hat{\mathbf{x}}^{\mathbf{m}} = \mathbf{T}\hat{\mathbf{x}}$ where $\mathbf{T}$ is a truncation matrix that truncates the full spectral vector $\hat{\mathbf{x}}$ to the analysed spectral vector $\hat{\mathbf{x}}^{\mathbf{m}}$. The model representation of the actual state can be expressed as

$$\mathbf{x}^{\mathbf{m}} = \mathbf{F}^{\mathbf{m}}\hat{\mathbf{x}}^{\mathbf{m}}, \tag{5}$$

where $\hat{\mathbf{x}}^{\mathbf{m}}$ is a vector of length $M_m = 2K_m + 1$ of spectral coefficients and $\mathbf{F}^m$ is a Fourier transform matrix of dimension $N_m \times M_m$ with elements defined as in Eq. (4) but with no terms with wave

number higher than K_m. The variables to be analysed are the Fourier spectral coefficients from $-K_m$ to K_m, $K_m < K$.

We define the observations by

$$y(r_o) = \int_{-a\pi}^{a\pi} x(r)w(r - r_o)dr. \tag{6}$$

Here the observations are defined as if they have been measured at point r_o by a remote sensing instrument. The choice of the weighting function $w(r)$ determines the type of observing instrument. Writing Eq. (6) in spectral space allows us to write the p error free observations as

$$\mathbf{y} = \mathbf{F_p W}\hat{\mathbf{x}}, \tag{7}$$

where $\mathbf{F_p}$ is a $p \times M$ Fourier transform matrix and $\mathbf{W}$ is a $M \times M$ diagonal matrix with elements $\hat{w}_k$, the spectral coefficients of the weighting function $w(r)$. $\mathbf{F_p W}$ is an exact observation operator in spectral space. The measurement vector $\mathbf{y^o}$ is given by,

$$\mathbf{y^o} = \mathbf{F_p W}\hat{\mathbf{x}} + \boldsymbol{\epsilon}^{\mathbf{I}}, \tag{8}$$

where $\boldsymbol{\epsilon}^{\mathbf{I}}$ is the instrument error.

The model representation of the observations is given by,

$$\mathbf{y^m} = \mathbf{F_p^m W^m T}\hat{\mathbf{x}}, \tag{9}$$

where $\mathbf{F_p^m}$ is the Fourier matrix with elements defined as in Eq. (4). $\mathbf{W^m}$ is a $M^m \times M^m$ diagonal matrix with elements $\hat{w}_k$, the spectral coefficients of the weighting function $w(r)$. This method assumes that the low resolution model is a truncation of the high resolution model. This allows representativity error to be considered in the perfect model case.

To obtain an equation for representativity error we assume the observation operator is linear and substitute the definitions of observations, Eq. (7), and model representation of the observation, Eq. (9), into Eq. (1) to give,

$$\boldsymbol{\epsilon}^{\mathbf{R}} = \mathbf{F_p W}\hat{\mathbf{x}} - \mathbf{F_p^m W^m T}\hat{\mathbf{x}}. \tag{10}$$

The expectation operation, denoted $E[.,.]$, is applied to give the representativity error covariance matrix $E[\boldsymbol{\epsilon}^{\mathbf{R}}\boldsymbol{\epsilon}^{\mathbf{R}^*}] = \mathbf{R^H}$,

$$\mathbf{R^H} = (\mathbf{F_p W} - \mathbf{F_p^m W^m T})\hat{\mathbf{S}}(\mathbf{F_p W} - \mathbf{F_p^m W^m T})^*, \tag{11}$$

where $\hat{\mathbf{S}} = E[\hat{\mathbf{x}}\hat{\mathbf{x}}^*]$ is the spectral covariance matrix for the high resolution state and $*$ denotes the complex conjugate transpose. The spectral covariance of the high resolution state, $\hat{\mathbf{S}}$, contains information on how different wave numbers are related. It can be calculated using

$$\hat{\mathbf{S}} = \mathbf{F^{-1}SF}, \tag{12}$$

where $\mathbf{F}$ is a Fourier transform matrix and $\mathbf{S} = E[\mathbf{x}\mathbf{x}^*]$ is the covariance matrix of the high resolution state in physical space.

We now have an equation that can be used to calculate the representativity error covariance matrix. We must now define the weighting matrices to be used and describe how the spectral covariance of the high resolution state can be calculated.

3 The Model and Data

In this study we calculate representativity error for both temperature and specific humidity over the UK. The calculation of representativity error by the method of Liu and Rabier [2002] assumes that the actual state can be taken from a high resolution model. As our actual state we take data from the Met Office UKV model. The UKV model is a variable resolution model that covers the UK; at its highest resolution the domain has 1.5km grid boxes. The model data used is at this resolution. The boundary data for the model is interpolated to the 1.5km grid from the 4km resolution regional model.

In this work we calculate representativity error using the assumption that the model state is a truncation of high resolution data. For the majority of our experiments we chose a truncation factor that gives a model grid spacing equivalent to the grid spacing that is used in the Met Office NAE model. The Met Office NAE model has a grid spacing of 12km (in mid-latitudes) and covers Europe and the North Atlantic.

3.1 The data available

We use temperature and humidity data over the UK available for two cases. The first case, Case 1, consists of data from 7 August 2007 at times 0830UTC, 0900UTC and 0930UTC on an area over the southern UK that covers -3.04^oW to 3.71^oE and 49.18^oN to 53.36^oN. In this case there are partly clear skies with convection occurring over the south east [Eden, 2007]. The second set of data, Case 2, is from 6 September 2008 at 1400UTC, 1430UTC and 1500UTC and covers -5.00^oW to 1.20^oE and 52.5^oN to 56.00^oN. In this case a deep depression is tracking slowly east-northeast across England [Eden, 2008]. The data is available on a 300×300 grid of latitude and longitude lines at each of 43 pressure levels.

3.2 Creating samples from the data

There are some limitations to the data. Data near the boundary is contaminated by the boundary conditions taken from the coarser model. We remove this data at each pressure level by reducing the grid to a 256×256 mesh centered on the main grid. We need to sample the data to calculate the covariance matrices for the actual state. The data we have is available on a 3D gridded domain. We are interested in calculating representativity error for individual pressure levels so for each experiment the data available id 2D; however, we are calculating representativity error on a 1D domain. To convert our data to 1D we take the individual latitude rows of the data from the 749hPa pressure level. We use this level as it is outside the boundary layer, but should still include the small scale features that are relevant when calculating representativity error. For both cases of temperature and specific humidity we have 768 samples to calculate the covariance matrices. A covariance calculated with this number of samples is dominated by sampling error and hence this is not a sufficient number of samples to calculate an accurate representation of the required covariances. A further problem is that the samples are not periodic, but the Liu and Rabier [2002] method assumes a periodic domain with a circulant covariance matrix $\mathbf{S}$. To overcome this we detrend and process the data.

Table 1 – Variances for the true state

	Temperature (K^2)	log(Specific Humidity) $((kg/kg)^2)$
Case 1	0.6638	0.0812
Case 2	0.1934	0.0178

3.3 Data processing

To create surrogate samples from each available sample the data must be detrended. Detrending gives data on a homogeneous field; this is required by our chosen method for calculating representativity error. Data is detrended by removing a best fit line using an appropriate polynomial of order no greater than 3 [Bendat and Piersol, 2011]. It is justifiable to detrend the data as only trends with large length scales are removed. All scales that contribute to the representativity error still remain. We detrend the 256 latitude samples at each available time, and remove a linear trend from the temperature data, and a cubic trend from the humidity data. This detrended data is now used to create new samples from each existing sample.

The method of Fourier randomization is used to generate surrogate samples from the same statistical disribution [Theiler et al., 1992, Small and Tse, 2002]. Fourier randomization consists of perturbing the phase of a set of data to create a new sample with a different phase, but where each wave number retains the same power. Here we calculate circulant samples, which corresponds to shifting the phase of the data. This also gives the data the required periodicity. A circulant sample is created by shifting each element of the sample one position and taking the final element and making it the first entry in the sample. Each element can be shifted to each position, which means a sample with n elements can be used to create n circulant samples. Therefore creating surrogate samples increases the number of available samples we have for calculating the covariance of the high resolution data. We have available 256 samples at three different times. Creating circulant samples gives us 65536 samples at each time, and a total of 196608 samples to estimate each of the covariance matrices, which is a sufficient number of samples.

4 Experiments

4.1 The covariance of the high resolution data

We calculate the sample covariance matrix, $\mathbf{S}$, of the high resolution data using,

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^*, \tag{13}$$

where $\mathbf{x_i}$ is the i^{th} sample vector and $\bar{\mathbf{x}}$ is a mean vector of the samples. We use this method and the model data to calculate the covariance matrix of the actual state, $\mathbf{S}$. We use the circulant samples calculated from the UKV model data to calculate the covariance matrices for the temperature and humidity fields for both cases. We give the variances in Table 1 and plot a row of each of the correlation matrices in Figure 1.

From Table 1 we see that the variances for Case 2 are smaller than those for Case 1. When considering the correlations plotted in Figure 1, we see that the temperature fields have larger correlations than humidity. For Case 1 the temperature correlations are very high; this is expected

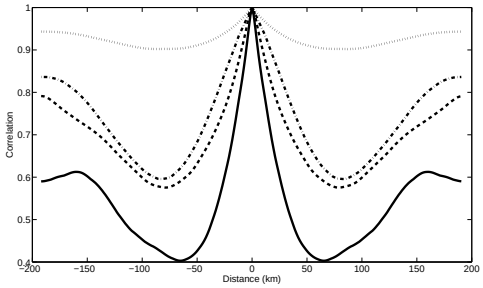


Figure 1 – Correlation structure for the true temperature and specific humidity fields. Temperature: Case 1 dotted line, Case 2 dot-dash line. Specific Humidity: Case 1 dashed line, Case 2 solid line

as, after detrending, this field is fairly constant. We also note that the correlations for Case 2 are smaller than the correlations for Case 1. This is due to the synoptic situation, since in Case 1 the field is more homogeneous with small scale features over the domain; but, in Case 2 the features are large scale and less homogeneous.

4.2 The observations

To calculate the representativity error we require pseudo-observations. We expect representativity error to depend on observation type. To calculate these observation types we use Eq. (7), which requires a weighting matrix. We choose the weighting matrices in Eq. (11) to correspond to different types of observing instruments. The elements of the weighting matrix are the spectral coefficients of the weighting function $w(r)$ that is used to define observations using Eq. (6). Pseudo-observations are created from the high resolution data using three weighting functions. The weighting functions used here are the same as those used in Liu and Rabier [2002]. Two of the weighting functions represent remotely-sensed observations. One follows a top-hat (uniform) function with a width of 5km. The other weighting function is calculated using a Gaussian curve with a width of 20km. We also consider in-situ measurements. For these direct observations the weighting function $w(r)$ in Eq. (6) becomes a Dirac delta-function. In this case the diagonal elements $\hat{w}$ of the weighting matrix are all unity.

Now we have the appropriate weighting matrices and the covariance matrices for the high resolution data at the 749hPa pressure level. This allows us to calculate representativity errors for temperature and log specific humidity. In the next section we present the results of our experiments.

5 Results

We now carry out a number of experiments to enable us to understand the nature of representativity error. The results for experiments carried out with data from Case 1 are given in Table 2, and for Case 2 in Table 3.

5.1 Temperature and humidity representativity errors

We first consider how the errors of representativity differ between the temperature and log humidity fields. We consider the representativity error for the case where the model has 32 points; this is

Table 2 – Representativity error (RE) variances for Case 1. The values given in brackets are a comparison of the representativity error variance to the high resolution data variance.

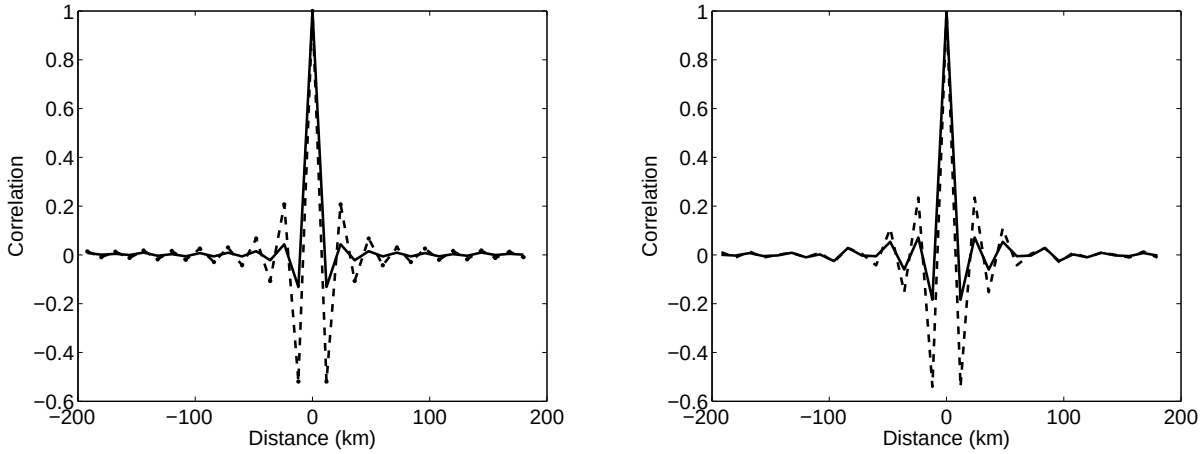
Experiment Number	Truncation	Number of Observations (p)	Observation type	Temperature RE variance ($(K)^2$)	Humidity RE variance $((kg/kg)^2)$
1.1	32	32	Direct	4.81×10^{-3} (0.7%)	1.51×10^{-3} (1.9%)
1.2	32	32	Uniform	2.71×10^{-3} (0.4%)	1.08×10^{-3} (1.3%)
1.3	32	32	Gaussian	8.99×10^{-4} (0.1%)	3.80×10^{-4} (0.5%)
1.4	32	16	Direct	4.81×10^{-3} (0.7%)	1.51×10^{-3} (1.9%)
1.5	64	64	Direct	2.13×10^{-3} (0.3%)	4.04×10^{-4} (0.5%)

Table 3 – Representativity error (RE) variances for Case 2. The values given in brackets are a comparison of the representativity error variance to the high resolution data variance.

Experiment Number	Truncation	Number of Observations (p)	Observation type	Temperature RE variance ($(K)^2$)	Humidity RE variance $((kg/kg)^2)$
2.1	32	32	Direct	2.21×10^{-3} (1.1%)	7.14×10^{-4} (4.0%)
2.2	32	32	Uniform	1.30×10^{-3} (0.7%)	4.84×10^{-4} (2.7%)
2.3	32	32	Gaussian	3.96×10^{-4} (0.2%)	1.60×10^{-4} (0.9%)
2.4	32	16	Direct	2.21×10^{-3} (1.1%)	7.14×10^{-4} (4.0%)
2.5	64	64	Direct	1.12×10^{-3} (0.6%)	2.50×10^{-4} (1.4%)

a truncation of a factor of eight from the high resolution model, which has 256 points. We start by assuming that we have direct observations. The values of the representativity error variance are given in Table 2 Experiment 1.1. We plot in Figures 2(a) and 2(b) (solid lines) the middle row of the representativity error correlation matrices for temperature and log specific humidity from Case 1 respectively.

When we compare the variance of representativity error against the variance of the actual states we see that representativity error is more significant for humidity than it is for temperature. We find that the representativity error variance for temperature is 0.7% of the high resolution temperature variance, whereas the humidity representativity error variance is 1.9% of the high resolution humidity field variance. When comparing the variance from this experiment to the same experiment carried out with Case 2 data (Table 3, Experiment 2.1), we see that the representativity error variances are smaller for Case 2. This is expected as there is less variance in the true fields in Case 2. These experiments show however, that the representativity error is more significant in this case, with representativity error for temperature being 1.1% of the high resolution temperature variance and humidity representativity error being 4.0% of the high resolution variance. For Case 1 from Figures 2(a) and 2(b) (solid line) we see that the correlation structure is similar for both temperature and specific humidity. The covariances rapidly decrease in magnitude as the separation distance increases. The correlations for specific humidity are slightly larger, and decay less rapidly than the correlations for temperature.



(a) Temperature (K)

(b) log(Specific humidity) (kg/kg)

Figure 2 – Representativity error correlations between observation center points for Case 1 with truncation to 32 points (12km resolution) with every model grid point observed using direct (solid line) and Gaussian-weighted (dashed line) observations. a) Temperature b) log(Specific humidity)

5.2 Changing the observation type

We now consider what happens where the observations are defined with a uniform weighting matrix. The variance of the representativity error is given in Table 2 Experiment 1.2. We see again, as expected, that the representativity error is more significant for humidity than it is for temperature. We see that the assumption of uniformly weighted observations has decreased the representativity error for both fields. This is as expected as the uniform observations do not capture all the small scales that the direct observations can. The correlations are larger than those when direct observations are used (result not shown). This is because two consecutive observations have some overlap in physical space. We see from Table 3 that Experiment 2.2 supports these results as the RE variance is smaller than those seen in Experiment 2.1.

We now consider what happens where the observations are defined using a Gaussian-weighting matrix. The results are given in Table 2 Experiment 1.3 and Table 3 Experiment 2.3. We plot the middle row of the representativity error correlation matrices for temperature and log specific humidity from Case 1 in Figures 2(a) and 2(b) (dashed lines). We find that the error variance is smaller than when either direct or uniform observations are assumed. Our Gaussian weighted observations capture fewer small scale features than the direct and uniform observations and hence the representativity error variance is smaller as the model captures a larger proportion of the scales captured by the observations. From the Figures we see that the correlations for the representativity error calculated with these Gaussian-weighted observations are larger than the representativity error correlations present when direct observations are used. This is due to the overlapping of the weighting functions in physical space of nearby observations.

By comparing the experiments with different weighting functions we see that the larger the weighting function length scale used to define the observation the lower the representativity error variance. Observations defined using weighting functions with larger length-scales capture fewer spatial scales. Therefore the difference between a larger length-scale observation and the model representation of the observation is smaller than a small length-scale observation and the model

representation of the observation. Hence observations defined using weighting functions with larger length scales result in smaller errors of representativity.

5.3 Number of observations

We now consider what happens when we calculate the representativity error where fewer direct observations are available. Experiments 1.4 in Table 2 and 2.4 in Table 3 show the error variance where only half the model grid points have associated direct observations. We see that having fewer observations available does not alter the variance of the representativity error. This is expected as representativity error applies individually to each observation and is independent of other observations. Experiments with uniform and Gaussian observations also support this conclusion. The Liu and Rabier [2002] method makes use of the Fourier transform; this leads to regularly spaced observations over the domain. The method used here leads to a class of correlation structures for representativity error that are dependent only on the distance between observations and not the number of observations available. We show this in Appendix A. Although the results in Appendix A are specific to the Liu and Rabier [2002] method, in general we expect that the representativity error should not be dependent on the number of available observations, only on the distance between them.

5.4 Number of model grid points

We now consider the results when the model has a larger number of 64 grid points. This is a higher truncation, so the model should be able to resolve more small scale features, and hence we expect the errors of representativity to decrease. We give the results for experiments with direct observations in Table 2 Experiment 1.5 and Table 3 Experiment 2.5. From these results we see that the representativity error variances have decreased. For direct observations the representativity error has been approximately halved. Experiments with uniform and Gaussian weighings are not shown but produce results that also support this conclusion.

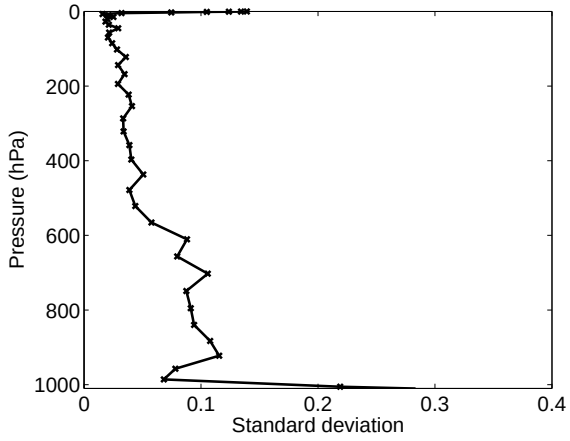
5.5 Representativity errors at different model levels

So far we have only considered the representativity error at the 749hPa model level height. We now calculate a representativity error for each pressure level of the model. This will allow us to consider the variation of representativity error with height. From this we can determine if one realisation of representativity error would be suitable at every pressure level, or if it is more appropriate to use the correct representativity error for each level.

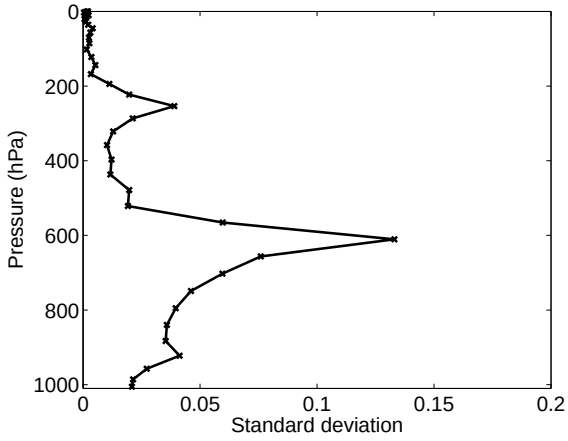
Before calculating the representativity error for each model level, we must first calculate the covariance matrices for the high resolution data for temperature and specific humidity for each pressure level. We use the same data, but at the correct pressure level, and the same preprocessing techniques described in section 3.

We consider the case where we have truncated to 32 grid points and have 32 direct observations available. We plot the standard deviation of representativity error for Case 1 in Figure 3(a) (temperature) and 3(b) (specific humidity) and for Case 2 in Figure 4(a) (temperature) and 4(b) (specific humidity).

From the figures we see that representativity error for temperature is more constant with height than specific humidity. The exception to this is in the boundary layer, where the temperature

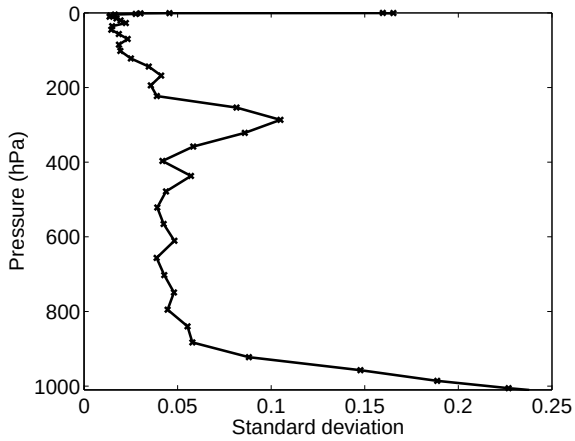


(a) Temperature (K)

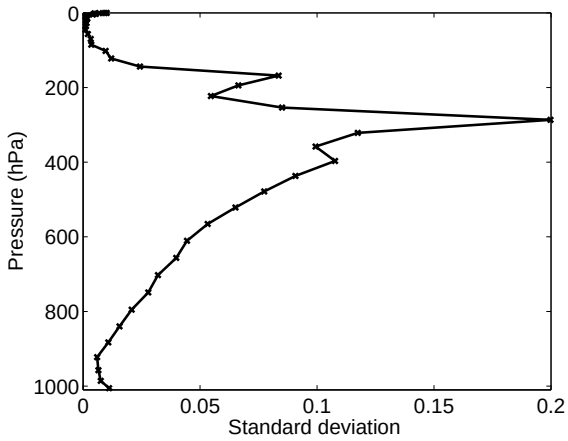


(b) log(Specific humidity) (kg/kg)

Figure 3 – Representativity error standard deviation with model level height. Case 1 with 32 direct observations (every model grid point observed). a) Temperature b) log(Specific humidity)



(a) Temperature (K)



(b) log(Specific humidity) (kg/kg)

Figure 4 – Representativity error standard deviation with model level height. Case 2 with 32 direct observations (every model grid point observed). a) Temperature b) log(Specific humidity)

representativity error is large. For specific humidity in Case 1 we see a large increase in the representativity error standard deviation between 749hPa and 610hPa. For Case 2 the largest peak in representativity error is seen at 300hPa. These levels are where cloud is seen and hence it is at these levels where the small scale humidity features exist; this results in the larger representativity error variances.

The results show that representativity error is not constant with height and in some cases it may be beneficial to have a representativity error matrix that varies with height. They also support our conclusions that representativity error is strongly case dependent. Experiments (results not shown) with uniform and Gaussian weighting functions and truncation to 64 points also show that representativity error varies with height.

6 Conclusions

In this study we wished to investigate whether significant correlations in the observation error matrix [Stewart et al., 2009, Stewart, 2010, Stewart et al., 2012b] could be attributed to representativity error, and whether representativity error is more significant for humidity than temperature. We calculated and compared representativity errors for temperature and specific humidity. Experiments using data from the Met Office UKV model showed that representativity error was more significant for humidity than temperature. This was determined by comparing how large the representativity error variance was in comparison to the variances of the high resolution states. This suggests that correlations found in previous approximations of the observation error covariance matrix, $\mathbf{R}$, such as those in [Stewart, 2010] and [Weston, 2011], are likely to be representativity error, at least in part. We calculated representativity error using data from two different cases and showed that representativity error is sensitive to the synoptic situation, which supports claims by Janjic and Cohn [2006]. We also found that as the number of model grid points is reduced the representativity error increases. This is because at lower resolution, the model is not able to resolve as many scales. We also found that using direct observations gave a higher representativity error than when uniform or Gaussian-weighted observations were used. This is because the direct observations contain more information on smaller scales than the uniform or Gaussian-weighted observations. Experiments showed that altering the number of observations used to calculate representativity error had no effect on the representativity error variance. We showed that this method leads to a class of correlation structures that depend only on the distance between observations and not the number of observations. We believe that in general the number of observations should not affect the representativity error correlation structure and that the structure is dependent only on the distance between the available observations. Finally we considered how the representativity error standard deviation varied at different pressure levels. We found that representativity does vary at different pressure levels and this means that assumptions such as those in Dee and Da Silva [1999], where errors at different model levels are fixed, may not be suitable when representativity error is taken into account. As it becomes more efficient to use correlated observation errors in data assimilation systems good approximations of the observation error covariance matrix will be required. Using the method of Liu and Rabier [2002] we have calculated representativity errors for specific fields and have shown that errors of representativity are correlated. However further work is required to determine whether the inclusion of these errors in an assimilation scheme improves the analysis.

Acknowledgments

The authors thank Laura Stewart for providing the data. This work was funded by NERC as part of the National Center for Earth Observation and the Met Office through a CASE studentship.

A Representativity error variance

Here we show why the representativity error variance does not alter when calculated with different numbers of observations. We do this by considering the calculation of the elements of the representativity error matrix.

Representativity error is calculated using Eq. (11). The matrices are:

- $\mathbf{F}$ with $j = 1 \dots p$ and $k = 1 \dots M$. Elements defined as in Eq. (4).
- $\mathbf{F^m}$ with $j = 1 \dots p$ and $k = 1 \dots M_m$. Elements defined as in Eq. (4).
- $\mathbf{W}$ with $j = 1 \dots M$ and $k = 1 \dots M$. Elements $\hat{w}_j$ when $j = k$, 0 otherwise.
- $\mathbf{W^m}$ with $j = 1 \dots M_m$ and $k = 1 \dots M_m$. Elements $\hat{w}_j$ when $j = k$, 0 otherwise.
- $\hat{\mathbf{S}}$ with $j = 1 \dots M$ and $k = 1 \dots M$. Elements $\hat{s}_j$ when $j = k$, 0 otherwise.
- $\mathbf{T}$ with $j = 1 \dots M_m$ and $k = 1 \dots M$. Elements 1 when $j = k$ and $k \leq M$, 0 otherwise.

First we calculate the elements of $\mathbf{A} = \mathbf{FW}$ and $\mathbf{B} = \mathbf{FWT}$. As many of the elements are zero we find that $A_{j,k} = F_{j,k}W_{j,j}$ and $B_{j,k} = F^m_{j,k}W^m_{j,j}$ when $k \leq M_m$, and 0 otherwise.

Next we calculate elements of $\mathbf{C} = \mathbf{A} - \mathbf{B}$.

$$C_{j,k} = \begin{cases} \exp(\frac{2ikj\pi}{p})\hat{w}_j & M_m < k \leq M \\ 0 & \text{Otherwise} \end{cases} \tag{14}$$

Now we calculate $\mathbf{R^H} = \mathbf{C\hat{S}C}^*$. Elements of $\mathbf{E} = \mathbf{C\hat{S}}$ are $E_{j,k} = C_{j,k}\hat{S}_{k,k}$. Finally we calculate $\mathbf{R^H} = \mathbf{EC}^*$, where $*$ is the complex conjugate transpose and $\bar{.}$ is the complex conjugate,

$$R^H_{j,k} = \sum_{l=0}^{M} E_{j,l}C^*_{l,k}, \tag{15}$$

$$= \sum_{l=1}^{M} C_{j,l}\hat{S}_{l,l}\bar{C}_{k,l}, \tag{16}$$

$$= \sum_{l=1}^{M} \exp(\tfrac{2ijl\pi}{p})\hat{w}_l\hat{s}_l\hat{w}_l \exp(\tfrac{-2ikl\pi}{p}). \tag{17}$$

We now show that the variance does not change when p changes, this is the case when $j = k$; that is,

$$R^H_{j,j} = \sum_{l=1}^{M} \exp(\tfrac{2ijl\pi}{p})\hat{w}_l\hat{s}_l\hat{w}_l \exp(\tfrac{-2ijl\pi}{p}), \tag{18}$$

$$= \sum_{l=1}^{M} \hat{w}_l\hat{s}_l\hat{w}_l. \tag{19}$$

This does not depend on p and hence we do not expect the variance to change when we use different numbers of observations to calculate representativity error.

We now show that the correlation structure depends only on the distance between observations and not the number of observations.

Our model has N_m grid points separated by a spacing Δx and we have p observations. The distance between consecutive observations is $\frac{N_m\Delta x}{p}$. Suppose we have two observations separated by a distance d and assume that these are observation j and observation k. Then we have

$$d = \frac{(jk)(N_m\Delta x)}{p}, \tag{20}$$

and hence

$$\frac{jk}{p} = \frac{d}{N_m\Delta x} \tag{21}$$

Substituting this into Eq. (17) we obtain

$$R^H_{j,k} = \sum_{l=1}^{M} \hat{w}_l\hat{s}_l\hat{w}_l \exp(\tfrac{2il\pi(j-k)}{p}), \tag{22}$$

$$= \sum_{l=1}^{M} \hat{w}_l\hat{s}_l\hat{w}_l \exp(\tfrac{2il\pi d}{N_m\Delta x}). \tag{23}$$

Hence the correlations depend only on the distance between the observations and not the number of observations.

References

J. S. Bendat and A. G. Piersol. *Random Data: Analysis and Measurement Procedures*. John Wiley and Son, fourth edition, 2011.

N. Bormann and P. Bauer. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136:1036 – 1050, 2010.

N. Bormann, S. Saariene, G. Kelly, and J. Thepaut. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Monthly Weather Review*, 131:706 – 718, 2002.

N. Bormann, A. Collard, and P. Bauer. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II :application to AIRS and IASI data. *Quarterly Journal of the Royal Meteorological Society*, 136:1051 – 1063, 2010.

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

R. Daley. Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophysicae*, 11:634–647, 1993.

D. P. Dee and A. M. Da Silva. Maximum-likelihood estimation of forecast and observation error covariance parameters. part I: Methodology. *Monthly Weather Review*, 127:1822 – 1843, 1999.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131: 3385 – 3396, 2005.

P. Eden. Weather log: August 2007. *Weather*, 62:i – iv, 2007.

P. Eden. Weather log: September 2008. *Weather*, 63:i – iv, 2008.

M. Fisher. Accounting for correlated observation error in the ECMWF analysis. Technical report, ECMWF, 2005. MF/05106.

S. B. Healy and A. A. White. Use of discrete Fourier transforms in the 1d-var retrieval problem. *Quarterly Journalof the Royal Meteorological Society*, 131:63 – 72, 2005.

F. Hilton, A. Collard, V. Guidard, R. Randriamampianina, and M. Schwaerz. Assimilation of IASI radiances at european NWP centres. *In Proceedings of Workshop on the assimilation of IASI data in NWP, ECMWF, Reading, UK, 6-8 May 2009*, pages 39–48, 2009.

A. Hollingsworth and P. Lönnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, 38A:111 – 136, 1986.

P. L. Houtekamer and H. L. Mitchell. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 133:3260 – 3289, 2005.

T. Janjic and S. E. Cohn. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, 134:2900 – 2915, 2006.

W. Lahoz, B. Khattatov, and R. Menard, editors. *Data Assimilation Making Sense of Observations*, chapter 2. Springer, 2010.

Z.-Q. Liu and F. Rabier. The interaction between model resolution observation resolution and observation density in data assimilation: A one dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128:1367–1386, 2002.

A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112:1177 – 1194, 1986.

M. Small and C. Tse. Applying the method of surrogate data to cyclic time series. *Physica D*, 164: 187–201, 2002.

L. Stewart, S. Dance, and N. Nichols. Data assimilation with correlated observation errors: analysis accuracy with approximate error covariance matrices. Technical report, University of Reading, 2012a. Department of Mathematics and Statistics Preprint MPS-2012-17, www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

L. Stewart, S. Dance, N. Nichols, J. R. Eyre, and J. Cameron. Estimating interchannel observation error correlations for iasi radiance data in the met office system. Technical report, University of Reading, 2012b. Department of Mathematics and Statistics Preprint, www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

L. M. Stewart. *Correlated observation errors in data assimilation*. PhD thesis, University of Reading, 2010.

L. M. Stewart, S. L. Dance, and N. K. Nichols. Correlated observation errors in data assimilation. *Int. J. Numer. Meth. Fluids*, 56:15211527, 2008.

L. M. Stewart, J. Cameron, S. Dance, S. English, J. R. Eyre, and N. Nichols. Observation error correlations in IASI radiance data. Technical report, University of Reading, 2009. Mathematics reports series, www.reading.ac.uk/web/FILES/maths/obs_error_IASI_radiance.pdf.

J. Theiler, S. Eubank, A. Longtin, G. B., and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.

P. Weston. Progress towards the implementation of correlated observation errors in 4d-var. Technical report, Met Office, UK, 2011. Forecasting Research Technical Report 560.

J. S. Whitaker, T. M. Hamill, X. Wei, Y. Song, and Z. Toth. Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, 136:463 – 482, 2008.