A pragmatic strategy for implementing spatially correlated observation errors in an operational system: an application to Doppler radial winds

Article

Accepted Version

A pragmatic strategy for implementing spatially correlated observation errors in an operational system: an application to Doppler radial winds

D. Simonin[a*], J. A. Waller[b], S. P. Ballard[a], S. L. Dance[b], and N. K. Nichols [b]

[a]*MetOffice@Reading, Meteorology Building, University of Reading, Reading, Berkshire, RG6 6BB, United Kingdom*

[b]*School of Mathematical, Physical and Computational Sciences, University of Reading, Reading, Berkshire, RG6 6BB, United Kingdom*

*Correspondence to: D. Simonin, MetOffice@Reading, Meteorology Building, University of Reading, Reading, Berkshire, RG6 6BB, United Kingdom. E-mail: david.simonin@metoffice.gov.uk

Recent research has shown that high resolution observations, such as Doppler radar radial winds, exhibit spatial correlations. High resolution observations are routinely assimilated into convection permitting numerical weather prediction models assuming their errors are uncorrelated. To avoid violating this assumption observation density is severely reduced. To improve the quantity of observations used and the impact that they have on the forecast requires the introduction of full, correlated, error statistics. Some operational centres have introduced satellite inter-channel observation error correlations and obtained improved analysis accuracy and forecast skill scores. Here we present a strategy for implementing spatially correlated observation errors in an operational system. We then provide the first demonstration of the practical feasibility of incorporating spatially correlated Doppler radial wind error statistics in the Met Office numerical weather prediction system.

Inclusion of correlated Doppler radial winds error statistics has little impact on the computation cost of the data assimilation system, even with a four-fold increase in the number of Doppler radial winds observations assimilated. Using the correlated observation error statistics with denser observations produces increments with shorter length scales than the control. Initial forecast trials show a neutral to positive impact on forecast skill overall, notably for quantitative precipitation forecasts. There is potential to improve forecast skill by optimising the use of Doppler radial winds and applying the technique to other observation types.

Key Words: Doppler radial wind, Observation error, spatial correlations, parallelisation, operational implementation

1. Introduction

Error characteristics of atmospheric observations are complex and not straightforward to derive. Each meteorological instrument is accurate to within a given tolerance subject to its engineering specifications. This is called instrument error. However, in the context of data assimilation, there is a representation error that arises in addition to the instrument error. The sources of representation error include the variability of the observed field at scales different from those resolved by the assimilating dynamical model, observation pre-processing and/or the approximation of the observation operator (Janjić *et al.* 2017). Therefore, the total observation error can be expressed as the sum of the instrument error and a representation error. It is generally assumed that instrument error is uncorrelated and unbiased (any existing biases are assumed to have been removed). In contrast the error of representation is generally correlated and state dependent (Waller *et al.* 2014).

Idealized studies have shown that incorporating correlated observation errors in data assimilation systems leads to a more accurate analysis (Stewart *et al.* 2013; Stewart 2010; Healy and White 2005) and to the inclusion of more observation information content (Stewart *et al.* 2008), particularly on small scales (Rainwater *et al.* 2015; Fowler *et al.* 2018). Studies with operational data have shown that satellite inter-channel errors can exhibit significant correlations (Stewart *et al.* 2009, 2014; Bormann and Bauer 2010; Bormann *et al.* 2010; Waller *et al.* 2016a), and accounting for them in the assimilation results in improvements in the forecast skill score (Weston *et al.* 2014; Bormann *et al.* 2016; Campbell *et al.* 2017), but may affect the number of iterations required to solve the variational minimization problem (Tabeart *et al.* 2018). More recent research has shown that observation errors can also be spatially correlated (Waller *et al.* 2016c,a; Cordoba *et al.* 2017).

The UK public weather service has an emphasis on accurate forecasts/nowcasts of strong convective storms which can be responsible for major flooding events. In response, the UK Met Office has an operational convection permitting numerical weather prediction (NWP) system using a 1.5km version of the Unified Model (UM) (Lean *et al.* 2008; Tang *et al.* 2013). Such a system requires the assimilation of new, high temporal and spatial resolution observations in order to provide an initial state that contains information at suitable scales (Gao and Stensrud 2012; Sun *et al.* 2014; Clark *et al.* 2015; Ballard *et al.* 2016). Such observations include, for example, mode-S aircraft data (e.g. de Haan and Stoffelen 2012; Strajnar *et al.* 2015; Lange and Janjić 2016), weather radar (e.g. Caya *et al.* 2005; Wattrelot *et al.* 2014; Wang and Wang 2017) or high resolution AMVs (e.g. Velden *et al.* 2017). However, due to the presence of correlated errors, there has been no attempt to operationally assimilate observations at a high spatial density. Instead, the observations are assumed to be spatially uncorrelated; the data is thinned to separation lengths where this assumption is understood to be reasonable and the error variances inflated to account for any neglected correlations (Buehner 2010). As a result, the quantity of high resolution observations, such as those provided by weather radar in the form of reflectivity, radial wind (Simonin *et al.* 2014) and refractivity, is severely reduced. This may result in a sub-optimal analysis and poorer forecasts. Therefore, in order to assimilate observations at a high spatial density the observation error correlations must be considered.

This work proposes a pragmatic strategy that allows the use of horizontally correlated observation errors. We describe the implementation of such a strategy within the Met Office operational variational assimilation scheme. Practical feasibility and possible impacts are demonstrated with NWP trial experiments using spatially correlated observation error for Doppler radial wind.

First, we present the current assimilation system used at the Met Office in Section 2. Subsequently, in Section 3, we describe in detail the implementation of the proposed strategy that allows use of correlated observation error statistics. After presenting the experimental details in section 4, section 5 shows the impact of the new scheme when it is applied to Doppler radial wind observations for the assimilation system, the analysis and the forecasts. Finally we conclude in Section 6.

2. The current Met Office approach

In this section we describe the current Met Office variational data assimilation system software (VAR) and its parallelisation. We also describe the current treatment of observation error statistics in the assimilation.

2.1. The data assimilation system

In this section we describe some pertinent features of the current Met Office variational data assimilation (VAR) software (Lorenc *et al.* 2000; Rawlins *et al.* 2007). These schemes are based on the incremental approach of Courtier *et al.* (1994) and are applicable to 3D-Var and 4D-Var. Here we document them following the notation of Ide *et al.* (1997).

Given a full resolution non-linear forecast model, incremental variational assimilation seeks a simplified, perturbation model state increment $\delta w^a \in \mathbb{R}^{n_s}$ to a full resolution guess field $\mathbf{x}^g \in \mathbb{R}^n$ such that the analysis at full resolution $\mathbf{x}^a \in \mathbb{R}^n$ at $t = T + 0$ is given by

$$\mathbf{x}^a = \mathbf{x}^g + \mathbf{S}^{-1}\delta\mathbf{w}^a. \quad (1)$$

Here, $\mathbf{S}^{-1}$ is the incrementing operator; it is the generalised non-linear inverse of a simplification operator $\mathbf{S}$ which reduces the full model's complexity and resolution to that of the perturbation (Ide *et al.* 1997). In the Met Office VAR schemes, where the full resolution non-linear model is the UM, the operator $\mathbf{S}$ is also used to simplify multiple moisture and cloud variables to a single variable (Rawlins *et al.* 2007). We find the perturbation model state, $\delta\mathbf{w}^a$, at $t = T + 0$ by minimizing a penalty function,

$$\begin{aligned} J(\delta\mathbf{w}) &= \frac{1}{2}\left(\delta\mathbf{w} - \delta\mathbf{w}^b\right)^T \mathbf{B}^{-1}\left(\delta\mathbf{w} - \delta\mathbf{w}^b\right) \\ &\quad + \frac{1}{2}\left(\mathbf{y} - \mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{y}^o\right) \\ &= J_b + J_o, \end{aligned} \quad (2)$$

where $\delta\mathbf{w} = \mathbf{S}(\mathbf{x}) - \mathbf{S}(\mathbf{x}^g)$ and $\delta\mathbf{w}^b = \mathbf{S}(\mathbf{x}^b) - \mathbf{S}(\mathbf{x}^g)$, $\mathbf{x}^b \in \mathbb{R}^n$ is the background model state, $\mathbf{B} \in \mathbb{R}^{n\times n}$ is the background error covariance matrix and $\mathbf{R} \in \mathbb{R}^{p\times p}$ is the observation error covariance matrix. The penalty function minimizes the fit of the model state to the background state, J_b, and observations, J_o. Note that the variational problem is solved iteratively using a conjugate gradient method.

This work was conducted using a 3D-Var assimilation system with a centered window using first guess at appropriate time (FGAT: Fisher and Andersson 2001; Lorenc and Rawlins 2005). The observations $\mathbf{y}^o \in \mathbb{R}^p$ are distributed within an assimilation time window $[T - t_\mathbf{w}, T + t_\mathbf{w}]$. The background model state is provided by a previous forecast and is interpolated in time to the observation time. Following Lorenc and Jardak (2018), the model prediction of the observations is given by $\mathbf{y} = H(\underline{\mathbf{G}}\mathbf{x}^g + \underline{\mathbf{G}}\mathbf{S}^{-1}\delta\mathbf{w})$ where $\underline{\mathbf{G}}$ is the linear time- and space-interpolation of the model generated field to the observation location and time and H is the non-linear observation operator.

In order to calculate the model prediction of the observations it is necessary to interpolate the primary variables of the forecast model and the perturbation forecast model to the observation locations. Therefore, for each observation we define:

- The array $C_x = \underline{\mathbf{G}}\mathbf{x}^g$ consisting of a vertical column of the primary variables of the forecast model, interpolated horizontally to the observation positions, valid at the observation time.
- The array $C_w = \underline{\tilde{\mathbf{G}}}\mathbf{S}^{-1}\delta\mathbf{w}$ consisting of a column of the primary variables of the perturbation forecast model, interpolated horizontally (and in time for 4D-VAR) to the observation positions. 3D-Var treats all increments at the same analysis time (in the middle of the window) so $\underline{\tilde{\mathbf{G}}}$ incorporates a space-interpolation only; FGAT is implemented by the time-interpolation to the exact time of each observation, in $\underline{\mathbf{G}}$.
- The array $\widehat{C}_w$, the derivative of the observation penalty function (J_o) with respect to the primary variables of the perturbation forecast model (C_w).

2.2. Parallelisation

The current approach to the parallelisation of the VAR code follows the Data Parallel paradigm (Pacheco 1997, section 2.2.3): all the processing elements (PEs) carry out the same operations on different portions of the data set (figure 1 top panel). The data is split into a number of geographical regions; this is known as Domain Decomposition.

For VAR, the domain decomposition splits the C_w columns such that each PE has information containing all the vertical levels but only for a specified area of the horizontal-plane. The

PE IDs assigned to each column of C_w are stored in the vector $Cw_{PE} \in \mathbb{R}^p$.

As the observations are assumed to be independent and uncorrelated, they are simply spread across processors following the same regional decomposition as the model information (shown in figure 1). In this approach, the costs of the observation calculations are small compared to other components because there are no inherent message-passing or synchronisation delays. This advantage outweighs the inefficient load-balancing, for the domains typically used. The allocated PE ID for each observation is stored in the vector $Obs_{PE} \in \mathbb{R}^p$. This strategy is applied to all observation types and to the model information such that $Obs_{PE} = Cw_{PE}$.

2.3. *Treatment of observation error statistics*

Observation errors are typically assumed to be temporally uncorrelated, and with no correlations between observation types, so that $\mathbf{R}$ is block-diagonal. This allows J_o to be calculated independently for each observation type and hence reduces the cost of the matrix-vector products in equation (2). Also, for many observation types, it is assumed that the observation errors are independent, Gaussian white noise, so that the associated observation error covariance sub-matrix for a given observation in equation (2) is diagonal (no cross-correlation) and contains the sum of instrument and representation errors $\mathbf{R} = \mathbf{E} + \mathbf{F}$ (Lorenc *et al.* 2000). In this case the matrix-vector product simplifies to a series of scalar multiplications. There is one exception to this description. The current system accounts for correlated satellite inter-channel errors (Weston 2011; Weston *et al.* 2014). In this case, sets of observations with inter-channel error correlations provide information related to a single model column; hence the inclusion of correlated inter-channel error matrices is compatible with the current parallelisation strategy where observation and vertical model columns are distributed together between supercomputer processors (see Section 2.2 for a more detailed description). However, in the case of horizontally correlated observation error statistics, the existing data-structures do not allow the computation of the required matrix-vector products without excessive communication between processors.

3. The new approach

In this section we describe how the current Met Office variational data assimilation system software (VAR) has been adapted to exploit, and allow for, horizontal correlated observation error statistics.

3.1. *Parallelisation*

As shown in section 2.2, the assimilation system is using the same domain decomposition for observations as model. However, in order to make use of a full observation error covariance matrix, $\mathbf{R}_s$, (i.e. variance and correlation), it is necessary to gather error-correlated observations, and their model equivalent, on a single processor as shown in the bottom panel of figure 1.

To accommodate full observation error covariance matrices, the parallelization has been modified for observations that have mutually correlated errors. These observations are assigned to a *family* and sent to a single PE (figure 1 bottom panel) and are no longer distributed on a PE according to its geographical location but according to its family instead: $Obs_{PE} \neq Cw_{PE}$. If no family has been defined (observations with uncorrelated errors shown as blue dots in the bottom panel of 1), then the distribution of the information across the numerous processors is done in the traditional way (i.e domain decomposition $Obs_{PE} = Cw_{PE}$).

If some observations are believed to be correlated and associated to families, the main steps of the algorithm are:

- Each family of observations is assigned to a unique processor, following the Obs_{PE} assignment.
- The C_w's are still distributed using the domain decomposition (following the Cw_{PE} assignment), to allow horizontal interpolation to be a local operation on each PE.
- At each iteration, all the C_w's associated with a family of observations are gathered into the processor assigned to this family.
- The observation penalty (J_o) is calculated.
- The last step is to redistribute the $\widehat{C}_w$'s to their original location according to the Cw_{PE} assignment.

This new approach could significantly increase the communication between processors. However, the added communications are not all-to-all; a set of lookup tables have been implemented to ensure a "link" between Obs_{PE} and Cw_{PE}. This restricts the communication to a minimum. In addition, the dissociation between the Obs_{PE} and Cw_{PE} offers the opportunity to improve the load balancing. Observations are rarely uniformly distributed across the model domain, which implies that some processors will have more work than others if a domain decomposition is used. With this new approach, families can be allocated to the least loaded processor and improve the overall load balancing of the system. The only real limitation of this approach is in the definition of families. For observation types such as radar observations, or GPS, where natural groupings exist, it is relatively easy to use. However for observations such as geostationary satellite imagery, where the entire model domain is covered by one single image, the creation of families is more difficult. One approach for this case is for families to represent a section of the domain, with extra observations forming a halo.

3.2. *Treatment of observation error statistics*

The proposed approach for using spatially correlated errors is to treat each family in a similar way to the current approach for inter-channel correlations mentioned in section 2: Since $\mathbf{R}$ and its inverse change each assimilation due to the quality control process and observation availability, a Cholesky decomposition method is used to calculate the observation penalty, J_o as described in Weston *et al.* (2014) . This avoids the need to compute the inverse observation error covariance matrix directly. The method requires positive definite symmetric matrices, which covariance and correlation matrices are by definition, and is computationally cheaper than alternatives such as Gaussian elimination. This approach for handling correlated observation errors relies on the full $\mathbf{R}$ being block diagonal, otherwise it may be necessary to use an approximation method such as Yaremchuk *et al.* (2018).

For each family it is necessary to determine the full spatial observation error correlation matrix $\mathbf{C}$ and a matrix of standard deviations $\mathbf{D}$. For families containing fixed observations (observations at the same locations at each assimilation step) it may be possible to store a single fixed observation error covariance matrix. However, as mentioned earlier, due to quality control procedures and the intermittent nature of most observations, the observation error covariance matrix for each family will change at each assimilation step. It therefore makes sense to derive $\mathbf{C}$ dynamically by simply providing a correlation function and a pre-derived correlation length scale for each type of family. For example $\mathbf{C}$ may be derived using,

$$\mathbf{C}_{i,j} = e^{\left(\frac{-|\Delta y_{i,j}|}{L_r}\right)}. \tag{3}$$

where for a given family, $\Delta y_{i,j}$ is the separation distance between a pair of observations $\mathbf{y}_i$ and $\mathbf{y}_j$ and L_r is the correlation lengthscale. Similarly $\mathbf{D}$ is constructed using pre-derived variance for each family.

After determining the full spatial observation error correlation matrix and matrix of standard deviations, the observation error covariance matrix $\mathbf{R}_f = \mathbf{DCD}$ and the observation penalty (J_o) can be calculated as follows:

1. Calculate a vector of observation minus model equivalent differences $d_o^b = (\mathbf{y}^o - \mathcal{H}(\mathbf{x}))$.
2. Calculate the sensitivity $\mathbf{q} = \mathbf{R}_f^{-1}(\mathbf{y}^o - \mathcal{H}(\mathbf{x}))$ using a Cholesky decomposition (Golub and Van Loan 1996). The Cholesky decomposition avoids the need to invert the observation error matrix. Instead the sensitivity is calculated by first factorizing $\mathbf{R}_f = \mathbf{U}\mathbf{U}^T$, where $\mathbf{U}$ is an upper triangular matrix, then solving for $\mathbf{q}$ using forward, and backward substitution.
3. The total observation penalty J_o for the family is calculated by multiplying the sensitivity by the observation minus model equivalent differences,

$$J_o = \frac{1}{2}\left(\mathbf{y} - \mathcal{H}(\mathbf{x})\right)^T \mathbf{R}_f^{-1}\left(\mathbf{y} - \mathcal{H}(\mathbf{x})\right). \tag{4}$$

The gradient of J_o needed for the variational minimisation is calculated using,

$$\frac{\partial J_o}{\partial \mathbf{x}} = \mathbf{H}^T\mathbf{R}_f^{-1}\left(\mathbf{y}^o - H\left(\mathbf{x}\right)\right). \tag{5}$$

4. Experimental details

The model used in this study is the operational UKV model. It is a variable-resolution version of the nonhydrostatic UM (Davies *et al.* 2005), allowing an explicit representation of convective processes as described in Lean *et al.* (2008). The horizontal grid has a 1.5-km fixed resolution on the interior surrounded by a variable-resolution grid that increases smoothly in size to 4 km and has 70 vertical levels. The variable-resolution grid allows the downscaled boundary conditions, taken from the global model, to spin up before reaching the fixed interior grid. The initial conditions are provided from the operational 3 hourly 3D variational assimilation scheme that uses an incremental approach (Courtier *et al.* 1994) and is a limited-area version of the Met Office variational data assimilation scheme (Lorenc *et al.* 2000; Rawlins *et al.* 2007). The assimilation uses a vertical adaptive mesh that allows the accurate representation of boundary layer structures (Piccolo and Cullen 2011, 2012).

The background error covariance has been derived using the Covariances and VAR Transforms software (Wlasak and Cullen 2014), which is Met Office covariance calibration and diagnostic tool that analyses training data representing forecast errors using the National Meteorological Center (NMC) lagged forecast technique or ensemble perturbations. Here an NMC method has been applied to $(T+6h)-(T+3h)$ forecast differences to diagnose variances and correlation length scales.

For this study, we are using a 3DVar analysis system with first guess at appropriate time (FGAT). The background field is provided by a $T+3$ forecast (actually time interpolated to observation time using fields every 30 minutes in the $3h$ observation window for FGAT). In addition to Doppler radial winds used at the centre of the assimilation window, the analysis uses hourly surface synoptic observations of temperature, wind, pressure, humidity and visibility, hourly wind profiler data, hourly satellite radiances, satellite winds and aircraft data, radiosonde and hourly GPS water vapour paths (note that hourly referes to the frequency usage of the observation) . Radar-derived surface precipitation rates available every 15 min are included via latent heat nudging from T-0.5 h to T+0.5 h and hourly cloud-cover-derived 3D relative humidity profiles via moisture nudging (Jones and Macpherson 1997; Dixon *et al.* 2009). The nudging was done over a period surrounding the analysis time, in addition to incremental analysis updating of the 3D-Var analysis increments.

The Doppler radial winds are provided by 18 Doppler Weather radars spread over the United Kingdom. Each radar produces 5 plan position indicator (PPI) scans every 10 minutes. The Doppler radial winds are assimilated using a simple observation operator where the horizontal model background winds are projected onto the slant of the radar beam (vertical motion is ignored) (Simonin *et al.* 2014). To reduce the density of the observations, multiple observations are made into a single super-observation (3^o by $3km$) and then thinned using Poisson disk sampling, as described in Simonin *et al.* (2014).

The observation error correlation matrices are calculated dynamically as described in section 3. In the correlation matrices we only consider horizontal correlations; we neglect vertical correlations as there are unlikely to be multiple observations, and hence vertically correlated errors, in a single model column. Instead we assume that observation errors are correlated only if the observations are within a height band as described in Waller *et al.* (2016c). This assumption results in a sparse block diagonal observation error correlation matrix. Using this approach the number of observations in a family cannot exceed 2000. When Doppler radial wind observation errors are assumed uncorrelated, the standard deviations for the control experiment are based on those described in Simonin *et al.* (2014) and evolve with range, whereas when correlation is accounted for, the standard deviations and length scales L_r are based on those calculated in Waller *et al.* (2016c). A comparison of the variances from both observation error matrices ($\mathbf{R}$) is shown in figure 2 as a function of height for the 1^o, 2^o and 4^o radar beams. The length scales L_r have been determined by fitting Markov functions (eq. 3) to the estimated horizontal correlations. We note that the length scales L_r are dependent on both the height of the observation and the radar beam elevation. Neither the prescribed variances nor length scales differ between radars. However, due to the intermittent nature of the observations, the observation error covariance matrices do differ between radars; similarly, for any given radar the error covariance matrices differ at each assimilation time.

The impact of including horizontally correlated Doppler radial wind errors was investigated by running three experiments using data for the period 1-20 May 2016. As shown in table 1, the Control experiment uses a diagonal observation error matrix, whereas both experiments Corr-R-3km and Corr-R-6km use a correlated observation error matrix. The Control run and the Corr-R-6km experiment use a $6km$ thinning distance whereas the Corr-R-3km experiment uses a $3km$ thinning distance. We note that the Control run and the Corr-R-6km experiment use the same set of observations; therefore, comparisons between these experiments allow us to determine the impact of including spatially correlated observation errors in the system. Comparisons between the Corr-R-6km and Corr-R-3km experiments allow us to assess the impact of including denser observations (permitted by the inclusion of the correlated errors). Results from an additional experiment using the control's set-up with 3km thinning instead of 6km will be presented periodically to add context. This experiment (Diag-R-3km) is known to be sub-optimal with the analysis degraded compared to the control. Comparison to the other experiments will positively bias the impact of correlated observation error; therefore, the authors limit the discussion of this experiment in the manuscript.

5. Initial Results

The initial impact of including the correlated observation error when assimilating Doppler radial wind has been assessed in three ways. First, we consider the computational performance of the system and its operational viability. Then we consider the relative impact on analysis and innovation accuracy by considering observation-minus-analysis and observation-minus-background statistics. Finally general forecast performance and specific quantitative precipitation forecast verification are presented.

5.1. *Variational data assimilation system performance*

This section focuses on the performance of the variational data assimilation system (VAR) during the trial.

Both the Control and Corr-R-6km experiments used a thinning distance of $6km$, which yield an average of 2000 Doppler radial wind observations per cycle. The Corr-R-3km experiment, however, use a reduced thinning distance of $3km$, which provides on average four times more (8000) Doppler radial wind observations per assimilation cycle.

Table 2 shows the average iteration and evaluation count for each experiment. The iteration and evaluation count from each run are very similar. (Note that one evaluation is one calculation of the penalty function, and one iteration is equivalent to one cycle of the minimisation algorithm). This result is most significant when considering that the Corr-R-3km experiment used four times more Doppler radial wind observations. When comparing the mean iteration/evaluation count to the standard deviation we find that for all experiments there are substantial differences observed between different assimilations. The large variance is expected since we are using a regional data assimilation system where the total number of observations can change significantly depending on the time of assimilation (e.g. day vs. night). We note that when comparing timeseries of iteration/evaluation counts there are minimal differences between the three experiments (not shown) and all follow a diurnal cycle.

Table 2 also shows the average and standard deviation of the of observation and background penalty values (J_o and J_b respectively). The changes in the mean value of J_o and J_b suggest that the overall observation weight is reduced and more importance is given to the background information as shown from theoretical studies by Seaman (1977) or Stewart *et al.* (2008). This is evident when Corr-R-6km is compared to the Control as both experiments use the same observation count. Corr-R-6km has an increased (reduced) observation (background) penalty. As values of J_o and J_b are affected by the observation count, Corr-R-3km needs to be compared to a Control experiment using $3km$ thinning (Diag-R-3km). The comparison of Corr-R-3km with Diag-R-3km gives similar results to the Corr-R-6km/Control comparison. The mean values of J_o and J_b for Diag-R-3km are equal to 9679.28 and 2277.59 respectively, whereas for Corr-R-3km these values are equal to 10134.63 and 2050.53. The decrease in background penalty between Diag-R-3km and Corr-R-3km more or less matches the increase in observation penalty between the two experiments.

Table 3 shows the performance of the assimilation over the trial period, as well as over 10 iterations, for the three experiments. Comparing the experiments we see that the increased communication did not impact on the performance of the code. The cost of computing J_o is minimal compared to that of J as wells as the wall-clock time. The cost of J_o remains minimal and there is little change in the total cost of J even when correlated observation error are used and observation count is increased.

Overall results show that the proposed strategy to introduce correlated observation error statistics does not diminish the computational performance of the assimilation system. Furthermore, denser observations with correlated errors can be assimilated without increasing the computational cost.

5.2. *Impact on the analysis*

Residual $(O - A)$ and innovation $(O - B)$ statistics provide a quantitative measurement of the impact of the correlated observation error upon the analysis for individual observation types. The $O - A$ are retrieved from the assimilation system as the residual at the end of the minimisation. First, we note that the mean bias from the innovation or the residual for this Doppler radial wind will always tend toward zero for a large quantity of observations due to the radial nature of the observation (Salonen *et al.* 2007). Instead figure 3 shows the $O - B$ and $O - A$ standard deviation (σ_{O-B} and σ_{O-A} hereafter) from the Control, Corr-R-6km and Corr-R-3km as a timeseries for each cycle over the of the trial. The time series of Doppler radial wind's σ_{O-B}, yield similar results, with mean σ_{O-B} over the length of the trial for the Control, Corr-R-6km and Corr-R-3km is equal to 2.77, 2.76 and 2.73 respectively. Here the background is a $T + 3$ forecast from the previous cycle (3 hourly data assimilation system). More pronounced differences between the control and the experiments are visible in the Doppler radial wind's σ_{O-A} (figure 3). In the case of Corr-R-6km (figure 3-a), the values of σ_{O-A} are consistently slightly higher than those for the Control. In the case of Corr-R-3km (figure 3-b) the σ_{O-A} are comparable to the values yielded by the Control.

The differences in σ_{O-A} between the two runs with identical observation count (i.e. Control and Corr-R-6km) confirm the results of the previous section. Despite the fact that the observation error matrix used in Corr-R-6km had smaller or equivalent variance compared to those prescribed for the Control experiment (figure 2), the weight of the Doppler radial wind observations was reduced in Corr-R-6km. This in turn reduces the fit to the observations and increases analysis error. This increase in analysis error is not seen in the Corr-R-3km's experiment where the σ_{O-A} shows similar values compared to the Control. The reduction in the observation weighting has been reversed by the increased observation count. This is supported by considering the additional Diag-R-3km experiment, where σ_{O-A} is consistently lower (mean value of 1.20) compared to the Control (mean value of 1.57).

Before considering the impact on other observation types, we first consider how the structured wind increments may have been modified by the introduction of correlated observation error. Figure 4 shows the mean length scale, the mean variance and maximum values of the zonal wind increment at each model level over the trial. Length scale is simply defined as the fourth root of the ratio of the variance of a field (ϕ) and the variance of its Laplacian (calculated using a second-order finite difference approximation) (Descombes *et al.* 2015); that is

$$\mathbf{L} = \left(\frac{8 \cdot variance(\phi)}{variance(\nabla^2 \phi)} \right)^{1/4}. \tag{6}$$

We note that the mean increment can be related to systematic error in either observations or the model (Rodwell and Palmer 2007). However it has been shown that the Doppler radial wind observations used here are unbiased (Simonin *et al.* 2014). When the correlated observation error covariance matrix is introduced (Corr-R-6km) the zonal wind increment becomes smoother with smaller extremes at all model levels. The introduction of correlation acts as a low-pass filter, reducing the weight from individual observations and increasing the importance of the background information. This is consistent with the results from the σ_{O-A}. However, increasing the observation density (Corr-R-3km) counterbalances the effect of the correlated $\mathbf{R}$, by increasing the amplitude and the variance of the increment values at all levels so that the values are closer to the Control experiment. It produces increments with smaller length scales than the Control from the assimilation of denser observations which are more able to represent smaller scale features.

We now consider the impact from the introduction of the correlated observation error covariance matrix for the Doppler radial wind on the fit to other observations assimilated during the trial. Figure 5 shows the trial average ratio of σ_{O-A} and σ_{O-B}, between the experiments and the Control for all the wind observation types used in the trial. The error bars shown in figure 5 and subsequent figures, represent the 95% confidence level. Due to high statistical variability between cycles, one should only consider the significant values to assess the impact. Most trial average innovation and residual standard deviations from the Corr-R-6km and Corr-R-3km yield smaller values compared to the Control, with Corr-R-3km outperforming Corr-R-6km. This trend is not completely homogeneous with, for example, mixed impact for the scatterometer wind (not statistically significant). Although not significant, the $O-A$ and $O-B$ from Corr-R-6km, exhibit a degradation for meridional wind from Sonde and Aircraft respectively. Note that Sonde U and V account for radiosonde profiles as well as wind-profiler observations. For all wind observations, the additional Diag-R-3km produces innovation and residual values (figure 5) that never improve upon the results of either the Corr-R-6km or Corr-R-3km. Furthermore, the innovation and residual values are significantly worse compared to the Control, with a decrease in analysis and background accuracy reaching 1% and more on a few occasions. For example, compared to the Control, U and V wind from Sonde are degraded by 2% in the σ_{O-A} and at least 0.5% in the σ_{O-B}. Note that Corr-R-3km shows an improvement of 0.5% in the σ_{O-A} and at least 0.75% in the σ_{O-B}.

The reduction in analysis error and improved innovations are equally visible when considering the results from satellite observations (figure 6). Again the general impact is stronger for Corr-R-3km. For the rest of the surface and upper-air observations (figure 7), the impact seen when considering Corr-R-6km is very much neutral, whereas Corr-R-3km still shows benefit. The statistics for relative humidity and potential temperature observations from sonde are neutral to negative in the σ_{O-A} (7-a), but improve in the σ_{O-B} (7-a). Again the additional Diag-R-3km (not shown) produces residual values that are worse compared to the Control with a maximum increase of 3% for relative humidity. Also, the σ_{O-B} do not outperform the Corr-R-6km or Corr-R-3km experiments. For satellite observations, Diag-R-3km has small overall improvement in comparison to the Control (0.3% in σ_{O-B}). However, Corr-R-6km or Corr-R-3km experiments are still significantly superior.

We now summarize the results from this section. From the analysis of $O-B$ and $O-A$ it is clear that the introduction of correlated observation error for the Doppler radial winds had a general benefit in reducing the analysis error. From the results of the residual statistics (σ_{O-A}) and the shape of the wind increments, we see that the introduction of correlated observation error has a multi faceted effect (Daley 1991). The changes in the σ_{O-A} from Corr-R-6km experiment compared to the Control, as well as the observation and background penalty values, demonstrated that the Control experiment settings were producing an analysis that was over-fitting the Doppler radial wind. The use of a diagonal observation-error covariance matrix when observation errors are clearly horizontally correlated (Waller *et al.* 2016c) produced a suboptimal analysis (Liu and Rabier 2003). When the observation errors are correlated with a length scale of 20-30km (Waller *et al.* 2016c), thinning the data to a 6km spacing does not result in negligible error correlations between assimilated observations. By introducing correlated observation error statistics in the assimilation algorithm (Corr-R-6km), the assimilation algorithm acts like a low-pass filter on the observation increments. Reducing the thinning distance shows benefit only when the correlation in the observation errors are accounted for as demonstrated by results from Corr-R-3km. Omitting the correlation when using a dense network of observations, only produces a sub-optimal system, where dense observations are over-fitted and the general analysis error is increased. By contrast accounting for correlation when using a dense network of observations, increases the potential number of neighbour observations y_j to an observation y_i, allowing for synergy between more pairs of observations, as described by Fowler *et al.* (2018), as well as allowing the information content from smaller scales to be exploited. This transforms the assimilation algorithm and allows it to behave more like a high pass filter compared to the Corr-R-6km setting.

We support these results using simple model experiments (details are given in the appendix). We designed three experiments to imitate the changes in observation density between the Control, Corr-R-6km and Corr-R-3km experiments. Figure 8 shows the eigenvalues of the analysis error covariance matrix in observation space for the three simple model experiments:

- The Control experiment is qualitatively similar to the simple model experiment shown as a black curve in figure 8. Here the simple model is using a diagonal observation error covariance but the true observation error covariance contains some correlation and the state is half observed.

- The Corr-R-6km experiment has a similar character to the simple model experiment shown as a gray curve in figure 8, where a correlated observation error covariance is used and the state is half observed.
- The Corr-R-3km experiment is qualitatively similar to the simple model experiment shown as a black dashed curve in figure 8. A correlated observation error covariance is used and the state is fully observed.

Figure 8 shows that assimilating observations with the correct error statistics reduces the analysis uncertainty at all scales compared to the case when the observation error correlations are neglected. However, in the case where the observation density is coarse, most of the reduction in analysis uncertainty is seen at large scale (grey curve of figure 8). Increasing the observation sampling when correlated observation errors are used, further reduces the analysis uncertainty. However, this time the additional reduction in uncertainty takes effect at small scales (dashed curve of figure 8), which is consistent with the analysis of our experiments.

5.3. *Forecast performance*

This section focuses on the impact on the forecast from the introduction of Doppler radial wind's correlated observation error.

5.3.1. *Overall forecast performance*

In order to quantify forecast skill of a variable such as temperature, wind or cloud cover it is possible to check the root mean square (RMS) or the equitable threat score (ETS) difference (Ebert *et al.* 2003) between an observed quantity and its forecast equivalent at a range of lead times, from T+6 to T+36 at 6-hour intervals. The forecast value at observation locations is calculated from a simple bilinear interpolation of the forecast taking a distance-weighted average of the four surrounding grid point values. From the values derived following the above process an index that summarizes this skill score can be determined so as to make it easier to tell how a given trial experiment is performing with respect to the Control.

The Met Office's UK NWP Index is defined as a weighted average of T+6 to T+36 skill scores over the UK domain, for 1.5m temperature, 10m wind (speed and direction), precipitation (equal to or greater than 0.5, 1.0 and 4.0 mm over the preceding 6 hours), total cloud amount (equal to or greater than 2.5, 4.5 and 6.5 oktas), cloud base height (given at least 2.5 oktas and equal to or less than 100, 500 and 1000 m above ground) and near-surface visibility (equal to or less than 200, 1000 and 4000 m).

Table 4 shows the results of the surface verification as percentage of improvements. For the Corr-R-6km UK NWP index changes by -0.005% compared to the Control run. For the Corr-R-3km UK NWP index changes by $+0.021\%$ compared to the Control run. Neither trial presents statistically significant differences in skill with respect to the Control run.

5.3.2. *Impact on precipitation*

So far we have concentrated our effort on the validation of the forecast performance overall. However, one of the main motivations of convective-scale assimilation is to improve short-term quantitative precipitation forecasts (QPF). Verification methods have conventionally been designed to assess the model forecast at point locations only. However, the temporal and spatial intermittent nature of a parameter such as rain makes these approaches unsuitable in general (Droegmeier 1997). This problem is amplified in this study because the 1.5 km model resolution is high enough to represent small-scale features and local variability. In response to this problem, a growing list of methods to verify precipitation forecasts based on the physical realism or spatial closeness to observations have been developed (Gilleland *et al.* 2009). Some techniques have concentrated on object verification (Ebert and MacBride 2000; Davis *et al.* 2006; Johnson and Wang 2013) by classifying rain features according to their characteristics. Other methods have focused on the spatial error and one such score is the Fractions Skill Score (FSS) introduced by Roberts and Lean (2008). The FSS provides a measure of the spatial agreement between two fields by comparing the fractional differences in the coverage of rain over differing sized squares (neighborhoods) centered at every grid square. More about the definition and use of the FSS can be found in the original paper by Roberts and Lean (2008) and then subsequently in Roberts (2008), Mittermaier and Roberts (2010) or Skok (2015). Here the two fields of hourly accumulations of surface precipitation are from the forecast itself and the radar composite of derived rain rate.

Figure 9 shows the difference in fraction skill score between the experiments (Corr-R-6km: figure 9-a; Corr-R-3km: figure 9-b) and the Control as a Hinton diagram for different forecast lead times and different thresholds of hourly rainfall accumulation. The sign and the amplitude of the change in FSS values (ΔFSS) are shown with the color and size of the square respectively: positive values (positive impact) are shown as grey squares, whereas negative values (negative impact) are shown as black squares. The introduction of the correlated observation error only (Corr-R-6km) does not show any real impact on precipitation (figure 9-a). The ΔFSS values are small ($max|\Delta FSS| = 0.009$) with an almost homogeneous distribution of positive and negative impact. The results are more promising when the correlated observation error is introduced in association with an increase in observation density (figure 9-b). The ΔFSS values are larger compared to the previous comparison, but more importantly, a positive impact can be seem until $T + 7$ forecast lead time. The biggest positive impact is found for low threshold values (e.g. $0.2mmh^{-1}$ and 90^{th} percentile).

Note that the FSS values for the Control forecasts, for a neighbourhood size of 41 grid boxes, were all well above 0.6 indicating an already skillful forecast; although little impact can be seen over the entire period of the fully cycled trial, individual cycles do show some improvements. Figure 10 gives an example of the sort of differences that can be seen and shows an hourly accumulated precipitation $T + 3$ forecast valid at 1500 UTC on the 7th of April 2016, for Control, Corr-R-6km, and Corr-R-3km. During the 7th of April 2016, a band of showers developed and moved southwards, producing heavy precipitation on the east and central part of the UK. Compared to the observed radar derived hourly rain accumulation (figure 10-a), the Control (figure 10-b) produced showers that were typically too sparse and locally far too intense. The Corr-R-6km improved the shower coverage, but the real benefit of including correlated observation error is visible in the Corr-R-3km experiment (figure 10-d), where shower coverage and intensity was noticeably improved. This is supported by the FSS value shown in figure 10-e and f.

The improvement seen in this particular forecast can be attributed to the change in observation weight. When accounting for correlated observation error the observation uncertainty information is no longer mutually independent. This results in a small down-weighting of the observations as demonstrated by the Corr-R-6km experiment (Figure 4). This effect results in a small benefit to the forecast and FSS (Figure 10c and e) as the Control experiment was over-fitting the Doppler radial wind producing broad analysis increments (Figure 4). Increasing the observation density in conjunction with correlated observation errors negates the smoothing effect seen in Corr-R-6km. The use of more accurate error statistics enables an improved representation of the small scale information content from the observation resulting in a more balanced analysis increment (Figures 4 and 8). Over time the small scale information propagates through the system and produces improved forecasts as seen in Figures 10d.

6. Conclusions

In this work, we provide a pragmatic strategy that allows the use of correlated observation errors in a high dimensional data assimilation system. We describe the implementation of this strategy in the Met Office system and then present a study demonstrating the practical feasibility of including horizontally correlated Doppler radial wind observation error statistics and its impact using an operational NWP system. The new strategy was achieved by altering the usual Data Parallel paradigm; rather than distributing observations with correlated errors using a domain decomposition, the observations are instead distributed in families that have mutually correlated errors as described in section 3.1. The second significant change relates to the actual use of the horizontally correlated observation errors statistics in the derivation of the observation penalty. This was implemented following the description presented in section 3.2.

A trial has been run for 20 days using the Met-Office UKV model configuration and 3D-Var, including a Control experiment with the operational settings (diagonal $\mathbf{R}$), an experiment using the operational settings with a correlated observation error covariance matrix, and an experiment using correlated observation error statistics with increased observation density.

Analysis of the impacts from the introduction of Doppler radar radial wind horizontal correlated observation errors on the data assimilation system and forecast skill have also been presented. The introduction of correlated observation error has changed the response from the variational system as follows:

- Introducing horizontal correlated observation errors improves the $O-A$ and $O-B$ statistics in both experiments. This suggests that the Control settings with a diagonal observation error covariance matrix causes the assimilation to over-fit Doppler radial observations.
- The positive impact of the introduction of correlated observation error is stronger on the $O-A$ and $O-B$ statistics when dense observations are used. We showed that introducing the correlated observation error, whilst keeping the observation density fixed, has little impact on the analysis uncertainty at the small scales. However this is remedied by increasing the observation density that introduces additional observation information at small scales.
- The inclusion of correlated observation error statistics allows dense observations to be assimilated without detriment to the analysis quality.
- We showed that by accounting for the correlation in the Doppler radial wind observation error, observation density can be increased without any degradation to the computational speed of the assimilation system.

Our results suggest that the use of a diagonal $\mathbf{R}$ (Control experiment) created a suboptimal system, where a $6km$ observation thinning distance was too dense (e.g. Liu and Rabier (2003) or Stewart *et al.* (2008)). As shown for example by Daley (1991) or Fowler *et al.* (2018) the system's responses to correlated observation error are complex and make use of observation information at specific scales. However, we showed that the general behavior of the data assimilation system is comparable to what can be expected using a simple model.

The impact on the forecast is more subtle. A small positive signal can be seen when the observations are compared to the model background within the assimilation system. This indicates that the impact on the forecast lasts long enough to improve the model background and consequently benefit the assimilation in a cycling NWP system. Regarding the conventional verification scores, the results indicate that over the length of the forecast there is a small positive impact, if any. A stronger signal is visible in the QPF scores. A positive impact can be seem until a forecast lead time of $T+7$. The biggest positive impact is found at low threshold values, which implies an improvement in the location of the rain. For all verification scores, the increase in the observation density yields better results.

To the best of our knowledge this is the first operational implementation of horizontal correlation observation errors in a data assimilation system for numerical weather prediction. Despite a marginal impact on the forecast, the introduction of the correlated observation error allows the assimilation to make better use of the observations by allowing the assimilation of very dense observation networks, such as radar, without any cost (no significant increase of wall clock time) to the assimilation. We note that we have only considered the impact for a single case study (20 days). Furthermore, the only alteration in the experiments has been the inclusion of the correlated observation errors. Further studies are required to analyse the impact for different meteorological conditions. Improved settings for operational parameters associated with Doppler radial wind assimilation may also benefit the forecast. This may include testing for statistical consistency of background and observation errors using the diagnostic of (Desroziers *et al.* 2005). In addition, since this work, the Met-Office operational system for convective scale numerical weather prediction system has been upgraded to 4D-VAR. Therefore this system is now being extended to the 4D-VAR framework.

Appendix

Here we present a simple example to help explain the results given in Section 5.

In statistical linear estimation theory, the analysed state, x^a, is given by

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a = x^b + \widetilde{\mathbf{K}}\mathbf{d}_b^o, \qquad (7)$$

where $\delta\mathbf{x}^a$ is the analysis increment, x^b is the background state, $\mathbf{d}_b^o$ the innovation vector,

$$\mathbf{d}_b^o = y^o - \mathbf{H}x_b, \tag{8}$$

and $\widetilde{\mathbf{K}}$ the gain matrix,

$$\widetilde{\mathbf{K}} = \widetilde{\mathbf{B}}\mathbf{H}^T(\mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}})^{-1}. \tag{9}$$

The matrices $\widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{R}}$ are the assumed background and observation error covariance matrix respectively.

To understand the impact on the analysis of using a sub-optimal observation error correlation matrix we consider the analysis error covariance matrix, $\mathbf{A}$. If we know the exact background error statistics, $\widetilde{\mathbf{B}} = \mathbf{B}$, but are knowingly using an incorrect observation error covariance matrix, $\widetilde{\mathbf{R}} \neq \mathbf{R}$ then the analysis error covariance matrix is given by,

$$\mathbf{A} = (\mathbf{I} - \widetilde{\mathbf{K}}\mathbf{H})\mathbf{B} + \widetilde{\mathbf{K}}(\mathbf{R} - \widetilde{\mathbf{R}})\widetilde{\mathbf{K}}^T. \tag{10}$$

We consider the impact on the analysis error covariance using three simple model experiments. We assume that our background is evaluated on 128 equally spaced points in a 1D periodic domain, (-32π,32π]. In order to compare with the results given in Section 5 we consider two different observation operators, one in which the full state is observed and one where the state at every other grid point is observed. We further assume that the true observation and background error statistics are homogeneous and are defined, as in Waller *et al.* (2016b), using a second order auto regressive function with length scales 5 and 10 respectively. For our first experiment we assume that we observe half the state and only know the observation error variance and hence neglect the correlations i.e. $\widetilde{\mathbf{R}} = \mathbf{I}$. For the second experiment we observe half the state, but this time use the correct $\mathbf{R}$ matrix. Finally we increase the observation density and observe all grid points and assume the correct $\mathbf{R}$ matrix.

In all three experiments the matrices $\mathbf{R}$, $\widetilde{\mathbf{R}}$ and $\mathbf{HBH}^T$ are circulant matrices. Since the sums, products and inverses of circulant matrices are circulant, $\mathbf{HAH}^T$ is also circulant. The eigenvalues of circulant matrices are positive and can be found using a discrete Fourier transform and consequently may be ordered according to wave number. In this case the order of the eigenvalues has a relation to the scales in the analysis error in observation space. Therefore, the eigenvalues of the analysis error covariance in observation space allows us to understand the uncertainty we have at different scales in the analysis in observation space. The k^{th} eigenvalue, ϕ_k, of a circulant matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$ associated with frequency $\omega_k = \frac{2\pi k}{\Delta_x N}$, and sampling interval Δ_x, is,

$$\phi_k = \sum_{n=0}^{N-1} c_n e^{-\frac{2\pi k n i}{N}}, \tag{11}$$

where c_n is the n^{th} coefficient of the first row of the circulant matrix. In our experiments, due to the different number of observations, the size of $\mathbf{HAH}^T$ changes. However, by analysing the results as a function of wavenumber we can compare physically consistent quantities. The results for our experiments are plotted in Figure 8 and discussed in Section 5.

Acknowledgement

Susan P. Ballard passed away after a long illness on 12 July 2018, during the manuscript drafting process.

J. A. Waller, S. L. Dance and N. K. Nichols were funded in part by the NERC grants NE/K008900/1 Forecasting Rainfall exploiting new data Assimilation techniques and Novel observations of Convection (FRANC) and Towards END-to End flood forecasting and a tool for ReaL-time catchment susceptibilitY (TENDERLY) and in part by the NERC National Centre for Earth Observation.

The authors would also like to thank Alison Fowler, Ross Bannister, Andrew Lorenc and Stephen Oxley for their valuable discussions.

The data used in this study may be obtained on request, subject to licensing conditions, by contacting the corresponding author.

References

Ballard SP, Li Z, Simonin D, Caron JF. 2016. Performance of 4D-Var NWP-based nowcasting of precipitation at the Met Office for summer 2012. *Quarterly Journal of the Royal Meteorological Society* **142**(694): 472–487, doi:10.1002/qj.2665.

Bormann N, Bauer P. 2010. 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**(649): 1036–1050, doi:10.1002/qj.616.

Bormann N, Bonavita M, Dragani R, Eresmaa R, Matricardi M, McNally A. 2016. Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quarterly Journal of the Royal Meteorological Society* **142**(697): 1767–1780, doi:10.1002/qj.2774.

Bormann N, Collard A, Bauer P. 2010. 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**(649): 1051–1063, doi: 10.1002/qj.615.

Buehner M. 2010. Error statistics in data assimilation: Estimation and modelling. In: *Data Assimilation Making Sense of Observations*, Lahoz W, Khattatov B, Mènard R (eds), ch. 4, Springer, p. 99.

Campbell WF, Satterfield EA, Ruston B, Baker NL. 2017. Accounting for correlated observation error in a dual-formulation 4d variational data assimilation system. *Monthly Weather Review* **145**(3): 1019–1032, doi: 10.1175/MWR-D-16-0240.1.

Caya A, Sun J, Snyder C. 2005. A comparison between the 4dvar and the ensemble kalman filter techniques for radar data assimilation. *Monthly Weather Review* **133**(11): 3081–3094, doi:10.1175/MWR3021.1.

Clark P, Roberts N, Lean H, Ballard S, Charlton-Perez C. 2015. Convection-permitting models: A step-change in rainfall forecasting. *Meteorological Applications* **23**(2): 165–181, doi: 10.1002/met.1538.

Cordoba M, Dance S, Kelly G, Nichols N, Waller J. 2017. Diagnosing atmospheric motion vector observation errors for an operational high resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **143**(702): 333–341, doi:10.1002/qj.2925.

Courtier P, Thepaut J, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var,using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* **120**(519): 1367–1387, doi:10.1002/qj.49712051912.

Daley R. 1991. *Atmospheric data analysis*. Cambridge University Press: Cambridge, UK.

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the met office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society* **131**(608): 1759–1782, doi:10.1256/qj.04.101.

Davis C, Brown B, Bullock R. 2006. Object-based verification of precipitation forecasts. part i: Methodology and application to mesoscale rain areas. *Monthly Weather Review* **134**(7): 1772–1784, doi:10.1175/MWR3145.1.

de Haan S, Stoffelen A. 2012. Assimilation of high-resolution mode-s wind and temperature observations in a regional nwp model for nowcasting applications. *Weather and Forecasting* **27**(4): 918–937, doi:10.1175/WAF-D-11-00088.1.

Descombes G, Auligné T, Vandenberghe F, Barker DM. 2015. Generalized background error covariance matrix model (GEN_BE v2.0). *Geosci. Model Dev.* **8**(3): 669–696, doi:10.5194/gmd-8-669-2015.

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society* **131**(613): 3385–3396, doi:10.1256/qj.05.108.

Dixon M, Li Z, Lean H, Roberts N, Ballard S. 2009. Impact of data assimilation on forecasting convection over the united kingdom using a high-resolution version of the met office unified model. *Monthly Weather Review* **137**(5): 1562–1584, doi:10.1175/2008MWR2561.1.

Droegmeier K. 1997. The numerical prediction of thunderstorms: Challenges, potential benefit, and results from real-time operational test. *WMO Bulletin* **46**(4): 324–336.

Ebert EE, Damrath U, Wergen W, Baldwin ME. 2003. The WGNE assessment of short-term quantitative precipitation forecasts. *Bulletin of the American Meteorological Society* **84**: 481492.

Ebert EE, MacBride JL. 2000. Verification of precipitation in weather systems: Determination of the systematic errors. *J. Hydrol.* **239**: 179202.

Fisher M, Andersson E. 2001. Developments in 4D-Var and Kalman Filtering. Technical report, ECMWF. ECMWF Technical Memorandum 347.

Fowler AM, Dance SL, Waller JA. 2018. On the interaction of observation and prior error correlations in data assimilation. *Quarterly Journal of the Royal Meteorological Society* **144**(710): 48–62, doi:10.1002/qj.3183.

Gao J, Stensrud DJ. 2012. Assimilation of reflectivity data in a convective-scale, cycled 3dvar framework with hydrometeor classification. *Journal of the Atmospheric Sciences* **69**(3): 1054–1065, doi:10.1175/JAS-D-11-0162.1.

Gilleland EB, Ahijevych D, Brown BG, B C, E EE. 2009. Intercomparison of spatial forecast verification methods. *Weather Forecast* **24**: 1416–1430, doi:10.1175/2009WAF2222269.1.

Golub GH, Van Loan CF. 1996. *Matrix computations*. Baltimore: The Johns Hopkins University Press, third edn.

Healy SB, White AA. 2005. Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quarterly Journal of the Royal Meteorological Society* **131**(605): 63–72, doi:10.1256/qj.03.193.

Ide k, Courtier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan* **75**(1B): 181–189, doi:n/a.

Janjić T, Bormann N, Bocquet M, Carton JA, Cohn SE, Dance SL, Losa SN, Nichols NK, Potthast R, Waller JA, Weston P. 2017. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society* **n/a**(n/a): n/a – n/a, doi:10.1002/qj.3130.

Johnson A, Wang X. 2013. Object-based evaluation of a storm scale ensemble during the 2009 noaa hazardous weather testbed spring experiment. *Mon. Wea. Rev.* **141**(n/a): 1079–1098.

Jones CD, Macpherson B. 1997. A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Applications* **4**(3): 269277.

Lange H, Janjić T. 2016. Assimilation of Mode-S EHS aircraft observations in COSMO-KENDA. *Monthly Weather Review* **144**(5): 1697–1711, doi:10.1175/MWR-D-15-0112.1, URL http://dx.doi.org/10.1175/MWR-D-15-0112.1.

Lean H, Clark P, Dixon M, Roberts N, Fitch A, Forbes R, Halliwell C. 2008. Charictaristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review* **136**(9): 3408–3424, doi:10.1175/2008MWR2332.1.

Liu ZQ, Rabier F. 2003. The potential of high-density observations for numerical weather prediction: A study with simulated observations. *Quarterly Journal of the Royal Meteorological Society* **129**(594): 3013–3035, doi:10.1256/qj.02.170.

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society* **126**(570): 2991–3012, doi:10.1002/qj.49712657002.

Lorenc AC, Jardak M. 2018. A comparison of hybrid variational data assimilation methods for global nwp. *Quarterly Journal of the Royal Meteorological Society* **0**: In press, doi:10.1002/qj.3401.

Lorenc AC, Rawlins F. 2005. Why does 4D-Var beat 3D-Var? *Quarterly Journal of the Royal Meteorological Society* **131**(613): 3247–3257, doi:10.1256/qj.05.85.

Mittermaier MP, Roberts N. 2010. Inter-comparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting* **25**: 343–354.

Pacheco PS. 1997. *Parallel programming with mpi*. Morgan Kaufmann: San Francisco, USA.

Piccolo C, Cullen M. 2011. Adaptive mesh method in the Met Office variational data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656): 631–640, doi:10.1002/qj.801.

Piccolo C, Cullen M. 2012. A new implementation of the adaptive mesh transform in the Met Office 3D-Var system. *Quarterly*

Journal of the Royal Meteorological Society **138**(667): 1560–1570, doi:10.1002/qj.1880.

Rainwater S, Bishop CH, Campbell WF. 2015. The benefits of correlated observation errors for small scales. *Quarterly Journal of the Royal Meteorological Society* **141**(693): 3439–3445, doi:10.1002/qj.2582.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society* **133**(623): 347–362, doi:10.1002/qj.32.

Roberts N. 2008. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an nwp model. *Meteorological Applications, Special Issue: Forecast Verification* **15**(1): 1631697.

Roberts N, Lean H. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review* **136**(1): 78–97, doi:10.1175/2007MWR2123.1.

Rodwell MJ, Palmer TN. 2007. Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society* **133**(622): 129–146, doi:10.1002/qj.23.

Salonen K, Jrvinen H, Eresmaa R, Niemel S. 2007. Bias estimation of doppler-radar radial-wind observations. *Quarterly Journal of the Royal Meteorological Society* **133**(627): 1501–1507, doi:10.1002/qj.114.

Seaman R. 1977. Absolute and differential accuracy of analyses achievable with specified observation network characteristics. *Monthly Weather Review* **105**(10): 12111222, doi:10.1175/1520-0493(1977)105⟨1211:AADAOA⟩2.0.CO;2.

Simonin D, Ballard SP, Li Z. 2014. Doppler radar radial wind assimilation using an hourly cycling 3d-var with a 1.5 km resolution version of the met office unified model for nowcasting. *Quarterly Journal of the Royal Meteorological Society* **140**(684): 2298–2314, doi:10.1002/qj.2298.

Skok G. 2015. Analysis of fraction skill score properties for a displaced rainband in a rectangular domain. *Met. Apps* **22**: 477484, doi:10.1002/qj.2298.

Stewart LM. 2010. Correlated observation errors in data assimilation. PhD thesis, University of Reading. Http://www.reading.ac.uk/maths-and-stats/research/theses/maths-phdtheses.aspx.

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

Stewart LM, Dance SL, Nichols NK. 2008. Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids* **56**(8): 1521–1527, doi:10.1002/fld.1636.

Stewart LM, Dance SL, Nichols NK. 2013. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A* **65**(1): 19 546, doi:10.3402/tellusa.v65i0.19546.

Stewart LM, Dance SL, Nichols NK, Eyre JR, Cameron J. 2014. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Quarterly Journal of the Royal Meteorological Society* **140**(681): 1236–1244, doi:10.1002/qj.2211.

Strajnar B, agar N, Berre L. 2015. Impact of new aircraft observations mode-s mrar in a mesoscale nwp model. *Journal of Geophysical Research: Atmospheres* **120**(9): 3920–3938, doi:10.1002/2014JD022654.

Sun J, Xue M, Wilson JW, Zawadzki I, Ballard SP, Onvlee-Hooimeyer J, Joe P, Barker DM, Li PW, Golding B, Xu M, Pinto J. 2014. Use of NWP for nowcasting convective precipitation: Recent progresses and challenges. *Bulletin of the American Meteorological Society* **95**(3): 409–426, doi:10.1175/BAMS-D-11-00263.1.

Tabeart J, Dance SL, Haben SA, Lawless AS, Nichols NK, Waller JA. 2018. The conditioning of least squares problems in variational data assimilation. *Numerical Linear Algebra with Applications* doi:10.1002/nla.2165.

Tang Y, Lean HW, Bornemann J. 2013. The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications* **20**(4): 417–426, doi:10.1002/met.1300.

Velden C, Lewis WE, Bresky W, Stettner D, Daniels J, Wanzong S. 2017. Assimilation of high-resolution satellite-derived atmospheric motion vectors: Impact on hwrf forecasts of tropical cyclone track and intensity. *Monthly Weather Review* **145**(3): 1107–1125, doi:10.1175/MWR-D-16-0229.1.

Waller JA, Ballard SP, Dance SL, Kelly G, Nichols NK, Simonin D. 2016a. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sensing* **8**(7): 581, doi:10.3390/rs8070581.

Waller JA, Dance SL, Lawless AS, Nichols NK, Eyre JR. 2014. Representativity error for temperature and humidity using the Met Office high-resolution model. *Quarterly Journal of the Royal Meteorological Society* **140**(681): 1189–1197, doi:10.1002/qj.2207.

Waller JA, Dance SL, Nichols NK. 2016b. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Quarterly Journal of the Royal Meteorological Society* **142**: 418–431, doi:10.1002/qj.2661.

Waller JA, Simonin D, Dance SL, Nichols NK, Ballard SP. 2016c. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Monthly Weather Review* **144**(10): 3533–3551, doi:10.1175/MWR-D-15-0340.1.

Wang Y, Wang X. 2017. Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in the gsi-based envar system: Methodology and experiment with the 8 may 2003 oklahoma city tornadic supercell. *Monthly Weather Review* **145**(4): 1447–1471, doi:10.1175/MWR-D-16-0231.1.

Wattrelot E, Caumont O, Mahfouf JF. 2014. Operational implementation of the 1d+3d-var assimilation method of radar reflectivity data in the arome model. *Monthly Weather Review* **142**(5): 1852–1873, doi:10.1175/MWR-D-13-00230.1.

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

Weston PP, Bell W, Eyre JR. 2014. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society* **140**(685): 2420–2429, doi:10.1002/qj.2306.

Wlasak MA, Cullen MJF. 2014. Modelling static 3d spatial background error covariances - the effect of vertical and horizontal transform order. *Advances in Science and Research* **11**(1): 63–67, doi:10.5194/asr-11-63-2014.

Yaremchuk M, D'Addezio JM, Panteleev G, Jacobs G. 2018. On the approximation of the inverse error covariances of high-resolution satellite altimetry data. *Quarterly Journal of the Royal Meteorological Society* : In pressdoi:10.1002/qj.3336.

g

Table 1. Experiment details

Experiment	Doppler wind observation error matrix	Doppler wind super-observation thinning distance
Control	Diagonal observation error matrix (Operational)	6 km
Corr-R-6km	Correlation observation error matrix	6 km
Corr-R-3km	Correlation observation error matrix	3 km
Diag-R-3km	Diagonal observation error matrix (Operational)	3 km

Table 2. Trial average (μ) and standard deviation (σ) of various parameters measuring the performance of the assimilation system.

Experiments	Iteration count		Evaluation count		J_b		J_o	
	μ	σ	μ	σ	μ	σ	μ	σ
Control	27.4	15.1	40.8	23.5	1752.16	526.1	9207.53	2707.16
Corr-R-6km	27.7	14.6	41.6	22.8	1722.43	500.37	9235.66	2732.84
Corr-R-3km	28.2	14.4	40.9	21.6	2050.53	761.25	10134.63	3175.32
Diag-R-3km	29.1	14.9	41.5	23.8	2277.59	910.74	9679.28	2900.41

Table 3. Computational cost in seconds. The first row shows the trial average wall-clock time ($\overline{W}^{(trial)}$). Subsequent rows show the average wall-clock time ($\overline{W}^{(10)}$), the average cost for calculating J ($\overline{J}^{(10)}$), and the average cost for calculating J_o ($\overline{J}_o^{(10)}$), over 10 iterations for the 12 Z run on the 7th of April 2016.

	Control	Corr-R-6km	Corr-R-3km
$\overline{W}^{(trial)}$ [s]	272	293	288
$\overline{W}^{(10)}$ [s]	73	72	73
$\overline{J}^{(10)}$ [s]	22.16	23.83	23.43
$\overline{J}_o^{(10)}$ [s]	0.81	2.21	2.23

Table 4. Surface verification scores and UK NWP index. All the values are presented as a percentage (positive values show improvement over the Control).

Score	Corr-R-6km	Corr-R-3km
Visibility	+0.027	+ 0.046
Precipitation	-0.063	-0.050
Cloud cover	+0.047	+0.012
Cloud base height	-0.013	-0.005
1.5m temperature	-0.014	+0.005
10 m wind	+0.010	+0.013
UK index	-0.005	+0.021

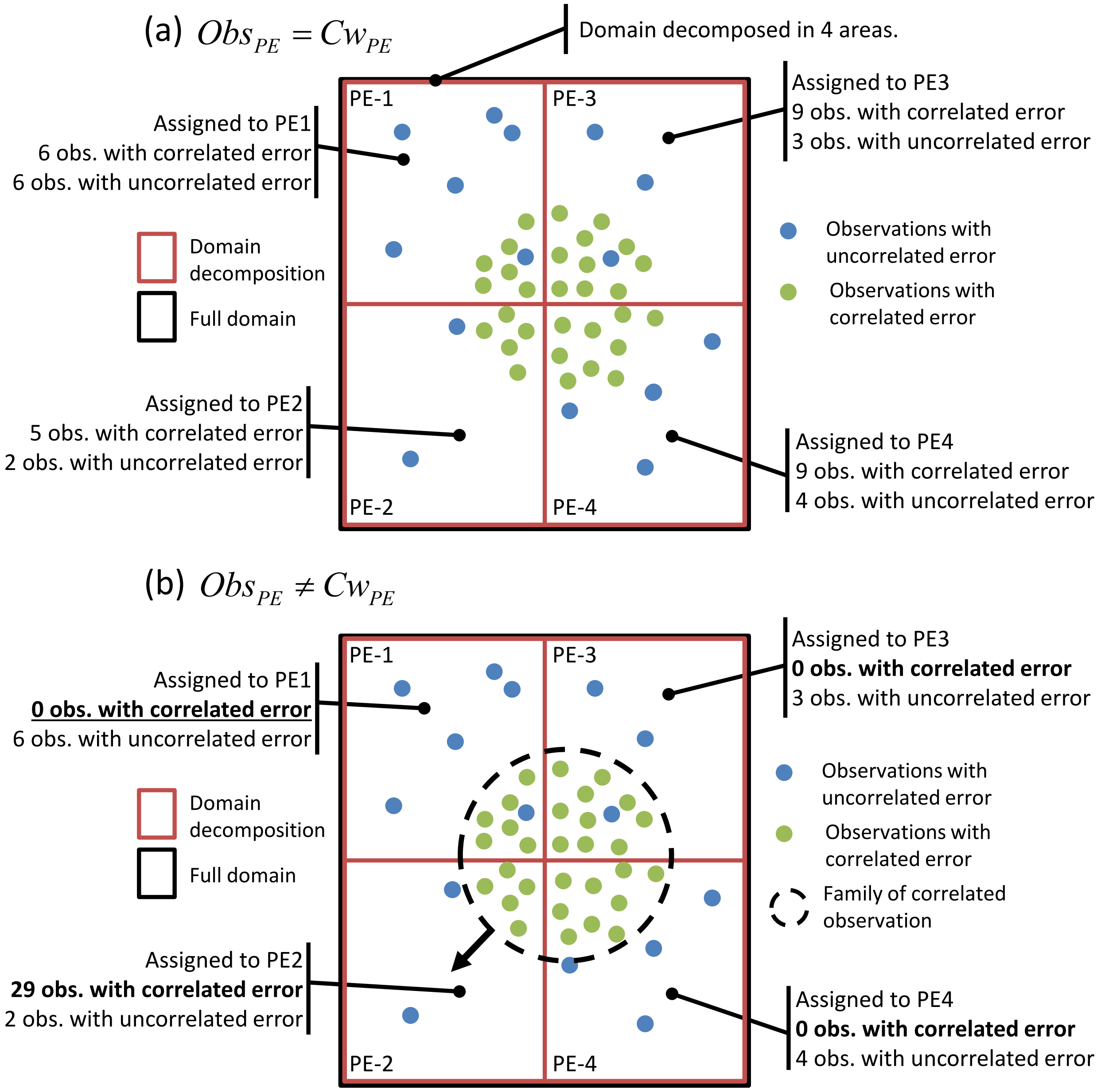

Figure 1. Example of the observation parallelisation in VAR for two observation types (uncorrelated and correlated error) with a 4 PE decomposition. (a) Conventional approach, i.e. without accounting for the horizontal correlation of the observation error. Each observation (with uncorrelated and correlated errors) is distributed between the 4 PE according to its geographical location. (b) The new approach i.e. accounting for the horizontal correlation of the observation error. As before, each observation with uncorrelated error statistics is distributed between the 4 PE according to its geographical location. However, this time all the observations with mutually correlated errors are assigned to a single family and sent to PE 2 regardless of their physical location. In both panels the model columns (C_w) are distributed according to their geographical location. This implies that the distribution of each C_w and observation is equivalent in (a) that is $Obs_{PE} = Cw_{PE}$, and different in (b) i.e: $Obs_{PE} \neq Cw_{PE}$.

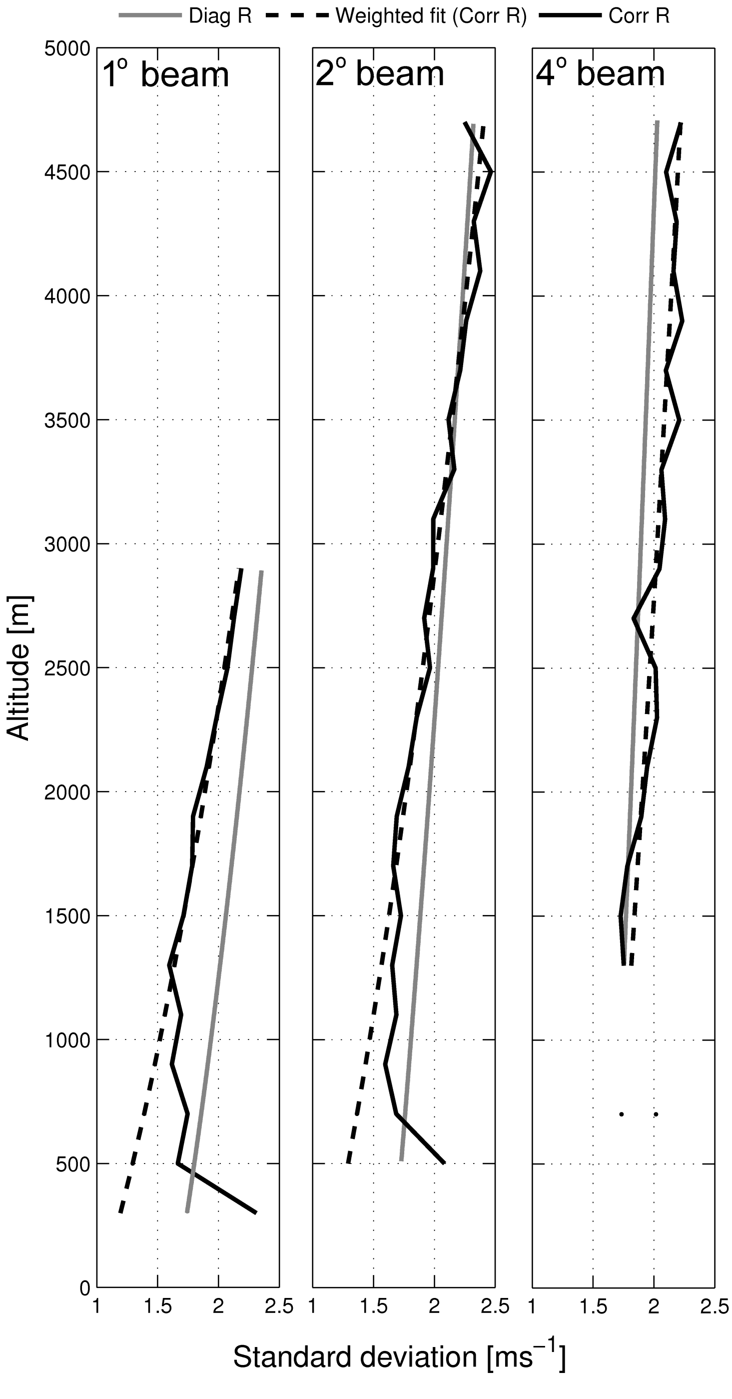

Figure 2. Error variance as function of height for three radar's beam elevation (1^o, 2^o and 4^o). (Grey curve) operational error variance used in the Control experiment when the observation error covariance matrix is assumed to be diagonal. (Black curve) error variance for the diagnosed correlated observation error covariance matrix. (Black dash curve) weighted least square fit of the error variance for the diagnosed correlated observation error covariance matrix used in the Corr-R-6km and Corr-R-3km experiments.

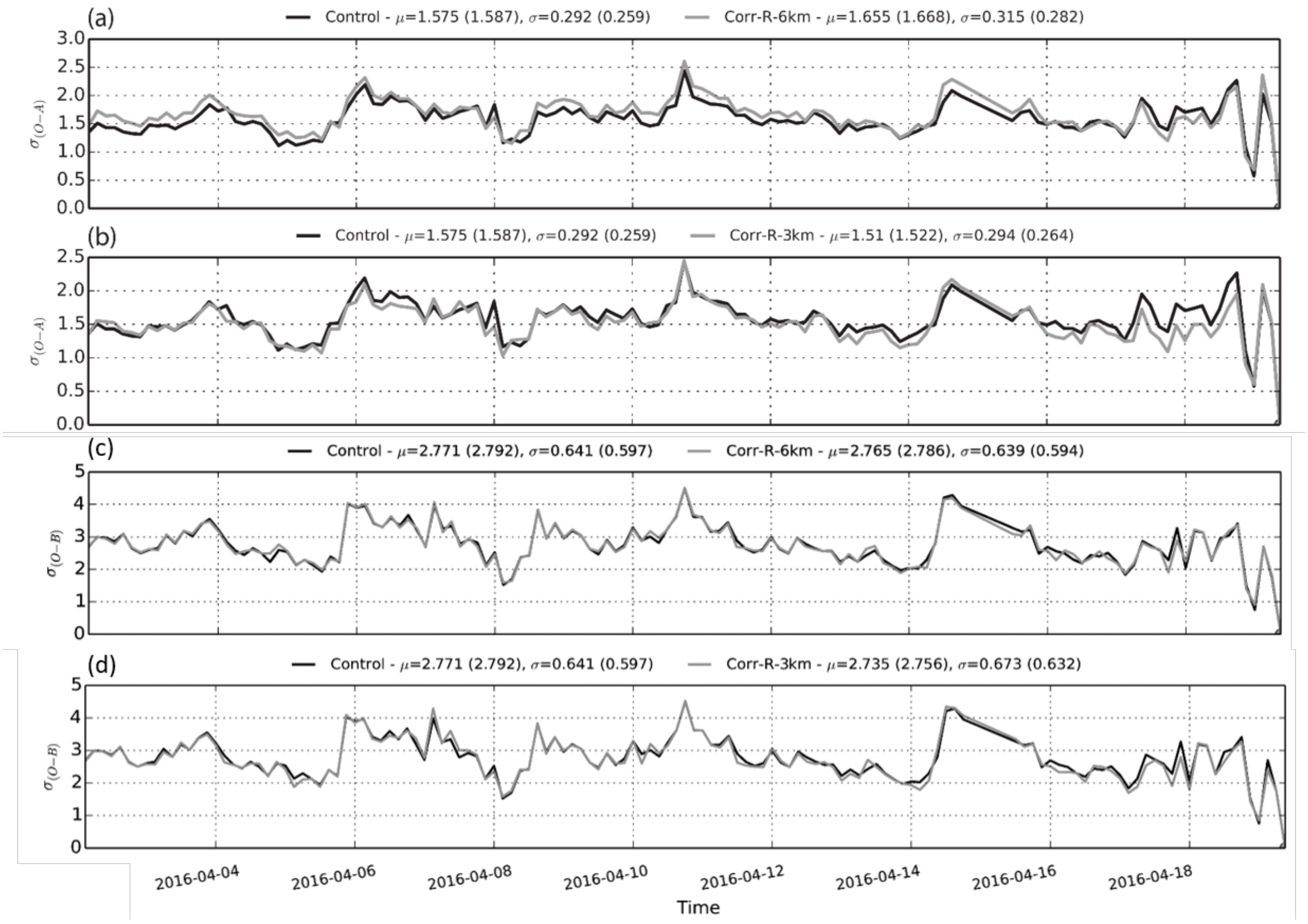

Figure 3. Time series standard deviation of Doppler radial wind $O - A$ (a,b) and $O - B$(c,d) for the Control and Corr-R-6km (a,c) and for the Control and Corr-R-3km (b,d). In both panels the Control is in black and the experiment in grey.

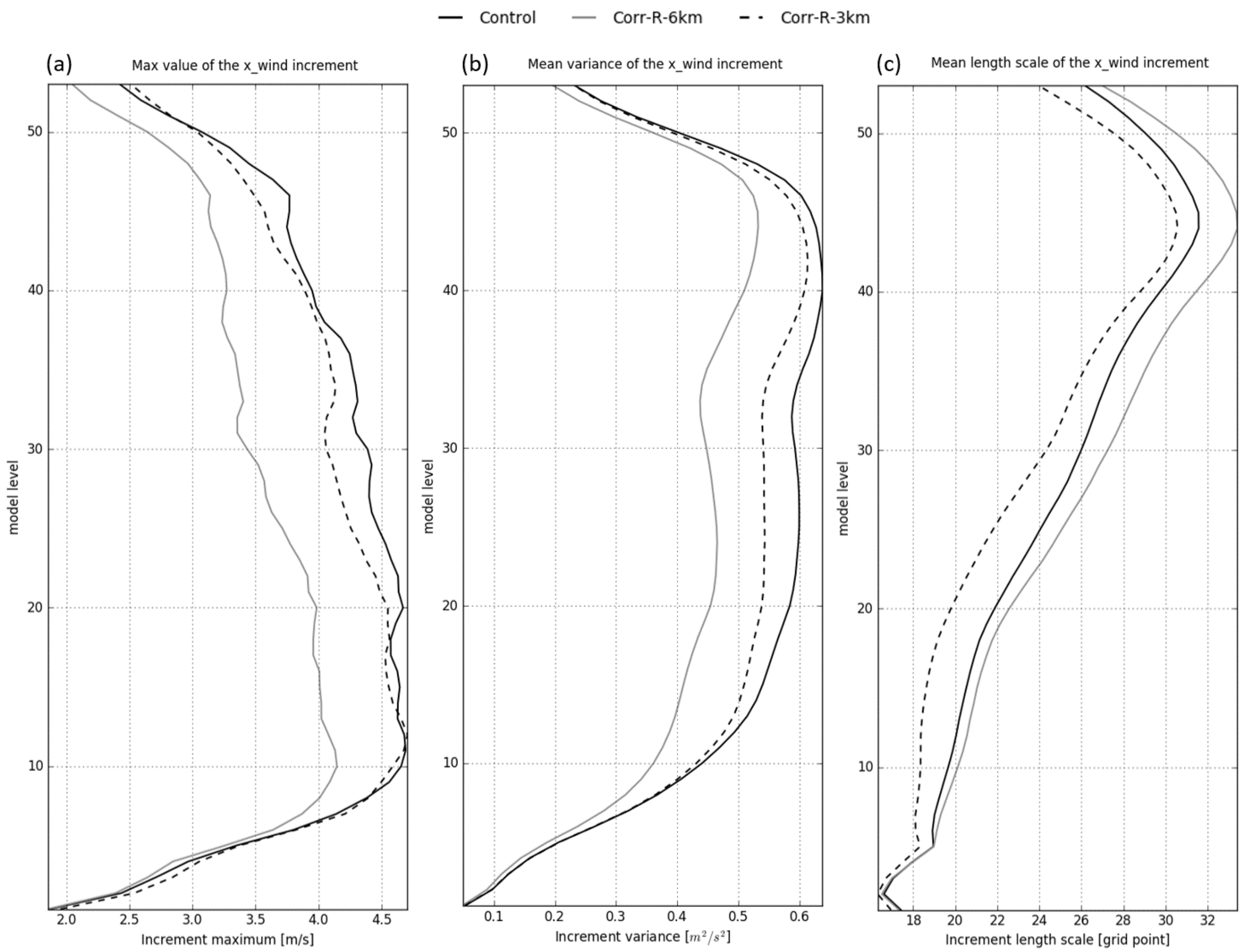

Figure 4. Trial average (a) maximum (b) variance and (c) length-scale (eq 6) for the zonal wind increment against model levels. (Black Curve) Control experiment; (grey curve) Corr-R-6km experiment and (black dash curve) Corr-R-3km experiment.

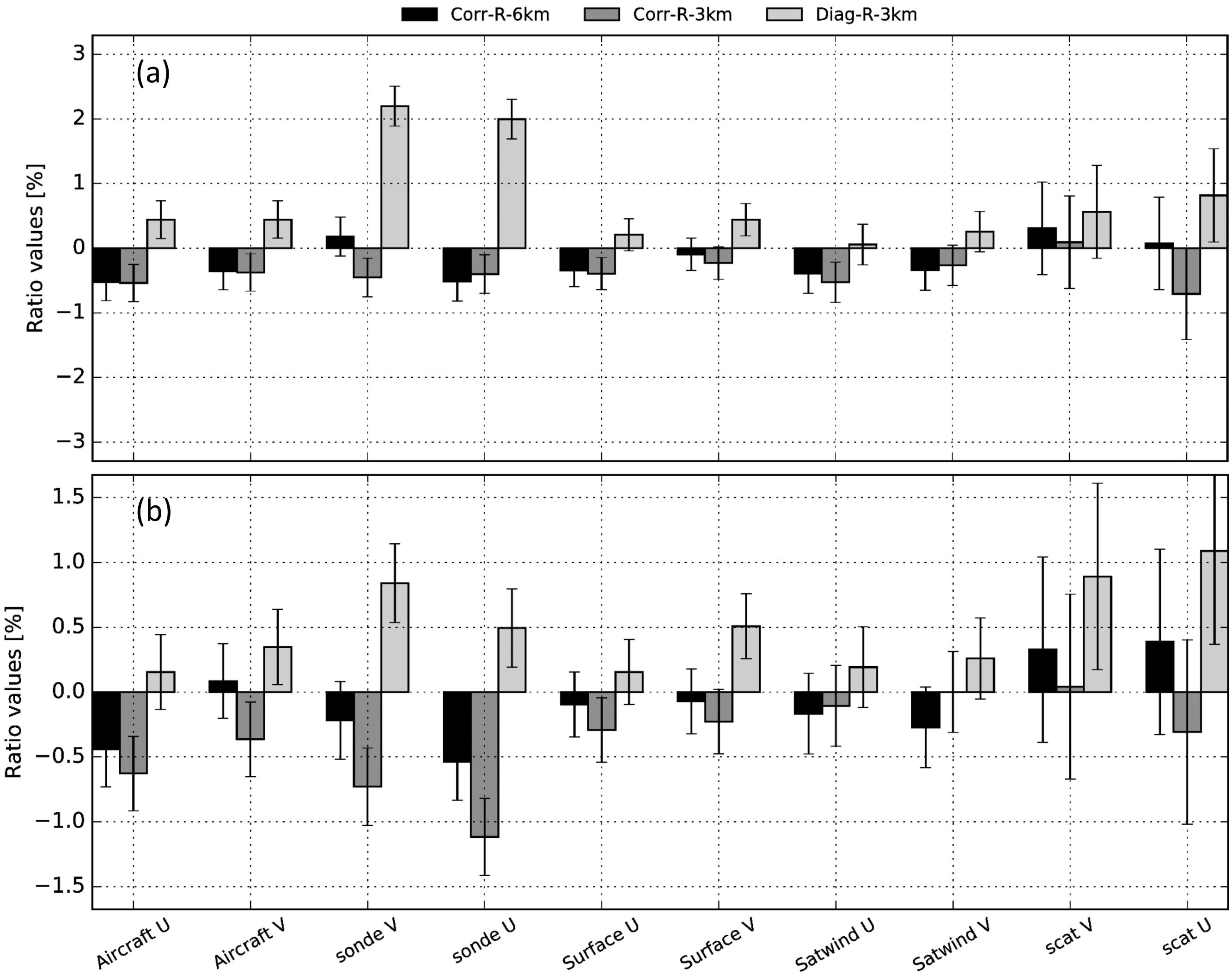

Figure 5. Wind observations (a) $O - A$ and (b) $O - B$ trial average standard deviation ratio between the experiments and the Control expressed as percentage and scaled to show positive impact as negative values (i.e. $\frac{\sigma_{(O-A)_{exp}}}{\sigma_{(O-A)_{ctrl}}} - 1$ and $\frac{\sigma_{(O-B)_{exp}}}{\sigma_{(O-B)_{ctrl}}} - 1$). In black exp =Corr-R-6km, in dark grey exp =Corr-R-3km and in light grey exp =Diag-R-3km . The error bars represent the 95% confidence level.

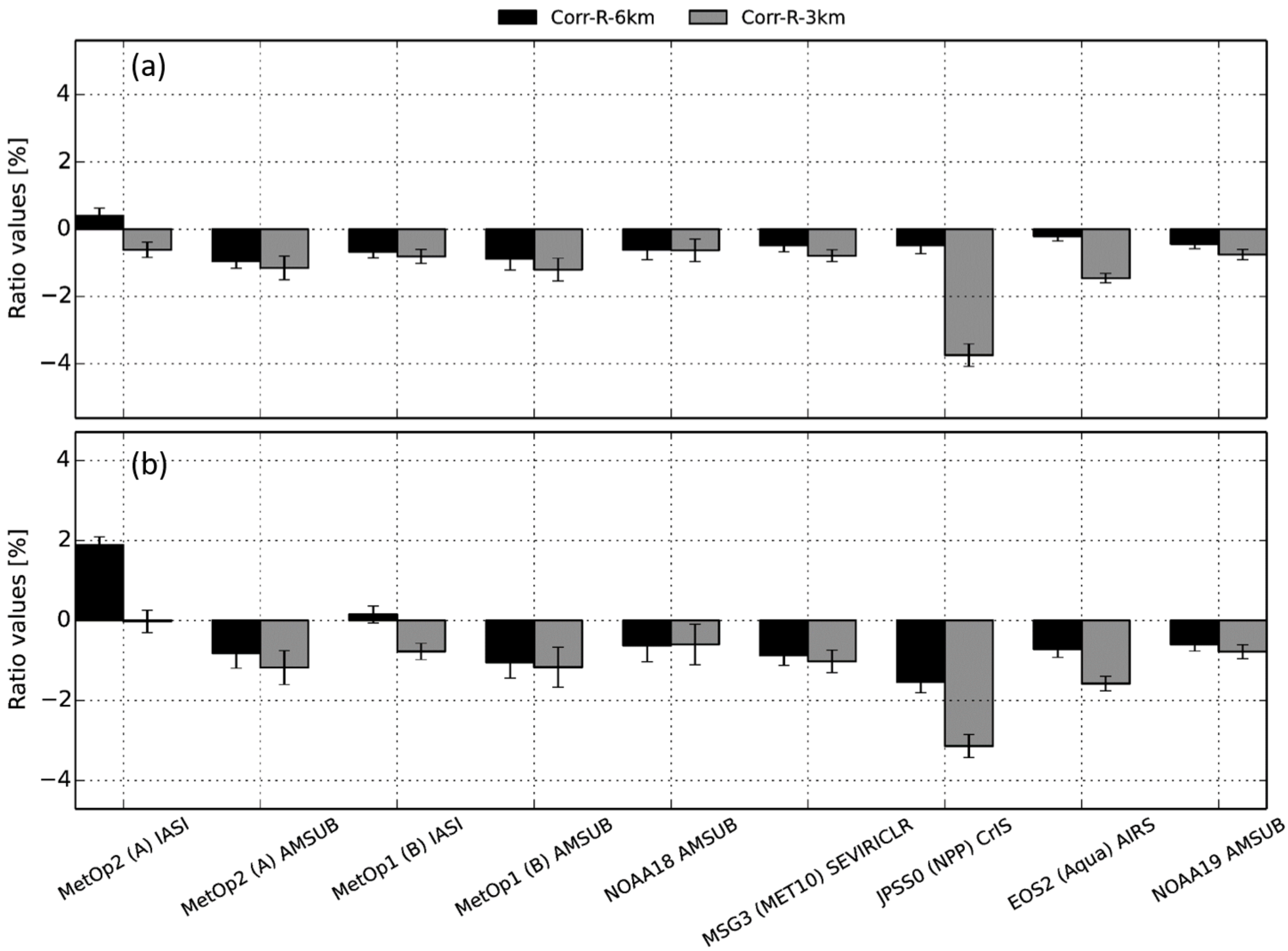

Figure 6. Similar to 5 but for Satellite observations. In black exp =Corr-R-6km and in dark grey exp =Corr-R-3km.

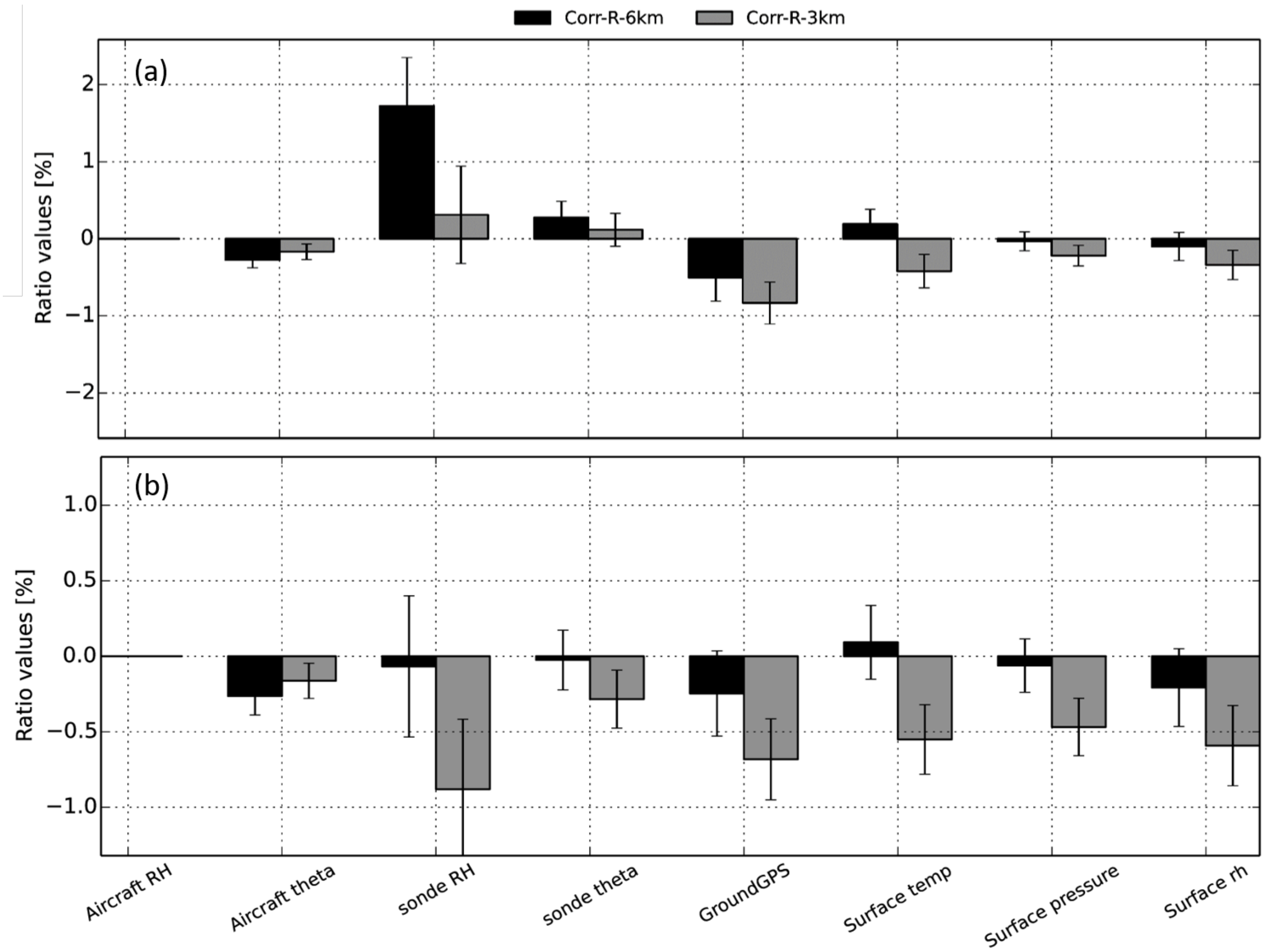

Figure 7. Similar to 6 but for the rest of the observations used.

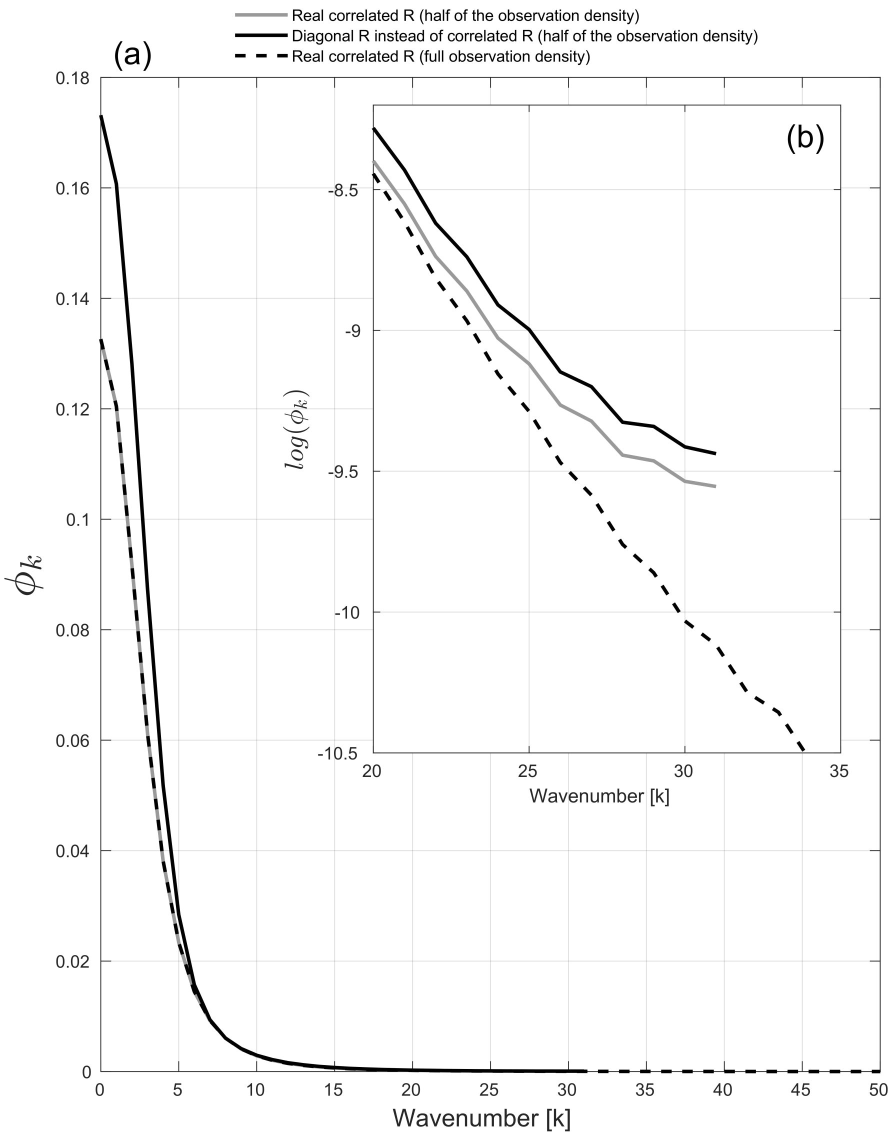

Figure 8. Panel (a) shows the eigenvalues of the analysis error covariance matrix (ϕ_k) in observation space (see appendix for details) against wavenumber. Insert (b) shows ϕ_k in log space for wavenumber ranging from 20 to 35. (Black curve) Φ_k using a diagonal observation error covariance when the true observation error covariance contains some correlation with the state being half observed. (Grey curve) ϕ_k using a correlated observation error covariance with the state being half observed. (Black dashed curve) ϕ_k using a correlated observation error covariance with the state being fully observed.

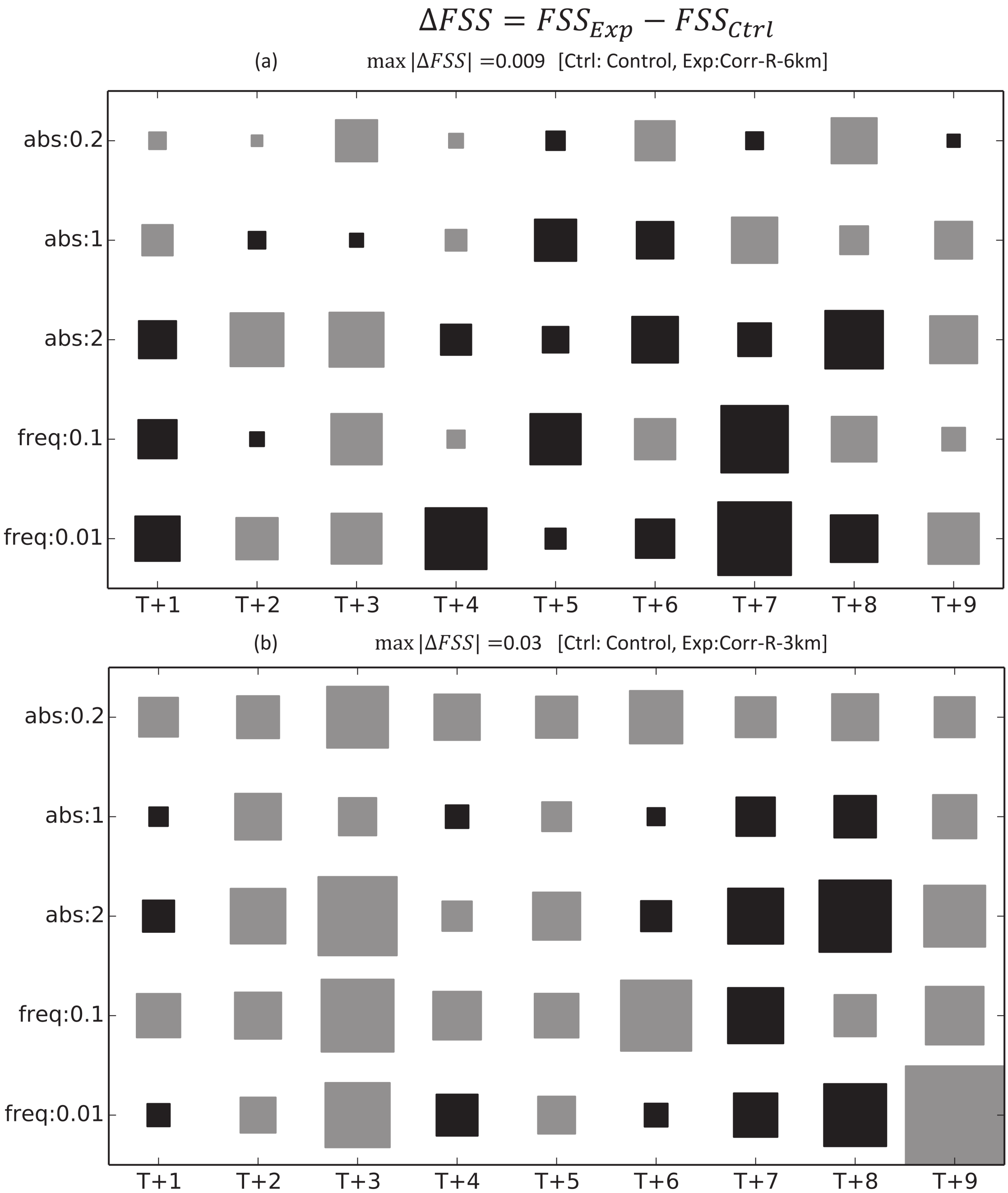

Figure 9. Hinton diagram showing the trial average FSS differences between the Corr-R-6km and the Control experiment (a) and between the Corr-R-3km and the Control experiment (b) for different forecast lead time and hourly rainfall accumulation thresholds with a neighborhood size of 41 grid-boxes. The sign and the amplitude of the change in FSS are shown with the color and size of the square respectively: positive values (positive impact) are shown as gray squares, whereas negative values (negative impact) are shown as blacks square. The rainfall accumulation thresholds on y-axis are $0.2mmh^{-1}$ (abs:0.2), $1.0mmh^{-1}$ (abs:1), $2.0mmh^{-1}$ (abs:2), the 90^{th} percentiles (freq:0.1) and the 99^{th} percentiles (freq:0.01).

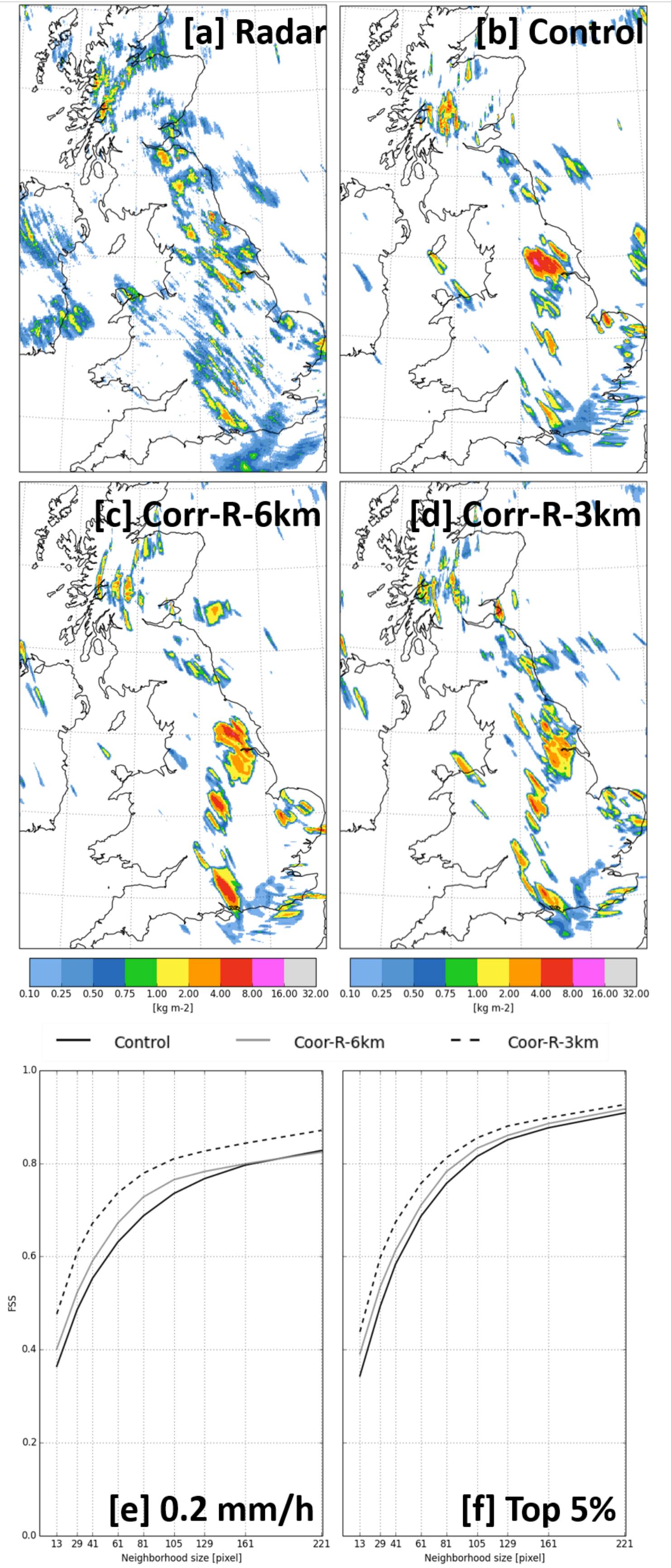

Figure 10. Hourly accumulated precipitation forecasts for 1500 UTC on the 7th of April 2016, for Control [b], Coor-R-6km [c] and Coor-R-3km [d] at T+3. Panel [a] shows the observed radar derived hourly rain accumulation at 1500 UTC. Panels [e] and [f] show the FSS as a function of neighbourhood size for the forecast experiments using thresholds of 0.2mm/h and top 5% (95th percentile) respectively.

