Met Office

Meteorology Research and Development

An evaluation of the utility of ETKF sensitive area predictions for observation targeting

Technical Report No. 513

Keir Bovis

email:nwp_publications@metoffice.gov.uk

An evaluation of the utility of ETKF sensitive area predictions for observation targeting

Keir Bovis

19th March 2008

Abstract

In this paper we evaluate the impact of targeting adaptive observations in a regional model during a one-month long trial period. Targeting guidance is obtained by generating a Sensitive Area Prediction (SAP) from different configurations of the Ensemble Transform Kalman Filter (ETKF). We evaluate the utility of each ETKF configuration in terms of forecast error and overall skill score by running a series of Observation System Experiments (OSE) and assimilating the routine and targeted observations. We include the evaluation of the targeting guidance in the presence of tropical cyclones during the 2005 North Atlantic hurricane season.

1 Introduction

Targeted observations collectively obtained from mobile observation platforms have historically formed part of an adaptive observing network. Such a network is designed to complement the routine observing network used by Numerical Weather Prediction (NWP) models. Targeted observations may comprise additional radiosonde ascents, Aircraft Meteorological Data Relay (AMDAR) aircraft observations, Automatic Shipboard Aerological Program (ASAP) ships balloon ascents and additional observations from drifting buoys. Each observation is obtained from, or targeted in, a previously defined *targeting region*. It is hypothesised that the assimilation of extra observations in these regions will lead to a reduction in forecast error of NWP models within a pre-defined *verification region*. The process of selecting targeting regions is termed Sensitive Area Prediction (SAP) and historically two competing approaches have been proposed, singular vectors and ensemble-based methods.

The application of singular vectors for targeting observations is discussed in detail in Palmer *et al.* [16]. Singular vectors have been used to provide a method for estimating the growth of small perturbations in NWP models over time. By associating them with the notion of predictability, studies have investigated the application in the estimation of initial errors prior to the production of a forecast. The utility of singular vectors in generating a SAP has been investigated in various observation system tests, the North Pacific Experiment (NORPEX) [10], the Fronts and Atlantic Storm-Track EXperiment (FASTEX) [1] and the Atlantic THORPEX Regional Campaign (ATReC) [11, 18].

An alternative to the singular vector approach is an ensemble-based method for selecting targeting regions, proposed by Bishop *et al.* [3]. The ensemble transform technique seeks a deployment of adaptive observations that minimise the expected prediction error associated with each possible deployment. An important assumption is that the ensemble used in initialisation gives an accurate estimate of the distribution and range of possible prediction errors associated with each targeted observation deployment. This approach has been used in conjunction with the singular vector method during the NORPEX and FASTEX campaigns. Bishop *et al.* extend their technique further by proposing the Ensemble Transform Kalman Filter (ETKF) [2]. The ETKF approach attempts to estimate the reduction in forecast error associated with the deployment of targeted observations using a set of ensemble member perturbations weighted by a transform matrix. By computing the transform matrix in a lower dimensional space, the deployment of observations that maximise the reduction in forecast error can be identified. The ETKF technique was successfully applied during the 2003 Atlantic THORPEX Regional Campaign (ATReC).

Comparisons between the two different approaches for generating SAPs have been attempted [13, 5, 12], though in reality when used in field campaigns adaptive networks are designed on the basis of predictions from all competing methods. This paper builds on the previous work [15] undertaken at the UK Met Office (UKMO) using the ETKF during the 2003 ATReC. It aims to investigate the different configurations of the ETKF. Similar analysis of different ETKF configurations has been undertaken by Petersen *et al.* [17] although this work differs in that rather than statistically evaluating the quality of the SAP we quantify the impact in terms of forecast error using a regional NWP model. In this way, the utility of different ETKF configurations and resulting SAPs are presented during a

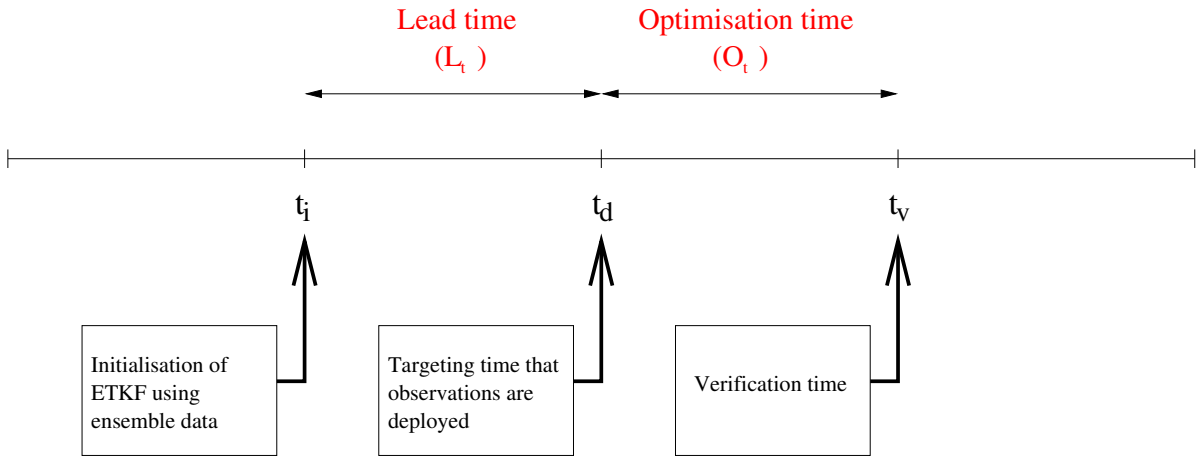


Figure 1: The targeting time-line used in the ETKF

one month-long trial period. This paper is organised as follows, Section 2 presents a brief overview of observation targeting and the formulation of the ETKF, Section 3 describes the experimental set-up. Sections 4 and 5 describe experiments undertaken to evaluate the effect of varying ensemble size and lead time together with a summary of results obtained. Targeting case studies in the presence of tropical cyclones are presented in Section 6 and a discussion of the how the ensemble affects the ETKF is given in Section 7. Conclusions from all experiments are presented in Section 8.

2 Observation targeting and the ETKF

In this section we introduce the observation targeting problem and give a brief overview of the discrete Kalman Filter. Using this, we describe how the Kalman Filter is used in the ETKF and can be applied to observation targeting. Throughout this section we follow the notation used in Ide *et al.* [8].

2.1 The observation targeting problem

Observation targeting is initiated by the design of an optimal adaptive observing network at some initial time (t_i), that complements a routine network for deployment at targeting time (t_d), subsequently assimilated into an NWP model. With this resulting analysis, forecasts are produced with the aim of reducing the forecast error at verification time (t_v) within a predefined verification region. The difference between t_i and t_d is referred to as the *lead time* (L_t), and that between t_d and t_v the *optimisation time* (O_t). The targeting time-line is shown in Figure 1. Accurate identification of the targeting area is critical in specifying the optimal adaptive observing network.

2.2 The discrete Kalman Filter

The Kalman Filter [9] has been successfully applied to various problems in many aspects of control theory. In its simplest form, it is a recursive data processing algorithm, built on Bayesian foundations, attempting to infer the variables of state from all available data. The Kalman filter estimates a process using a form of feedback control used in classic *predictor-corrector* algorithms:

Predict Estimate the process state x at a time step t_i using time update equations and obtain an *a priori* estimate of state $\mathrm{x}^{\mathrm{f}}(t_i)$ and covariance $\mathbf{P}^{\mathrm{f}}(t_i)$ for use in the next time step. This is achieved by projecting forward the current state analysis x^{a} and error covariances analysis $\mathbf{P}^{\mathrm{a}}$ in the presence of a process noise estimate $\mathbf{Q}$ by

$$\begin{aligned} \mathrm{x}^{\mathrm{f}}(t_i) &= \mathrm{x}^{\mathrm{a}}(t_{i-1}) \\ \mathrm{P}^{\mathrm{f}}(t_i) &= \mathbf{P}^{\mathrm{a}}(t_{i-1}) + \mathbf{Q}(t_{i-1}) \end{aligned} \tag{1}$$

Correct Obtain feedback in the form of measurement $\mathrm{y}_i^{\mathrm{o}}$, contaminated with process noise $\mathbf{R}_i$ at time step t_i. This is incorporated using a weighted innovation $\mathbf{K}_i$, into the *a priori* estimate to obtain an improved *a*

posteriori estimate of state $\mathrm{x}^{\mathrm{a}}(t_i)$ and covariance $\mathrm{P}^{\mathrm{a}}(t_i)$. H relates the state to the measurement $\mathrm{y}_i^{\mathrm{o}}$ at time step t_i.

$$\begin{array}{rcc} \mathbf{K}_i & = & \mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T\left[\mathrm{H}_i\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T+\mathbf{R}_i\right]^{-1} \\ \mathrm{x}^{\mathrm{a}}(t_i) & = & \mathrm{x}^{\mathrm{f}}(t_i)+\mathbf{K}_i\left(\mathrm{y}_i^{\mathrm{o}}-\mathrm{H}_i\mathrm{x}^{\mathrm{f}}(t_i)\right) \\ \mathbf{P}^{\mathrm{a}}(t_i) & = & (\mathbf{I}-\mathbf{K}_i\mathrm{H}_i)\,\mathbf{P}^{\mathrm{f}}(t_i) \end{array} \tag{2}$$

We make the following assumptions on the tuning of individual parameters. Using equation 2 we can derive a more complete definition of the *a posteriori* estimate of the covariance $\mathbf{P}^{\mathrm{a}}(t_i)$ by substituting in $\mathbf{K}_i$

$$\begin{array}{rcc} \mathbf{P}^{\mathrm{a}}(t_i) & = & (\mathbf{I}-\mathbf{K}_i\mathrm{H}_i)\,\mathbf{P}^{\mathrm{f}}(t_i) \\ & = & \left(\mathbf{I}-\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T\left[\mathrm{H}_i\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T+\mathbf{R}_i\right]^{-1}\mathrm{H}_i\right)\mathbf{P}^{\mathrm{f}}(t_i) \\ & = & \mathbf{P}^{\mathrm{f}}(t_i)-\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T\left[\mathrm{H}_i\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T+\mathbf{R}_i\right]^{-1}\mathrm{H}_i\mathbf{P}^{\mathrm{f}}(t_i) \end{array} \tag{3}$$

and apply this to NWP by assuming:

$\mathbf{x}^{\mathbf{f}}\left(t \mid \mathrm{H}_i^q\right)$ is a forecast estimate of the state of the atmosphere at time t conditional upon the qth sequence of observations available up to some time t_i.

$\mathbf{P}^{\mathbf{f}}\left(t \mid \mathrm{H}_i^q\right)$ is the associated forecast error covariance estimate at time t conditional upon the qth sequence of observations available up to some time t_i.

We define H_i^q as an operator that assimilates observations at initialisation time t_i associated with the qth feasible sequence of observations. Substituting these terms into Equation 3, we use the Kalman Filter to obtain an estimate of the analysis covariance matrix $\mathbf{P}^{\mathbf{a}}\left(t_{i+m} \mid \mathrm{H}_{i+m}^q\right)$ at any time t_{i+m} using the forecast error covariance matrix $\mathbf{P}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_{i+m-1}^q\right)$ as

$$\begin{array}{rcl} \mathbf{P}^{\mathrm{a}}(t_i) & = & \mathbf{P}^{\mathrm{f}}(t_i)-\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T\left[\mathrm{H}_i\mathbf{P}^{\mathrm{f}}(t_i)\,\mathrm{H}_i^T+\mathbf{R}_i\right]^{-1}\mathrm{H}_i\mathbf{P}^{\mathrm{f}}(t_i) \\ \mathbf{P}^{\mathbf{a}}\left(t_{i+m} \mid \mathrm{H}_{i+m}^q\right) & = & \mathbf{P}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_{i+m-1}^q\right) \\ & & -\mathbf{P}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_{i+m-1}^q\right)\mathrm{H}_{i+m}^{qT}\left[\mathrm{H}_{i+m}^q\mathbf{P}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_{i+m-1}^q\right)\mathrm{H}_{i+m}^{qT}+\mathbf{R}_{i+m}^q\right]^{-1} \\ & & \mathrm{H}_{i+m}^q\mathbf{P}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_{i+m-1}^q\right) \end{array} \tag{4}$$

2.3 Applications to observation targeting

The deployment of additional targeted observations is made with the aim of identifying an optimal deployment that will minimise the error covariance matrix $\mathbf{P}^{\mathbf{a}}\left(t_v \mid \mathrm{H}_i^q\right)$ at verification time t_v in a verification region. The qth considered deployment from Q total additional deployments that minimises $\mathbf{P}^{\mathbf{a}}\left(t_v \mid \mathrm{H}_{i+m}^q\right)$ can be identified by evaluating Equation 4 repeatedly. There is though, a huge cost in running a Kalman Filter for an operational NWP system with a routine observing network let alone utilising an additional one. In their derivation, Bishop *et al.* identify significant computational expense in computing the inversion of the residual covariance matrix $\left[\mathrm{H}_i^q\mathbf{P}^{\mathbf{f}}\left(t_i \mid \mathrm{H}_{i-1}^q\right)\mathrm{H}_i^{qT}+\mathbf{R}_i^q\right]^{-1}$ for two reasons

1. The matrix is very large of size $p \times p$ where $p \sim O\left(10^5\right)$ observations
2. The matrix is ill-conditioned, that is, application of an eigenvalue transformation will result in the largest eigenvalue being many orders of magnitude larger than the smallest.

2.4 Using an ETKF to approximate the Kalman Filter

There are three novel elements to the solution proposed by Bishop *et al.* to address these problems delivering a solution that is tractable.

1. An ensemble of forecasts. By making use of an ensemble $n = \{1 \ldots N\}$ of forecasts valid from initialisation time t_i, to deployment time t_d and out to verification time t_v, the outer product of differences (or perturbations $\mathbf{x}^{\mathbf{f}}\left(t \mid \mathrm{H}_{i+m}\right)$) of each nth forecast from the ensemble mean can be used to approximate the error covariance matrix $\mathbf{P}^{\mathbf{f}}\left(t \mid \mathrm{H}_{i+m}\right)$ at times $\{t_i, t_d, t_v\}$. This approach avoids the need to propagate covariance error matrices forward in time necessary for running the Kalman Filter.

2. Orthonormal transformations. Transformation of the perturbations and attaching variances to each of the direction vectors allows the ensembles to be used to describe error covariances within a vector subspace of perturbations. This subspace is significantly smaller making the computation possible. Each transformation can be computed and applied to the perturbations to give an estimate of the forecast error covariance obtained by different deployments of observations.

3. Serial processing theory. Further reductions in computation expense are obtained by appropriate use of serial processing theory. This is achieved by first estimating the effect of the routine observation network on the estimated error covariance before considering the effect of each feasible deployment of adaptive networks. *This assumes that adaptive observation errors are uncorrelated with the routine observation errors.*

In the next section we discuss how the error covariances for the routine and adaptive observing networks can be estimated and used to estimate a signal covariance. Minimisation of this will result in the identification of the optimal adaptive network.

2.5 Error covariances associated with the routine observing network

Weighting ensemble perturbations with an orthonormal transform allows the specification of error covariances at any time. For example the first-guess analysis error covariance matrix $\mathbf{P^a}\left(t_{i+1} \mid \mathrm{H}_i\right)$ can be written as the outer product of perturbations weighted by $\mathbf{T}_0$

$$\mathbf{P^a}\left(t_{i+1} \mid \mathrm{H}_i\right) = \mathbf{x^f}\left(t_i \mid \mathrm{H}_i\right) \mathbf{T}_0 \mathbf{T}_0^T \mathbf{x^f}\left(t_i \mid \mathrm{H}_i\right)^T \tag{5}$$

where $\mathbf{T}_0$ is equal to the identity matrix $\mathbf{I}$ at time t_{i+1} but at any future time t_{i+m}, the $N \times N$ transformation matrix $\mathbf{T}_{i+m-1}$ is generally not equal to the identity matrix. Accurate specification of the transformation matrix is vital in estimating the forecast error covariance matrix. Bishop *et al.* show that if an eigenvalue decomposition is undertaken yielding eigenvectors $\mathbf{E}^q$ and eigenvalues λ^q for the qth deployment, the residual covariance matrix in Equation 4 can be written

$$\left[\mathrm{H}_{i+m}^q \mathbf{P^f}\left(t_{i+m} \mid \mathrm{H}_{i+m-1}^q\right) \mathrm{H}_{i+m}^{qT} + \mathrm{I}^{p\times p}\right]^{-1} = \mathbf{E}^q\left(\lambda^q + \mathbf{I}^{p\times p}\right)\mathbf{E}^{qT} \tag{6}$$

This is the basis upon which the transformation matrix $\mathbf{T}^r$, capturing the impact of the routine observing network, is calculated. Actual implementation differs in that eigenvectors $\mathbf{E}^q$ are scaled by $\sqrt{\frac{1}{\lambda^q}}$ such that

$$\mathbf{T}^r = \mathbf{E}^r\left(\lambda^r + \mathbf{I}^{F\times F}\right)^{-\frac{1}{2}} \tag{7}$$

To calculate $\mathbf{T}^r$, an eigenvalue decomposition is undertaken using the ensemble perturbations interpolated to the locations of the routine observing network and a linearised observation operator $\mathbf{H}^r$ such that

$$\mathbf{T}^{rT}\mathbf{x^f}\left(t_{i+m} \mid \mathrm{H}_i^r\right)^T \mathbf{D}_A^{-1}\left(\mathbf{H}^r\right)\mathbf{x^f}\left(t_{i+m} \mid \mathrm{H}_i^r\right)\mathbf{T}^r = \mathbf{I} \tag{8}$$

$\mathbf{D}_A^{-1}$ lists the inverse of the best available estimates of analyses error variance, a measure of the process error, at time t_i. Current ETKF implementation uses values that relate NCEP analyses' fit to radiosonde observation taken from Majumdar *et al. [14]*. Further analysis is required to determine if these values approximate Met Office analyses' fit. The analysis error covariance matrix $\mathbf{P^a}\left(t_{i+m} \mid \mathbf{H}_i^r\right)$ associated with the routine observation network can now be written using Equations 5 and 7 as

$$\mathbf{P^a}\left(t_{i+m} \mid \mathbf{H}_i^r\right) = \mathbf{x^f}\left(t \mid \mathrm{H}_i^r\right)\mathbf{T}^r\mathbf{T}^{rT}\mathbf{x^f}\left(t \mid \mathrm{H}_i^r\right)^T \tag{9}$$

2.6 Error covariances associated with the adaptive observational network

Equation 4 can be re-written in terms of a signal covariance matrix $\mathbf{S}\left(t_{i+m} \mid \mathbf{H}_i^q\right)$ defined as the estimated outer product of signal realisations that would be produced by the qth deployment of supplementary *and* routine observation network if calculated using the error covariances of the ETKF

$$\mathbf{P^a}\left(t_{i+m} \mid \mathbf{H}_i^q\right) = \mathbf{P^f}\left(t_{i+m} \mid \mathrm{H}_i^q\right) - \mathbf{S}\left(t_{i+m} \mid \mathbf{H}_i^q\right) \tag{10}$$

This assumes that the targeted and routine analysis error covariances are both specified and uncorrelated with each other. The signal covariance in Equation 10 is equal to the reduction due to the qth adaptive deployment. The

signal covariance matrix can be expressed in a similar way to the analysis error covariance matrix in Equation 9 by utilising the ensemble of perturbations and serial processing theory

$$\mathbf{S}\left(t_{i+m} \mid \mathbf{H}_i^q\right) = \mathbf{x}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_i^r\right)\mathbf{V}^q\mathbf{V}^{qT}\mathbf{x}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_i^r\right)^T \tag{11}$$

Bishop *et al.* give a derivation for the $N \times N$ transformation matrix $\mathbf{V}^q$ that captures the impact of the deployment of the qth adaptive network in the presence of the routine observation network (captured in $\mathbf{T}^r$) as

$$\mathbf{V}^q = \mathbf{T}^r\mathbf{E}^q\left[\lambda^q\left(\lambda^q + \mathbf{I}^{F\times F}\right)^{-1}\right]^{\frac{1}{2}} \tag{12}$$

where $\mathbf{E}^q$ and λ^q are the eigenvectors and eigenvalues of the matrix product associated the matrix product specified in a similar way in Equation 8 such that

$$\mathbf{T}^{rT}\mathbf{x}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_i^r\right)\widehat{\mathbf{R}}^{q-1}\left(\mathbf{H}^q\right)\mathbf{x}^{\mathbf{f}}\left(t_{i+m} \mid \mathrm{H}_i^r\right)\mathbf{T}^r = \mathbf{I} \tag{13}$$

Ensemble perturbations are interpolated to the location of a test radiosonde simulating the deployment of an adaptive observation using a linearised observation operator $\mathbf{H}^q$. $\widehat{\mathbf{R}}^q$ *holds the pre-defined radiosonde observational error covariance matrix taken as specified in Majumdar* et al. [14].

2.7 Summary maps

A summary map is a sensitive area prediction generated by the ETKF synthesising the signal covariance matrix for a large number of different hypothetical adaptive networks in the presence of the routine observation network. We seek the impact of the routine and adaptive observing network on the analysis error covariance at verification time t_v, localised using our localisation operator L_v within the verification region. This can be specified using the ensemble of perturbations and Equation 11 as

$$L_v\left[\mathbf{S}\left(t_v \mid \mathbf{H}^q\right)\right] = L_v\left[\mathbf{x}^{\mathbf{f}}\left(t_v \mid \mathrm{H}_i^r\right)\right]\mathbf{V}^q\mathbf{V}^{qT}L_v\left[\mathbf{x}^{\mathbf{f}}\left(t_v \mid \mathrm{H}_i^r\right)^T\right] \tag{14}$$

To obtain an overview of all the information associated with the eigenvectors of $\mathbf{S}\left(t \mid \mathbf{H}^q\right)$ Majumdar *et al.* [14] propose calculating the signal variance at each point expressed in terms of a total perturbation energy as

$$\sigma = 0.5\left(u'^2 + v'^2 + \frac{c_p}{T_r}T'^2\right) \tag{15}$$

averaged over 850-, 500- and 200 hPa levels. The variables *(u', v', T')* represent the transformed wind component and temperature perturbations given in the $\mathbf{x}^{\mathbf{f}}\left(t_v \mid \mathrm{H}_v^r\right)\mathbf{V}^q$ matrix, c_p is the specific heat and T_r= 300 K is the reference temperature. An example summary map is shown in Figure 2 for targeting on 23 September 2005 12 UTC as Hurricane Rita is about to make landfall on the southern coastline of the United States. The ETKF generated SAP identifies that deployment of adaptive observations in an area centred 100^oW by 30^oN that is estimated to make an impact on a forecast verifying on 24 September 2005 12 UTC in the verification area defined by the blue box.

3 Experimental setup

The ETKF may be run with a number of different parameter settings and these are summarised in Table 1. By varying one or more of the parameters for each ETKF configuration, a different SAP is generated at a given deployment time t_d. One approach to assess the impact of these different configurations is to generate associated SAPs, then using these as guidance, deploy and assimilate the additional adaptive observations. This approach is known as an Observation System Experiment (OSE). By repeating this for a month long period and calculating the Root Mean Square (RMS) error for a basket of forecast fields at different forecast ranges, it is possible to obtain a quantitative measure of the performance at each ETKF configuration. In this study, targeting guidance used in all OSEs is derived from the ETKF initialised from the ECMWF ensemble [4]. The range of values considered for each ETKF parameter is shown in the last column of Table 1.

A Met Office Local Area Model (LAM) [7] is used to evaluate different configurations of the ETKF. The LAM domain definition is shown in Figure 3(a) together with the bottom left and top right points of the verification area ($V_a = \{35^oN/98^oW/45^oN/85^oW\}$) used to evaluate each configuration of the ETKF. The verification region

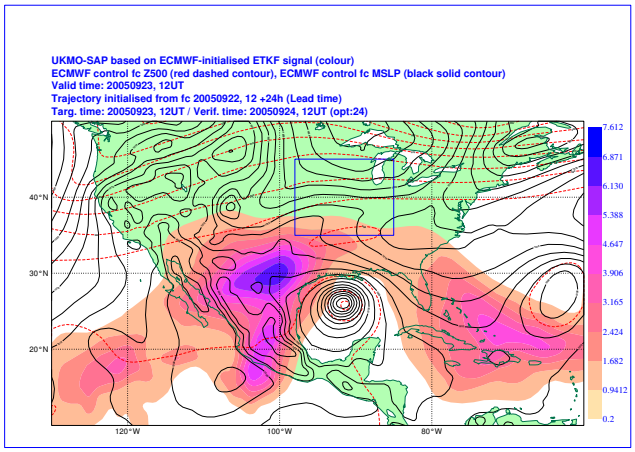


Figure 2: An example of a summary map showing the signal variance and the control forecast 500 hPa geopotential height and mean sea level pressure. The blue square is the verification area.

Parameter	Description	Range of values
E_n	The number of members comprising the ensemble used to initialise the ETKF.	$E_n = \{10, 20, 30, 40, 50\}$
L_t	The lead time (in hours) or difference between ensemble initialisation t_i and observation deployment time t_d.	$L_t = \{24, 48\}$
O_t	The optimisation time (in hours), i.e. the difference between observation deployment time t_d and forecast verification time t_v.	$O_t = \{24, 48\}$
V_a	The vector of grid points comprising the verification area.	$V_a = \{35^oN/98^oW/45^oN/85^oW\}$

Table 1: ETKF parameters

is chosen such that radiosondes are available for deployment upstream in any direction depending on the location of the sensitive areas. The LAM resolution is based on a 17 km rotated grid with 38 model levels and uses a 3D-variational assimilation scheme. It has been modified to generate forecasts in the ranges 24-hour, 36-hour and 48-hour at 00 UTC and 12 UTC of each trial day. At each 00/12 UTC cycle, each OSE is initialised from an identical observation and assimilation background valid at that time. By doing this, the impact of each deployment is apparent and not masked by background departures. Boundary conditions for the regional model are obtained from the Met Office's Global model. The OSE trial period spans 29 August 2005 12 UTC through to 1 October 2005 12 UTC and includes the four tropical cyclones identified in Table 2, though not all of them pass directly through V_a. Qualitative results for each these high impact events will be given later in this paper.

To construct the routine radiosonde network, stations are subjectively selected from the the existing North American network and then thinned to approximate resolution of one station per 10^o latitude/longitude box. Additional adaptive observations may then be added by selecting from those remaining not labelled as routine observing stations. No routine or adaptive observations are deployed within the verification area. Throughout the trial period all aircraft observations have been removed to improve the radiosonde observation targeting signal. The impact of aircraft observations has been shown to lead to a positive effect on the reduction of forecast error in regional models [6]. Re-running these experiments with the inclusion of these observations may lead to a reduction in any subsequently reported positive impact. Figure 3(a) shows the locations of surface stations that may be

Tropical cyclone	Landfall	Date
Hurricane Katrina	$30^oN/89^oW$	29 August 2005 12 UTC
Hurricane Ophelia	$34^oN/77^oW$	14 September 2005 00 UTC
Hurricane Rita	$30^oN/93^oW$	24 September 2005 12 UTC
Hurricane Otis	$20^oN/110^oW$	30 September 2005 12 UTC

Table 2: Tropical cyclones occurring during the trial period

OSE	(E_n)					(L_t)		(O_t)	
	10	20	30	40	50	24	48	24	48
10^{24}_{24}	★					★		★	
20^{24}_{24}		★				★		★	
30^{24}_{24}			★			★		★	
40^{24}_{24}				★		★		★	
50^{24}_{24}					★	★		★	

Table 3: ETKF configuration of OSEs evaluating the effect of ensemble size. For each OSE, the resolution of the ensemble used to initialise the ETKF is $2.5^o \times 2.5^o$.

considered for observation targeting.

For every 12-hour cycle during the trial period, each configuration of the ETKF is run. The ETKF signal variance is interpolated to the locations of the targetable radiosonde stations. Radiosonde observations from the top 10 ranked surface stations maximising the ETKF signal variance are automatically selected for assimilation into the NWP model. Figure 3(b) shows an example SAP for targeting date t_d= 29 August 2005 12 UTC together with the ETKF signal variance and forecast Mean Sea Level Pressure (MSLP) and geopotential height at 500 hPa (Z500). The 10 stations selected for targeting are shown in black together with the routine observing network in red.

Each OSE in this study is identified using the standard naming convention $E_n \begin{smallmatrix} O_t \\ L_t \end{smallmatrix}$, where E_n is defined as the size of the ensemble used to initialise the ETKF, L_t is the lead time and O_t is the optimisation time. For example, OSE 50^{48}_{24} refers to an OSE using ETKF targeting guidance produced from a 50 member ensemble with a lead time of 24 hours and an optimisation time of 48 hours.

The results presented in this study do not include the effect of varying V_a and it is assumed to be constant in all OSEs. The following sections describe experiments and result summaries for the effect of varying ensemble size and lead time.

4 The effect of varying ensemble size

This section presents the results obtained from running five different OSEs, each assimilating adaptive observations deployed using a SAP created from the ETKF initialised with different sized ensembles. The configurations of the ETKF used in each OSE is shown in Table 3. SAPs are generated for every 12 hour cycle throughout the trial period on the basis of which targeted observation are deployed and assimilated as previously described.

4.1 Evaluation of OSE SAPs

To quantitatively evaluate the similarity of a SAP pair, two metrics proposed by Majumdar *et al.* [13] are examined, the Modified Equitable Threat Score (METS) and a ranking of neighbouring regions.

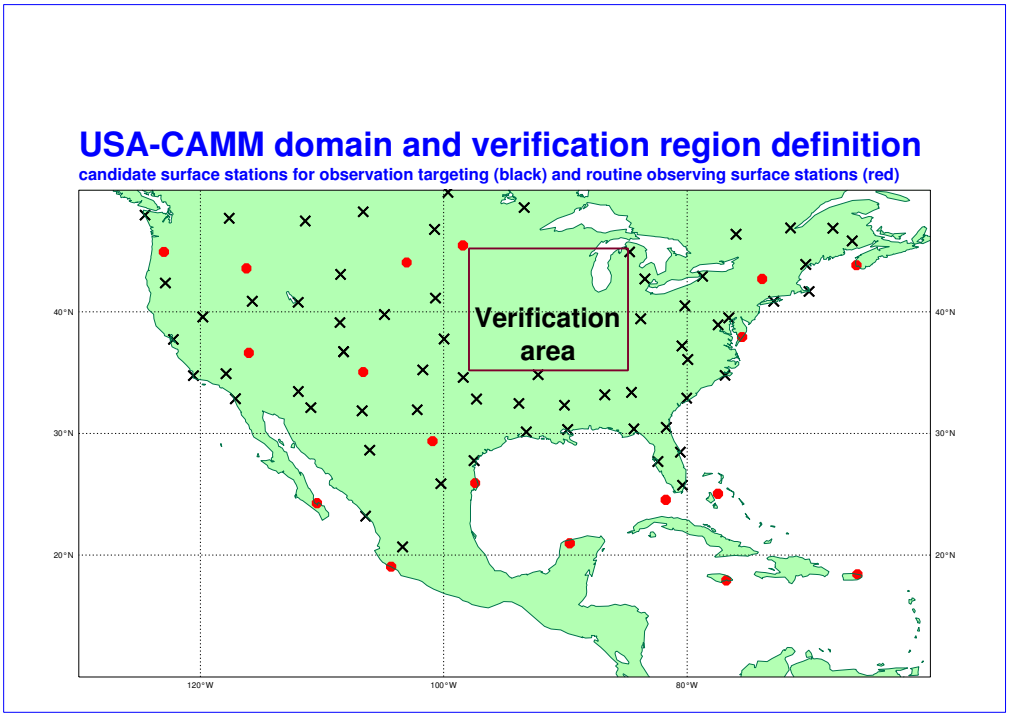
Modified Equitable Threat Score (METS)

This is a test of similarity between SAPs generated from different configurations of the ETKF. METS is calculated using Equation 16 by considering the highest X ranked grid points within a SAP pair. The number of common points contained within each pair is defined by C and $E[C]$ gives the expected number of grid points that may occur by chance.

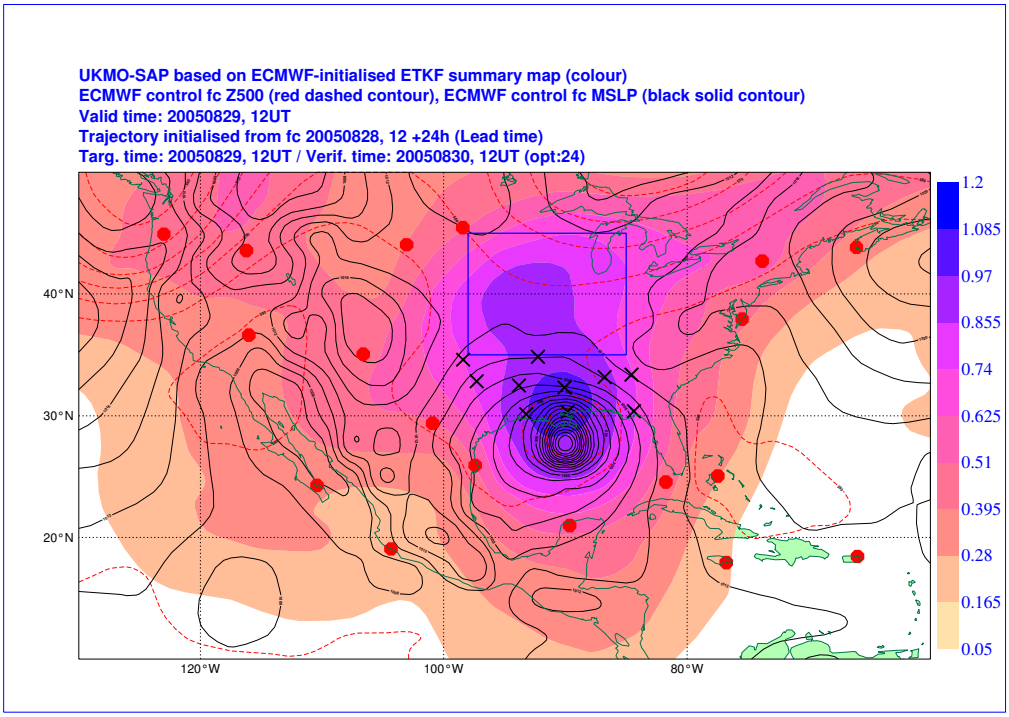
$$METS = \frac{C - E[C]}{2X - C - E[C]} \tag{16}$$

$E[C]$ is defined using a similar approach to Majumdar *et al.* [13] by considering on a leave-one-out basis the mean value of C over each combination of independent paired configurations of similar comparison. For each OSE defined in Table 3 , the expected value of $E[C]$ is computed.

For each OSE, Table 4(a) shows the percentage of the 70 cases where the calculated METS score was greater than zero for $X = 50$. From these results, the two largest ensembles used in OSEs 50^{24}_{24} and 40^{24}_{24} have 100% of cases with a METS score greater than zero. This percentage drops considerably when comparing the largest and smallest ensemble sizes (OSEs 10^{24}_{24} and 20^{24}_{24}) with only 47% of cases having a METS score greater than zero. These results indicate that larger ensemble pairs offer similar targeting guidance compared to smaller sensible pairs using this



(a)



(b)

Figure 3: (a) LAM domain and verification area (V_a) definition (candidate surface stations for observation targeting and routine surface network shown as black crosses and red circles respectively); (b) An example SAP (t_d= 29 August 2005 12 UTC) showing the ETKF signal variance and forecast Mean Sea Level Pressure (MSLP) and geopotential height at 500 hPa (Z500). Stations selected for targeting are shown as black crosses, routine observing network shown as red dots.

	40^{24}_{24}	30^{24}_{24}	20^{24}_{24}	10^{24}_{24}
50^{24}_{24}	100%	88%	74%	47%
40^{24}_{24}		94%	74%	52%
30^{24}_{24}			74%	57%
20^{24}_{24}				67%

(a)

	40^{24}_{24}	30^{24}_{24}	20^{24}_{24}	10^{24}_{24}
50^{24}_{24}	94%	91%	77%	57%
40^{24}_{24}		97%	87%	64%
30^{24}_{24}			90%	67%
20^{24}_{24}				77%

(b)

Table 4: Evaluation of the impact of ensemble sizes on a SAP. (a) Entries indicate OSE paired configurations considered for calculating $E[C]$. Actual values are the percentage of the cases in which the METS score is greater than zero for X=50. (b) Percentage of cases in which ranked correlation between any two sets of two targets are deemed similar at the 99% significance level, for $Y = 8$ boxes of size 9×9 at $Y - 2$ degrees of freedom.

measure. Percentage of cases with a METS score greater than zero for other OSE pairings vary according to the similarity in size of the ensemble used.

Ranking of neighbouring region

The second metric considered makes use of a box average of the ETKF signal variance used by each OSE. For each SAP, Y grid boxes of size $G \times G$ are defined over the whole area. The average ETKF signal variance in each box is calculated and a degree of similarity between any paired combination of OSEs is computed using the Spearman rank correlation coefficient, ρ. In this study, a box size of $G = 9$ corresponding to a lower correlation between adjacent regions at the expense of a smaller number of boxes is used.

Table 4(b) shows the percentage of ranked correlation of neighbouring regions that are deemed similar at the 99% significance level for each OSE defined in Table 3. In this test, the null hypothesis H_0= there is no correlation between the ranking of corresponding SAPs. Of the 70 cases, 94% of the paired OSEs $50^{24}_{24}, 40^{24}_{24}$ and 97% of the paired OSEs $40^{24}_{24}, 30^{24}_{24}$ correlate at the 99% significance level. Paired comparisons on disparate sized ensembles, for example $50^{24}_{24}, 10^{24}_{24}$ give the lowest percentage of cases correlating significantly.

4.2 Analysis of resulting forecast error

This section presents verification results from each OSE quantified in terms of mean forecast minus observation RMS error for the optimisation time. RMS error is calculated by verifying forecasts against sonde and surface observations within the verification area. Forecast fields used in the verification of each OSE are listed in Table 5.

Verification scores for each forecast field for the optimisation time ($O_t = 24$) are shown in Figures 4(a)-(f). These show the forecast error for each OSE together with an OSE labelled BASE containing a routine network deployment only. Verification of geopotential height is shown in Figure 4(a) and it can be seen that at the lower and upper levels there is little difference in the results from the different ensemble sizes. At mid-levels 500 hPa and 250 hPa, the larger ensemble OSE 50^{24}_{24} is marginally out-performing the smaller ensemble OSE 10^{24}_{24} with other configurations spread out between. The signal for temperature shown in Figure 4(b) is clearer with OSE 10^{24}_{24} giving a larger RMS error compared with OSE 50^{24}_{24} at all levels. At the mid-level 500 hPa the larger ensemble size, OSE 50^{24}_{24}, has a reduced RMS error compared with the other OSEs. Results for vector winds shown in Figure 4(c) indicate little clear difference between each of the OSEs with the exception of the 250 hPa level where the greatest variance in scores exists. The performance observed for OSE 50^{24}_{24} is inferior than that for 10^{24}_{24}in terms of RMS error. Verification obtained for relative humidity shown in Figure 4(d) shows greater separation for the OSE with the larger ensemble giving the best performance at all levels with the exception the 850 hPa level. A similar pattern is apparent when verifying 24 hour surface forecast fields as shown in Figures 4(e)-(f). The OSEs initialised with the larger ensembles perform better in terms of exhibiting a lower RMS error for $MSLP$ though this pattern is less obvious in W_{surf} at T+24.

Broadly speaking, we observe improved performance in terms of a reduction in RMS forecast error for the 24 hour mid-level pressure forecast and surface fields considered here. This improvement is attributable to the deployment of extra radiosonde observations in sensitive areas identified by the relevant ETKF configuration. By using all 50 members of the ECMWF ensemble in OSE 50^{24}_{24}, this permits the ETKF to sample the probability distribution functions (pdf) associated with the nine control variables more broadly. This in turn therefore leads to the generation of a more accurate estimate of the forecast error covariance at targeting and verification times.

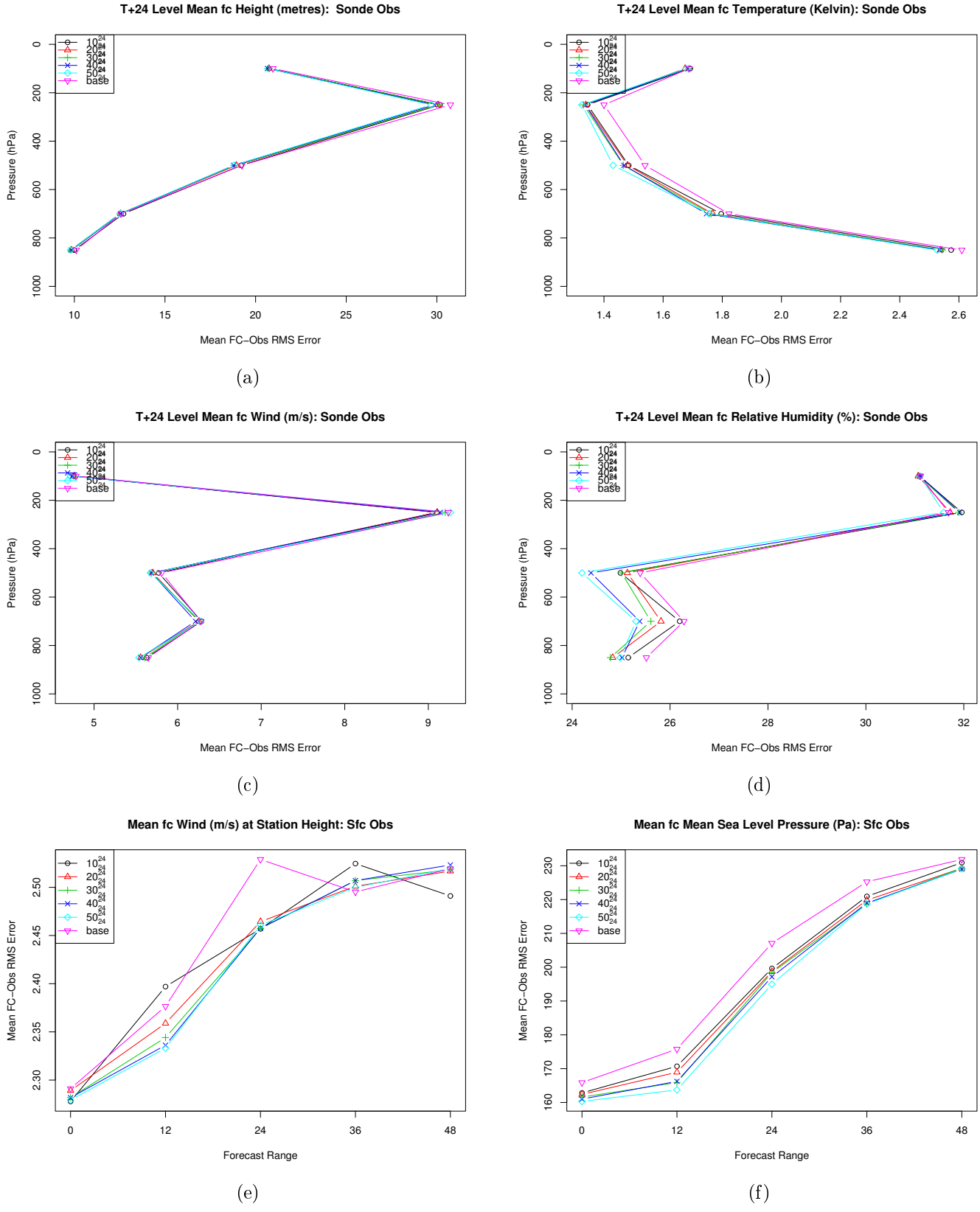


Figure 4: Mean forecast-observation RMS error obtained in the verification of (a)-(d) 24-hour geopotential height, temperature, wind vector and relative humidity against sonde observations at pressure levels $\{850, 700, 500, 250, 100\}$hPa; (e)-(f) surface wind vector and MSLP against surface observations.

Field	Pressure fields					Surface fields
	850	700	500	250	100	
Z	★	★	★	★	★	
T	★	★	★	★	★	
W	★	★	★	★	★	
RH	★	★	★	★	★	
$MSLP$						★
W_{surf}						★

Table 5: Forecast fields verified against observations in verification area. Geopotential height (Z), temperature (T), vector winds (W), relative humidity (RH), mean sea level pressure ($MSLP$) and surface vector winds (W_{surf}).

OSE	(E_n)	(L_t)		(O_t)		(E_r)	
	50	24	48	24	48	2.5	1.25
50^{24}_{24}	★	★		★		★	
50^{48}_{24}	★	★			★	★	
50^{24}_{48}	★		★	★		★	
40^{48}_{48}	★		★		★	★	

Table 6: ETKF configuration of OSEs for evaluating the effect of lead time. Each OSE is identified using the nomenclature $E_n \begin{smallmatrix} O_t \\ L_t \end{smallmatrix}$, where E_n is defined as the size of the ensemble used to initialise the ETKF, L_t is the lead time and O_t is the optimisation time defined in Figure 1. For each OSE, the resolution of the ensemble used to initialise the ETKF is $2.5^o \times 2.5^o$.

	50^{24}_{24}	50^{48}_{24}
50^{24}_{48}	70%	
50^{48}_{48}		54%

(a)

	50^{24}_{24}	50^{48}_{24}
50^{24}_{48}	60%	
50^{48}_{48}		50%

(b)

Table 7: Evaluation of the impact of lead time on a SAP. (a) Entries indicate OSE paired configurations considered for calculating $E[C]$. Actual values are the percentage of the cases in which the METS score is greater than zero for X=50. (b) Percentage of cases in which ranked correlation between any two sets of two targets are deemed similar at the 99% significance level, for $Y = 8$ boxes of size 9×9 at $Y - 2$ degrees of freedom.

5 The effect of varying lead time

This section describes the results obtained from a series of OSEs deploying and assimilating adaptive observations in SAPs created using the ETKF initialised with two different lead times ($L_t = \{24, 48\}$) at two different optimisation times ($O_t = \{24, 48\}$). Four separate OSEs are run using SAPs created from the four different configurations of the ETKF shown in Table 6. In a similar way to the evaluation undertaken of ensemble size, SAPs are generated for every 12-hour cycle throughout the trial period during which routine and targeted observation are assimilated. In evaluating we only compare ETKF configurations with the same optimisation time (O_t), that is, compare OSE pairs $50^{24}_{24}, 50^{24}_{48}$ and $50^{48}_{24}, 50^{48}_{48}$.

5.1 Evaluation of OSE SAPs

Modified Equitable Threat Score (METS)

For each OSE considered in Table 6 , the expected value of $E[C]$ is computed using Equation 16 described in Section 4.1. The results are presented in Table 7(a). The percentage of METS scores greater than zero for OSE pairs with the shorter optimisation time ($O_t = 24$) is 70%. Using the longer optimisation time ($O_t = 48$), the percentage of METS score greater than zero falls to 54%. Compared with the SAP compassion's derived from differing ensemble sizes, these results show that increasing the lead time results in SAPs with greater differences.

Ranking of neighbouring region

To analyse the box average of ETKF signal variance from different ETKF configurations we compute ρ for $G = 9$ as described in Section 4.1. Table 7(b) shows the percentage of ranked correlation of neighbouring regions that are deemed similar at the 99% significance level for each OSE defined in Table 6. For the 70 cases, 60% of the OSE pairs with the shorter optimisation time ($50^{24}_{24}, 50^{24}_{48}$) show a significant correlation compared to 50% with the longer optimisation time ($50^{48}_{24}, 50^{48}_{48}$). These results further support the impact that varying the lead time has on paired SAP comparisons.

5.2 Analysis of resulting forecast error T+24

Plots of forecast error are presented in Figures 5(a)-(f). The scores for OSE *BASE* without additional radiosonde deployments are shown as well for comparison. Little difference can be seen between forecast RMS error in Figure 5(a) for geopotential height. Differences are more apparent in Figure 5(b) for temperature at pressure levels 850

hPa and 250 hPa with a smaller RMS error apparent in OSE 50^{24}_{24} compared to 50^{24}_{48}. This reduction in RMS error is less apparent for vector winds for the OSE with the larger lead time (50^{24}_{48}) exhibiting a smaller RMS error at pressure levels 700 hPa and 250 hPa shown in Figure 5(c). By contrast, a marked reduction in RMS error can be seen from plots of T+24 relative humidity in Figure 5(d) below 250 hPa for OSE 50^{24}_{24}. Verification against surface observations gives a marginal reduction in RMS error for forecast field W_{surf} and MSLP at T+24 for OSE 50^{24}_{24} compared to 50^{24}_{48}.

The use of a shorter lead time leads to improved performance in terms of a reduction in RMS forecast error for the 24 hour mid-level pressure forecast and surface fields considered here. In these experiments the full ECMWF ensemble is used for each OSE. Therefore for each experiment, the pdf associated with the nine control variables is sampled equally. Use of a shorter lead time and therefore more up-to-date NWP forecast fields, leads to a more accurate estimate of the true pdf which is subsequently used to give an estimate of the future forecast error covariance.

5.3 Analysis of resulting forecast error T+48

Analysis of the verification for OSEs 50^{48}_{24} and 50^{48}_{48} with the larger optimisation time ($O_t = 48$) against sonde and surface observations are presented in Figures 6(a)-(f). The scores for OSE *BASE* without additional radiosonde deployments are shown as well for comparison. For all forecast fields, there is little difference between the two OSEs and clear differences sometimes harder to identify. For instance, little discernible difference in RMS error is apparent in the verification of geopotential height at any level in Figure 6(a). There is a small reduction in RMS error across all levels for the OSE with larger lead time (50^{48}_{48}) compared with the smaller (50^{48}_{24}) for temperature as shown in Figure 6(b). This signal is also apparent to a lesser extent for verification of vector winds shown in Figure 6(c). Verification of the OSEs against surface observations give contradicting results. Against forecast field W_{surf} shown Figure 6(e) at T+48, OSE 50^{48}_{24} shows a reduction if RMS error compared with the larger lead time (50^{48}_{48}) though this signal is not apparent for $MSLP$ as shown in Figure 6(f) with both OSEs giving identical RMS error scores.

For the mid-level and surface 48 hour forecast fields considered here, results for the different lead times appear mixed. With a longer forecast range and lead time, a total 96 hour ensemble range between ensemble initialisation time (t_i) and verification time (t_v) is employed. The ETKF utilises a linearised observation operator ($\mathbf{H}^q$) for the deployment of observations to interpolated ensemble perturbations at these times. It is unlikely that any linear assumptions will hold at this ensemble range thereby explaining the mixed results obtained in this series of experiments.

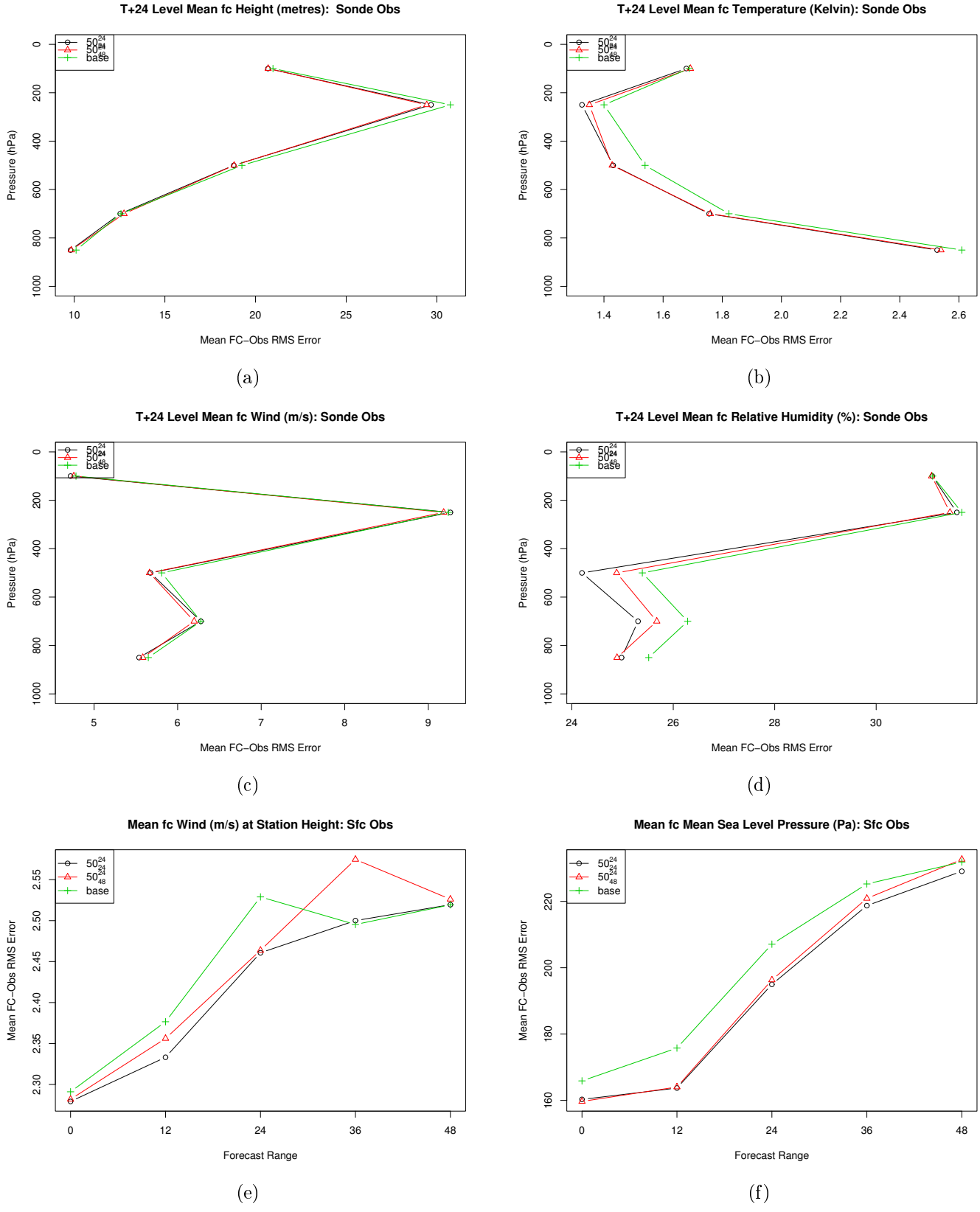


Figure 5: Mean forecast-observation RMS error obtained in the verification of (a)-(d) T+24 geopotential height, temperature, wind vector and relative humidity against sonde observations at pressure levels $\{850, 700, 500, 250, 100\}$hPa; (e)-(f) surface wind vector and MSLP against surface observations.

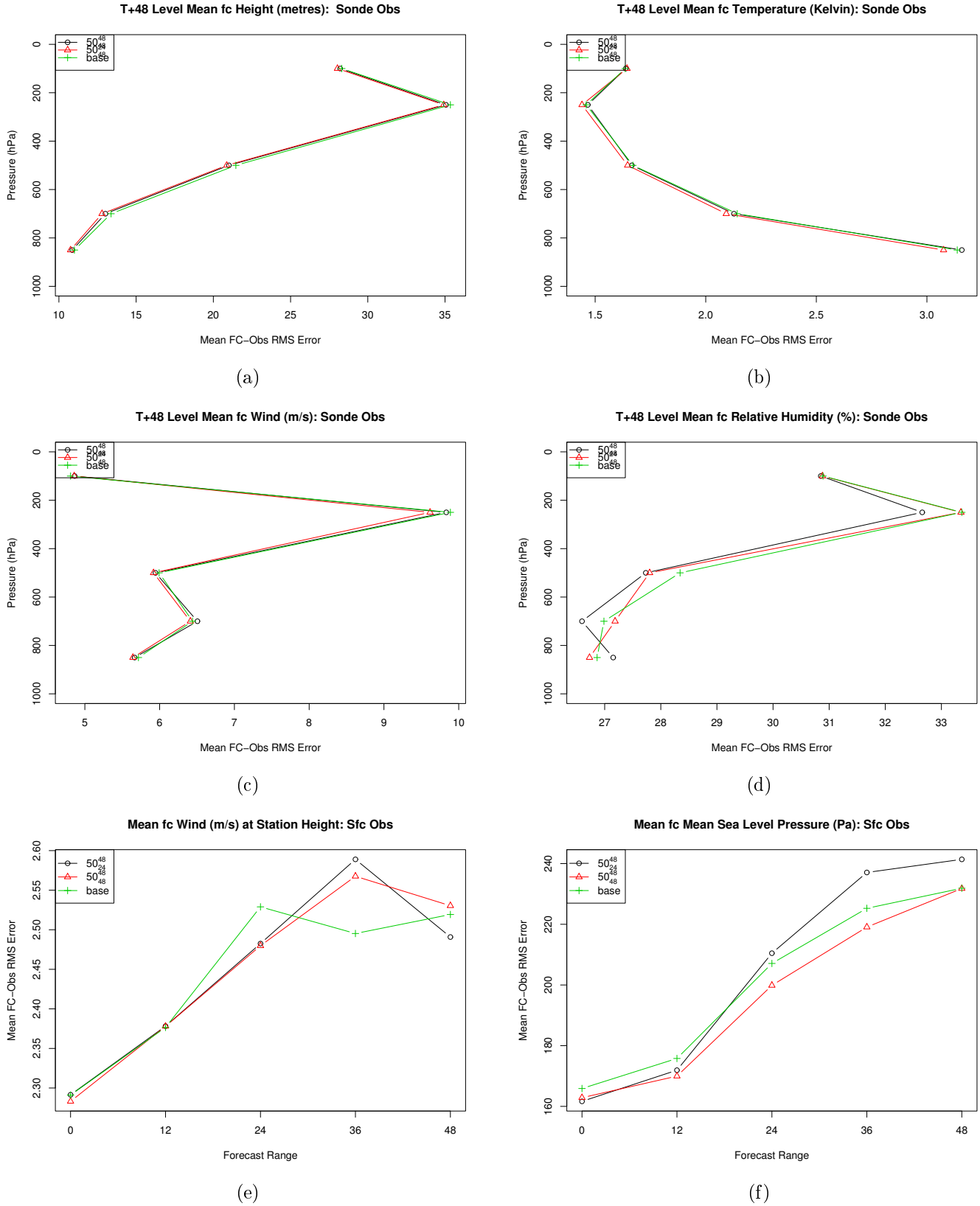
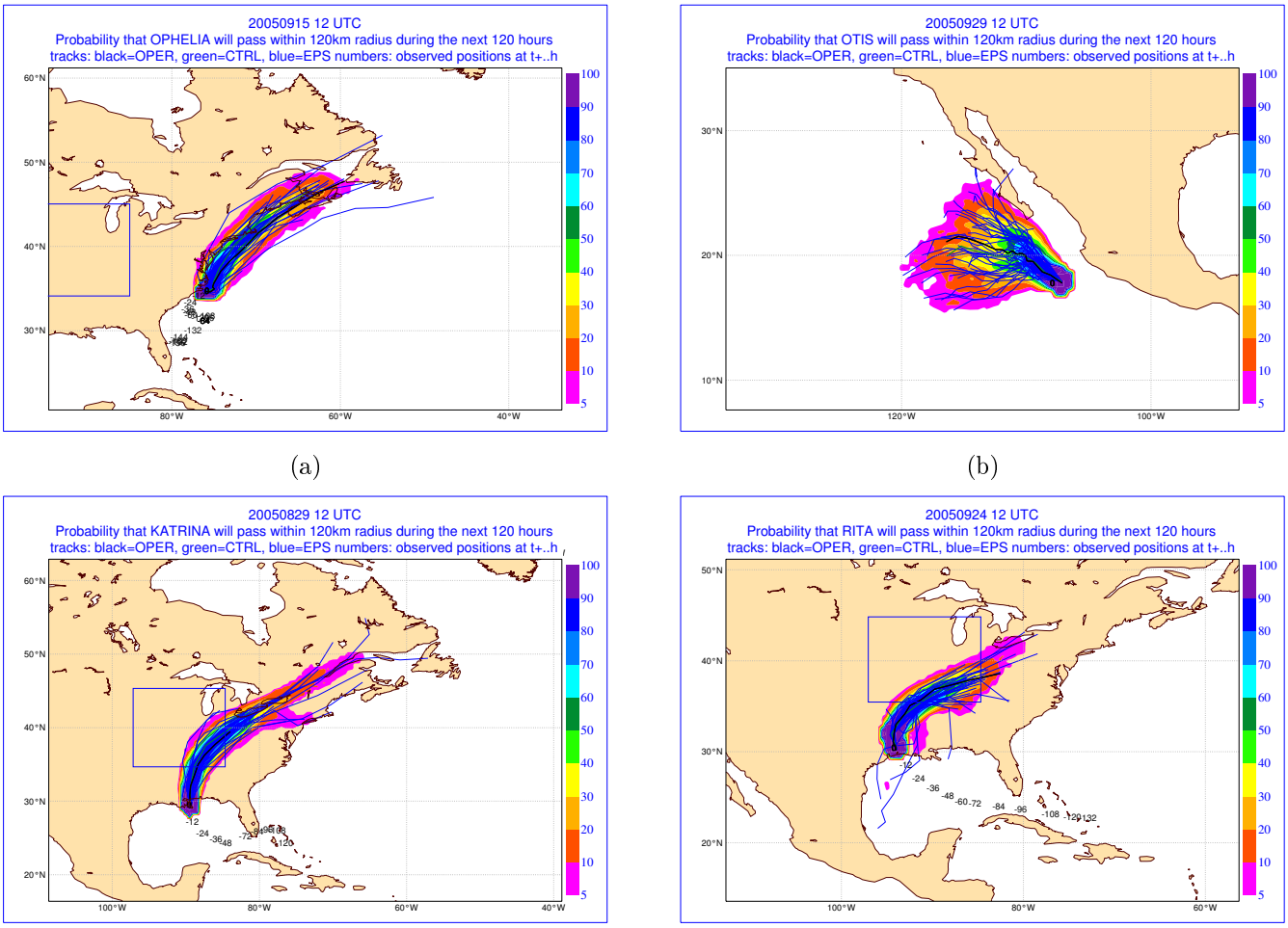


Figure 6: Mean forecast-observation RMS error obtained in the verification of (a)-(d) T+48 geopotential height, temperature, wind vector and relative humidity against sonde observations at pressure levels $\{850, 700, 500, 250, 100\}$hPa; (e)-(f) surface wind vector and MSLP against surface observations.



(a) (b)

(c) (d)

Figure 7: Forecast trajectory probability (coloured contour), verification area (blue box), ECMWF ensemble member trajectory (blue lines), ECMWF operational trajectory (black line) and observed position (digits) for hurricanes: (a) Ophelia on 15 September 2005 12 UTC; (b) Otis on 29 September 2005 12 UTC; (c) Katrina on 29 August 2005 12 UTC and (d) Rita on 24 September 2005 12 UTC.

6 Case studies

In this section we evaluate different ETKF configurations and present results of the impact of deploying targeted observations with the aim of improving 24-hour forecast of tropical cyclones entering the pre-defined verification region. In Section 7, we provide an analysis of these results and discuss them further. During this trial period four tropical cyclones are active within the LAM domain. We discount hurricanes Ophelia and Otis whose forecast trajectories are shown in Figures 7(a) and (b) and can be seen not to enter the verification region. In contrast, the tracks of hurricanes Katrina and Rita (Figures 7(c) and (d)) do cross the verification region and are considered further. For the Katrina and Rita cases, we present results from the verification of 1-day forecasts as each enter the verification region. An independent analysis is used in the verification of each case. We consider the paired verification of forecast $MSLP$ resulting from the deployment of targeted observations from ETKF guidance initialised from the smallest and largest ensembles $10^{24}_{24}, 50^{24}_{24}$ and from two different lead times $50^{24}_{24}, 50^{24}_{48}$.

6.1 Hurricane Katrina case

Figure 8(a) shows the verification of T+24 $MSLP$ for OSE pair $10^{24}_{24}, 50^{24}_{24}$ (left and right respectively) for the Katrina case. For this forecast field, the RMS error within the verification area is smaller in the 10-member ensemble ($10^{24}_{24} = 0.548$) compared with the 50-member ensemble ($50^{24}_{24} = 0.583$). A qualitative examination of these plots though, reveals that a larger area just south-west of the central pressure at $87^oW, 35^oN$ has a forecast

$MSLP$ between 4.6 and 6.4 Pa lower than the verifying analysis (shown in purple in Figure 8(a)) in OSE 10^{24}_{24} compared with OSE 50^{24}_{24}. Figure 8(b) shows the RMS error for OSE pair $50^{24}_{24}, 50^{24}_{48}$ (left and right respectively). The RMS error obtained using the longer lead time ($50^{24}_{48} = 0.514$) is approximately 12% smaller than that obtained using a shorter lead time ($50^{24}_{24} = 0.583$). In addition, a further qualitative evaluation of the forecast error highlights that the forecast $MSLP$ is between 4.6 and 6.4 Pa lower (shown in purple in Figure 8(b)) for a similar central area associated with Katrina in OSE 50^{24}_{48} compared with OSE 50^{24}_{24}.

6.2 Hurricane Rita case

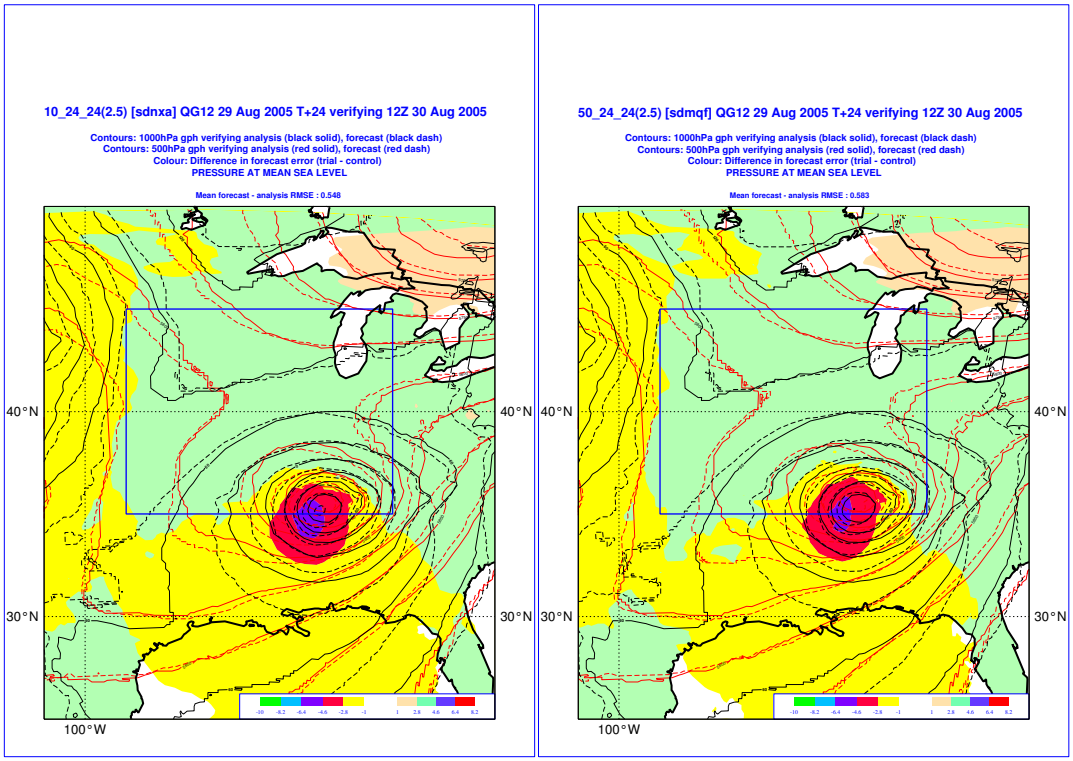
Figure 9(a) shows the verification of T+24 $MSLP$ for OSE pair $10^{24}_{24}, 50^{24}_{24}$ (left and right respectively) for the Rita case. The RMS error within the verification area is over 18% smaller in the 50-member ensemble ($50^{24}_{24} = 1.313$) compared with the 10-member ensemble ($10^{24}_{24} = 1.612$). Most of this reduction in RMS error is obtained in the north-west of the verification region when Rita is approaching its southern boundary. A qualitative analysis of this plot indicates that a larger annulus surrounding Rita's centre at $98^oW, 33^oN$ shown in red in Figure 9(a), corresponding to an area with a forecast $MSLP$ between 2.6 and 4.6 Pa lower than the verifying analysis exists in OSE 10^{24}_{24}. Figure 9(b) shows the RMS error for OSE pair $50^{24}_{24}, 50^{24}_{48}$ (left and right respectively) . The RMS error obtained from using the shorter lead time ($50^{24}_{24} = 1.313$) is approximately 11% smaller than that using the longer lead time ($50^{24}_{48} = 1.481$) and a further qualitative evaluation highlights that the larger annulus of incorrectly forecast $MSLP$ in OSE 50^{24}_{48}.

7 Discussion

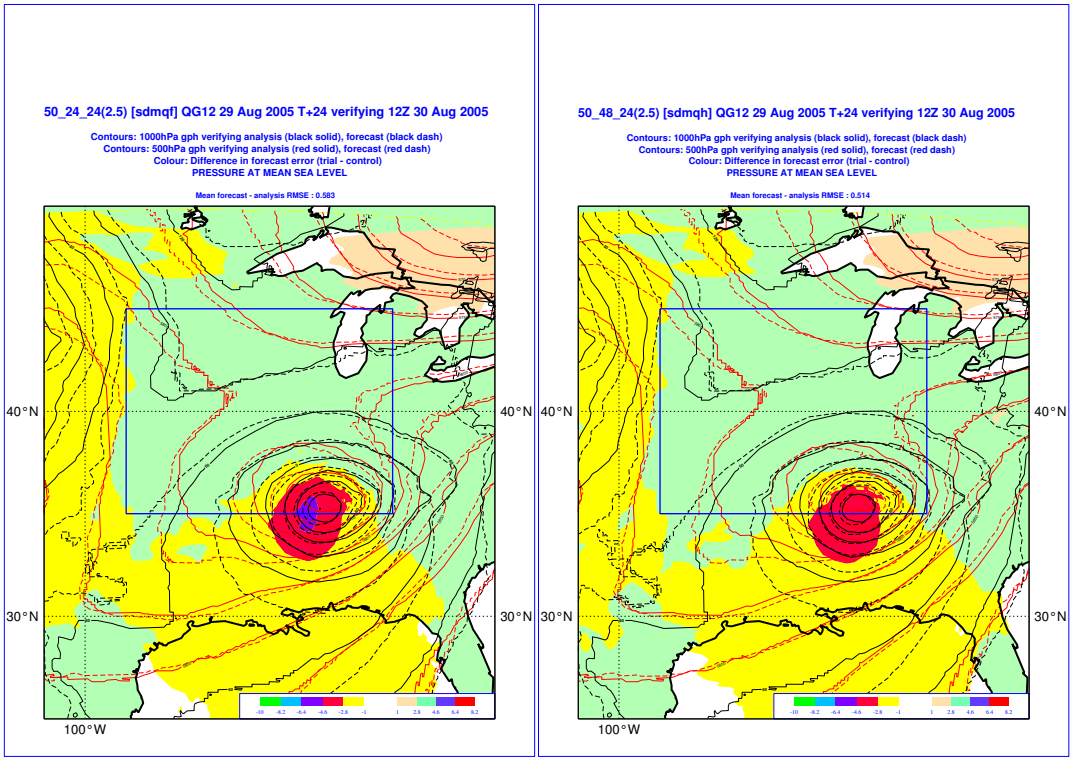
In Section 4.2 we presented mean verification results of experimentation investigating the effect of ensemble size on ETKF targeting guidance. Results presented indicate that a deployment of targeted observations in areas identified with ETKF guidance derived from a larger sized ensemble, leads to reduced forecast error. In contrast, the results for the Katrina case presented in Section 6.1 indicate that this is not necessarily true for an individual high-impact event using guidance with the same lead and optimisation times. Similarly results for the same case show that using a longer lead time, and therefore a less up-to-date forecast to initialise the ETKF, may not necessarily result in inferior targeting guidance in terms of RMS error. In this section we investigate the properties of the ensembles used by the ETKF to estimate the analysis error covariance matrix from which targeting guidance for the Katrina and Rita cases is derived.

Figures 10(a)-(d) show the time-series of T+24 forecast 500 hPa wind vector RMS error verified against sonde observations within the verification area for the Rita and Katrina cases. Figures 10(a) and (b) show results for the Katrina case verifying at 30 August 2005 12 UTC and Figures 10(c) and (d) show results for the Rita case verifying at 26 September 2005 00 UTC . In each plot the black vertical dashed lines in this figure indicate the period when the tropical cyclone enters and leaves the verification area. From the ensemble size experiments previously undertaken, targeting guidance generated from the smallest ensemble lead to the largest OSE forecast error. Results for the Katrina case in Figure 10(a) and the Rita case in Figure 10(c) show that OSE 10^{24}_{24} and 20^{24}_{24} respectively, give the smallest forecast error at verification time. Similarly, although the targeting guidance initialised with the shortest lead time used in 50^{24}_{24} out-performs that with the longer lead time, OSE 50^{24}_{48} for the Rita case, this is not true in the Katrina case. As all OSEs are initialised using an identical background and start dump, and the observations used only differ in those that have been targeted, observed differences in verification results are attributed to the configuration of the additionally deployed observations.

Figures 11(a)-(c) show the targeting guidance for improving T+24 forecasts used in OSEs 10^{24}_{24}, 50^{24}_{24} and 50^{24}_{48} respectively for the Katrina case. Figures 11(d)-(f) show the targeting guidance for improving T+24 forecasts used in OSEs 20^{24}_{24}, 50^{24}_{24} and 50^{24}_{48} respectively for the Rita case. In the absence of a diagnostic tool to estimate individual observation impact on short range forecast error, we seek to explain differences in forecast error by qualitatively assessing the location of targeted observation with respect to the underlying 500 hPa flow. For the Katrina case, the underlying mid-level flow is from an area north-west of the verification region. Accurate modelling of this air mass (shown as a red dashed ellipse) as it engages with the remnants of Katrina (shown as a black dashed ellipse) entering from the south might be one possible explanation for differences in forecast error observed. For the best performing OSE 50^{24}_{48} shown in Figures 11(c), three observations have been targeted in this area together with four in a region just ahead of Katrina to the south of the verification area. OSEs 50^{24}_{24} and 10^{24}_{24} have fewer targeted observations deployed in an airmass (shown as a red dashed ellipse) upstream of the verification region shown in Figures 11(b) and (a). Adopting a similar hypothesis that accurate modelling of the air mass upstream of the verification region that engages with the remnants of Rita (shown as a black dashed ellipse), may also

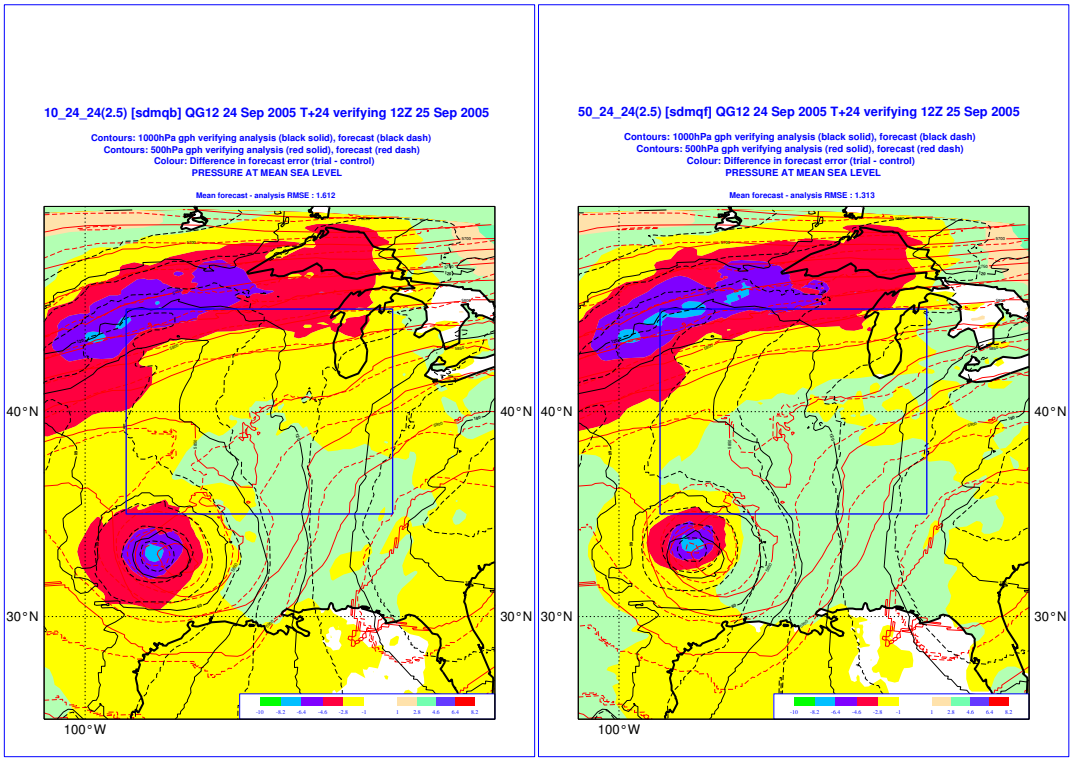


(a)

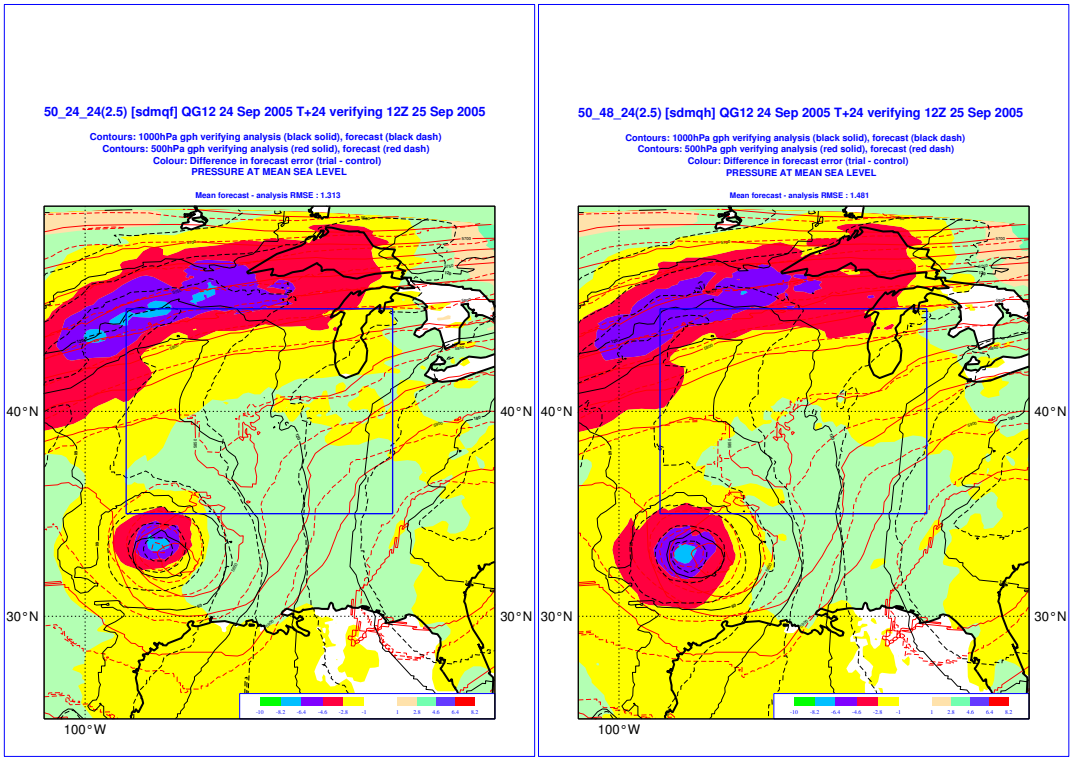


(b)

Figure 8: Verification of T+24 $MSLP$ RMS forecast error against an independent all data run analysis for ex-Hurricane Katrina verifying 30 August 2005 12 UTC for (a) OSE pair $10_{24}^{24}, 50_{24}^{24}$ (left and right respectively) and (b) OSE pair $50_{24}^{24}, 50_{48}^{24}$ (left and right respectively) . The verification region is shown as a blue box. In both plots, the filled purple region identifies a region at $87^oW, 35^oN$ where the forecast $MSLP$ is between 4.6 and 6.4 Pa too low compared with the verifying analysis.



(a)

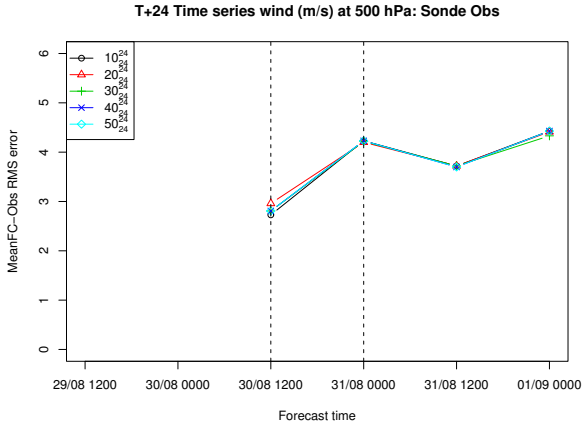


(b)

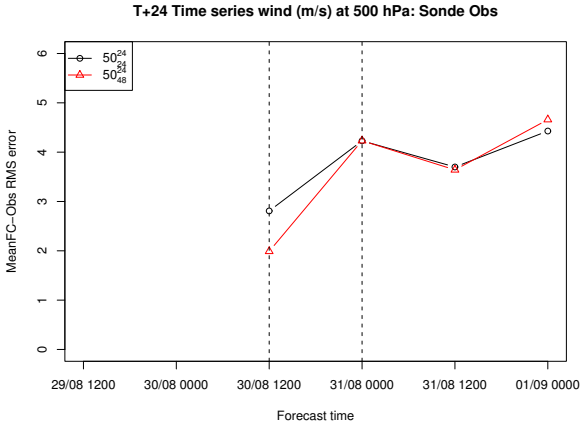
Figure 9: Verification of T+24 $MSLP$ RMS forecast error against an independent all data run analysis for ex-Hurricane Rita verifying 25 September 2005 12 UTC for (a) OSE pair $10_{24}^{24}, 50_{24}^{24}$ (left and right respectively) and (b) OSE pair $50_{24}^{24}, 50_{48}^{24}$ (left and right respectively) . The verification region is shown as a blue box. In both plots, the filled red region identifies a region at $98^oW, 33^oN$ where the forecast $MSLP$ is between 2.6 and 4.6 Pa too low compared with the verifying analysis.

explain differences in forecast error for this case. In the worst performing OSE for the Rita case, 50^{24}_{48}, no targeted observations were deployed around Rita directly south of the verification region although the upstream air mass was well sampled shown in Figure 11(f). By contrast, the second best performing OSE 50^{24}_{24} deploys five observations in the vicinity of Rita and five upstream although none are further west than 110^oW of longitude shown in Figure 11(e). The observations deployed in the best performing OSE for the Rita case, OSE 20^{24}_{24} are broadly similar but an additional observation deployment just ahead of the remnants of Rita in the black delineated ellipse shown in Figure 11(d). If we assume the mid-level flow is an important factor for the difference in impact seen in each OSE, we go on to examine the information contained within each ensemble to understand the data underpinning each set of targeting guidance generated. The ETKF uses the spread of perturbations from each ensemble member as a basis for model uncertainty at targeting and verification time. This is subsequently used an an estimate of the analysis error covariance matrix at the different times.

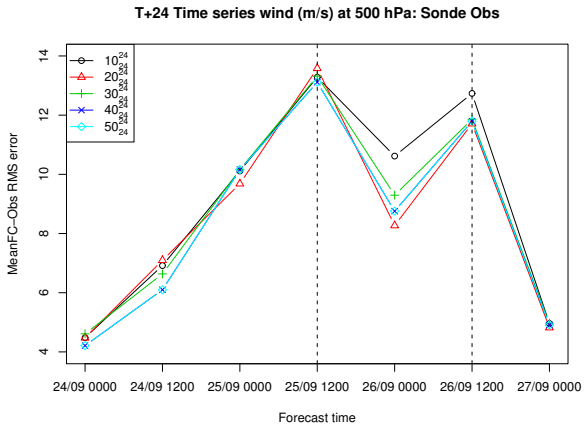
Figures 12(a)-(c) show the spread of ensemble member perturbations for forecast temperature at 500 hPa at targeting time, when Katrina is forecast to be in an area centred at 27^oN, 90^oW for OSEs 10^{24}_{24}, 50^{24}_{24} and 50^{24}_{48}. Figures 12(d)-(f) show the same information at verification time when Katrina is forecast to be centred in a similar area. In Section 2.7 we described how targeting guidance is generated from ETKF signal variance. It is this spread of ensemble member perturbations that the ETKF uses, in estimating the analysis error covariance matrix forming the basis of the summary maps. It can be seen that the spread of perturbations at targeting and verification time in Figures 12(c) and (f) are more localised at 27^oN, 90^oW in OSE 50^{24}_{48} compared with OSEs 10^{24}_{24} and 50^{24}_{24} in Figures 12(a) and (d) and Figures 12(b) and (e) respectively. For this case, from these plots it can be seen that the spatial structure of the spread of member perturbations is smaller when the ETKF utilises an ensemble with a longer lead time in OSE 50^{24}_{48}. This results in fewer observations being deployed in this area at verification time leaving more available for deployment upstream of the verification region in the air mass (not shown on these plots) that engages with Katrina. The Rita case by contrast highlights the effect of failing to capture the uncertainty associated with this tropical cyclone. OSE 50^{24}_{48} is the poorest performing OSE for this case failing to deploy any targeted observations south of the verification region as seen in Figure 11(f). The ensemble has failed to capture any perturbation spread in this area at targeting time as shown in Figure 12(i) or within the verification region at verification time in Figure 12(l). By contrast the marginally best performing OSE 20^{24}_{24} correctly captures a localised spread of perturbations due south of the verification region at targeting time shown in Figure 12(g) and yet still has the remnants of Rita in the verification region at verification time in Figure 12(j). OSE 50^{24}_{24} has a similar spatial distribution of ensemble member perturbations spread at targeting time shown in Figure 12(h) and at verification time in Figure 12(k). This explains the similar performance in terms of RMS error, seen in OSEs 20^{24}_{24} and 50^{24}_{24} verifying on 26 September 2005 00 UTC shown in Figures 10(c) and (d).



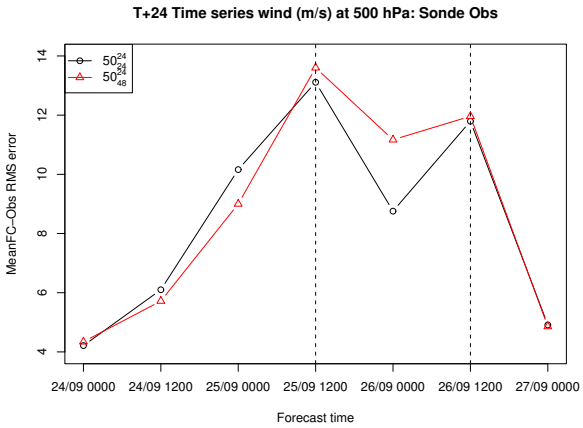
(a)



(b)



(c)



(d)

Figure 10: Time-series of mean forecast - observation RMS error for T+24 W500 verified against radiosonde observations within the verification region. Katrina case verifying 30 August 2005 12 UTC (a) for different ensemble sizes and (b) different lead times. Rita case verifying 26 September 2005 00 UTC (c) for different ensemble sizes and (d) different lead times. Black dotted line indicates period that ex-Rita enters the verification region.

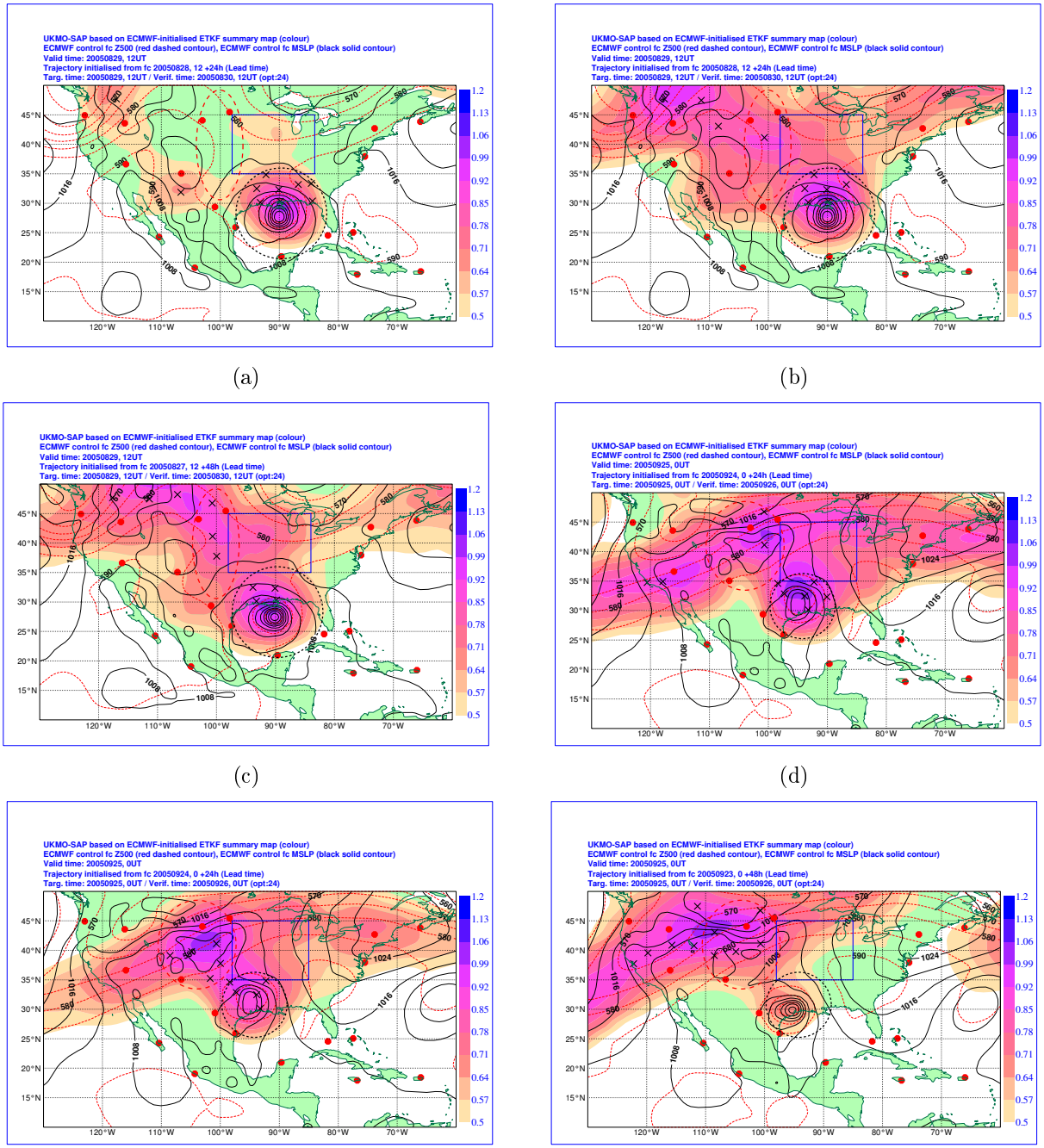


Figure 11: Katrina case ETKF targeting guidance at 29 August 2005 00 UTC targeting guidance with a lead time of 24 hours and optimisation time of 24 hours for (a) 10^{24}_{24}, (b) 50^{24}_{24}, (c) 50^{24}_{48}. Rita case ETKF targeting guidance at 25 September 2005 00UTC targeting guidance with a lead time of 24 hours and optimisation time of 24 hours for (d) 20^{24}_{24}, (e) 50^{24}_{24} and (f) 50^{24}_{48}. The 500 hPa mid-level flow area upstream of the verification region is shown by a red dashed ellipse. Remnants of each tropical cyclone are delineated by a dashed black ellipse.

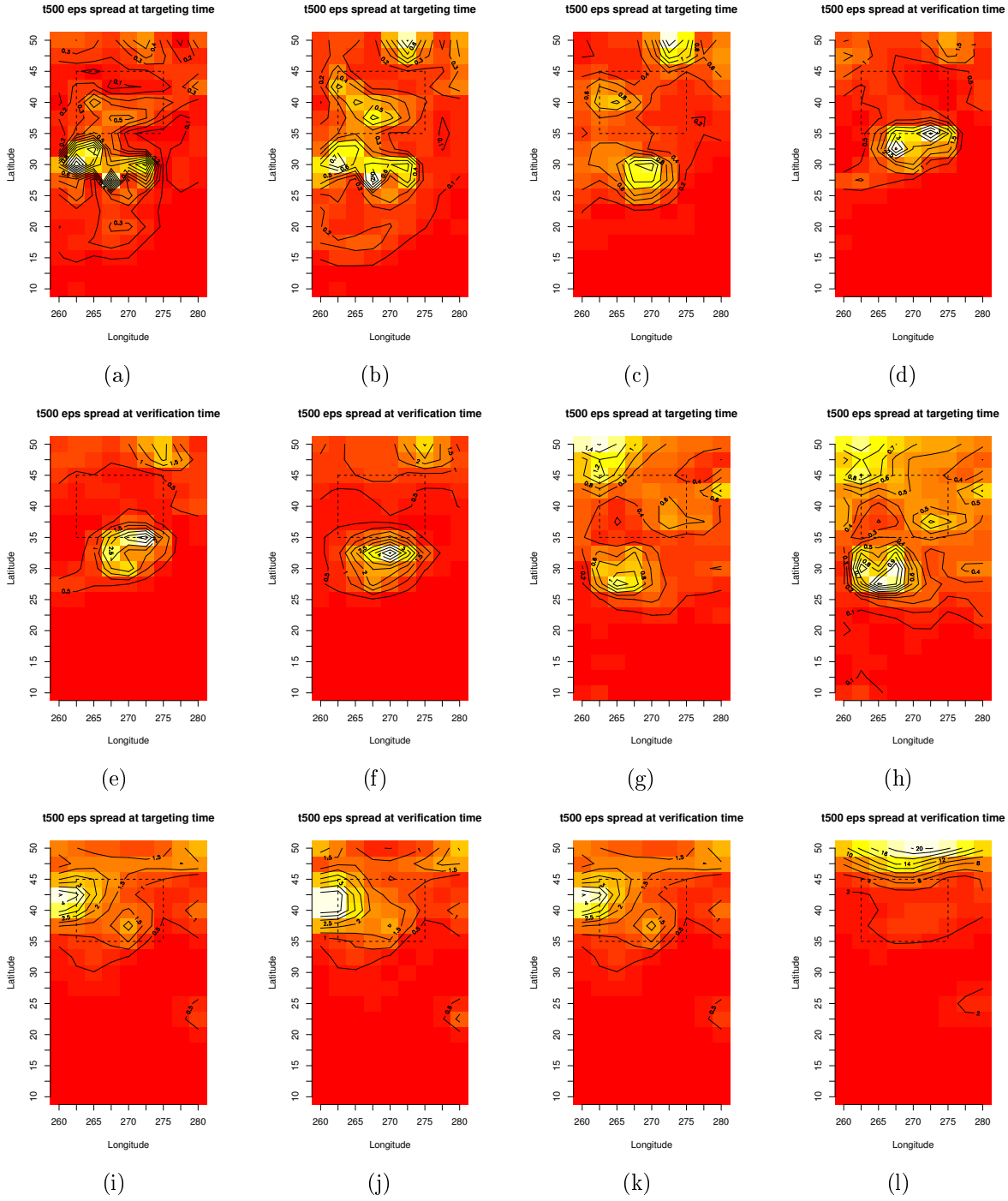


Figure 12: Plots of the spread of ensemble member forecast perturbations at 500 hPa at targeting time for Katrina case (a) 10^{24}_{24}, (b) 50^{24}_{24}, (c) 50^{24}_{48} and at verification time for (d) 10^{24}_{24}, (e) 50^{24}_{24}, (f) 50^{24}_{48}. Targeting time is 29 August 2005 12 UTC and verification time 30 August 2005 12 UTC. Plots of at targeting time for Rita case (g) 20^{24}_{24}, (h) 50^{24}_{24}, (i) 50^{24}_{48} and at verification time for (j) 20^{24}_{24}, (k) 50^{24}_{24}, (l) 50^{24}_{48}. Targeting time is at 25 September 2005 00 UTC and verification time on 26 September 2005 00 UTC. Dashed box denotes verification region.

8 Conclusions

In this study we have generated SAPs for each 12-hour cycle for a month-long trial period using different configurations of the ETKF. The utility of the targeting guidance from each configuration has been assessed by running a

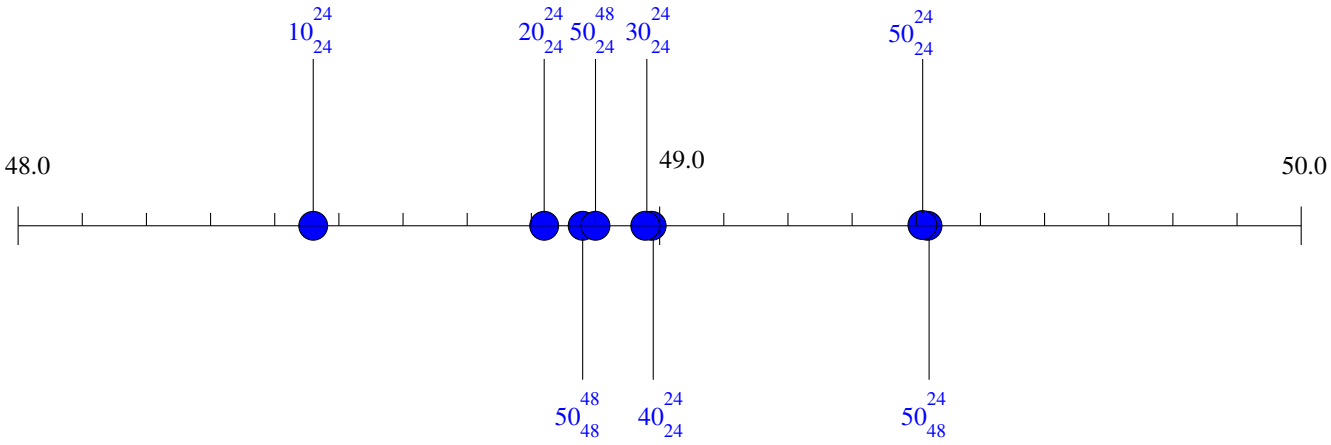


Figure 13: Visualisation of overall skill score obtained from each OSE.

series of OSEs. We have evaluated the performance of each OSE by examining a selection of forecast fields at the appropriate forecast ranges. The performance of each OSE can also be further assessed in terms of an overall skill score compiled from the complete set of forecast fields during the trial period verified against surface and sonde observations at different forecast ranges. These skill scores are shown in Figure 13 with the best performing OSEs lying to the right of the line and poorest performing OSEs to the left.

From our experimentation we have seen that larger ensembles of similar size have a larger percentage of cases with METS scores greater than zero. This indicates that these configurations of the ETKF are more likely to contain common grid points in their targeting guidance. This is in contrast to disparate sized ensemble configurations that exhibit a smaller percentage of cases with a METS score greater than zero indicating that they have fewer common targeting grid points. This pattern is also seen in the ranked correlation of boxed-average regions. Lower percentage METS scores greater than zero and significant ranked correlations of boxed-average regions, indicate that the targeted regions for the shorter optimisation time were more similar in these OSE pairs compared with the larger optimisation time. Smaller forecast error is seen in the OSEs with the shorter lead time for T+24 $MSLP$ and temperature at different pressure levels but for longer optimisation times, the converse is true with the longer lead times giving the better performance. The overall skill scores shown in Figure 13 confirm this result suggesting that better performance is obtained with a longer lead time for the shorter 24 hour optimisation time. It is important to bear in mind though that the overall skill scores are calculated from the forecast error of *all* candidate forecast ranges not simply that chosen as the optimisation time. There is no guarantee that targeting observations will necessarily improve the forecast error at all forecast ranges, but the overall skill score does serve as a cursory metric to rank the impact of each OSE.

Our analysis of two tropical cyclone cases studies has shown two important attributes of an ensemble used by the ETKF in targeting high-impact weather. Firstly, with a finite number of targeting observations available, the ability of an ensemble to generalise on the synoptic scale is vital. Larger ensembles possessing many more degrees of freedom and ensembles with shorter lead times possessing more accurate NWP, appear suitable to model small scale areas of uncertainty. In the case of Katrina, this was seen at the expense of failing to facilitate the deployment of a share of the targetable observations that adequately capture the broad scale. Clearly human forecaster intervention would prevent some of the localised deployment seen in this case. Secondly, whilst generalising, the ensemble must be able to maintain a reasonable level of sensitivity to small scale features. It is unclear whether increasing the resolution of the ensembles would improve this at the expense of generality, though there may be an optimal balance obtainable by a multi-model ensemble of different sized ensembles initialised with different lead times.

One possible enhancement under consideration in the generation of SAPs, is the adoption of sequential ETKF. In this approach error covariances associated with the adaptive network are re-evaluated after every individual targeted observation deployment. An overhead would be incurred from the iterative generation of a new sensitivity map for each targetable observations. This would be created in the presence of a routine network *and* the previously deployed targeting observations. A perceived benefit of this approach is that it may lead to more optimal guidance and avoid problems of targeting in one particular region. Conversely, it may be less easy to design a feasible flight plan for dropsonde deployment from such guidance due to the larger spread of sensitive regions.

References

[1] T. Bergot. Adaptive observations during FASTEX: A systematic survey of upstream flights. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, 125(561):3271–3298, 1999.

[2] C. H. Bishop, B. J. Etherton, and S. J. Majumdar. Adaptive sampling with the ensemble transform Kalman Filter. part I: Theoretical aspects. *MONTHLY WEATHER REVIEW*, 129(3):420–436, 2001.

[3] C. H. Bishop and Z. Toth. Ensemble transformation and adaptive observations. *Journal of the Atmospheric sciences*, 56(11):1748–1765, 1999.

[4] R. Buizza, J. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart. The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quarterly Journal of the Royal Meteorological Society*, 133(624):681–695, 2007.

[5] A. Doerenbecher, M. Leutbecher, and D. S. Richardson. Impact comparison of observation targeting predictions during the (North-) Atlantic TReC 2003. In *The First THORPEX International Science Symposium*, Montreal, Canada, December 2004.

[6] R. Dumelow. Space-terrestrial study at the Met Office - initial report for EUCOS on the summer period experiment. 2007.

[7] G. Greed. Upgrades to the Crisis Area Mesoscale Model service. Technical Report 464, Met Office Forecasting Research Technical Report, 2005.

[8] K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc. Unified notation for data assimilation: operational, sequential and variational. *Journal of the Meterological Society of Japan*, 75(1B):181–189, March 1997.

[9] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[10] R. H. Langland, Z. Toth, R. Gelaro, I. Szunyogh, M. A. Shapiro, S. J. Majumdar, R. E. Morss, G. D. Rohaly, C. Velden, N. Bond, and C. H. Bishop. The North Pacific experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bulletin of the American Meteorolical society*, 80(7):1363–1384, 1999.

[11] M. Leutbecher, A. Doerenbecher, F. Grazzini, and C. Cardinali. Planning of adaptive observations during the Atlantic THORPEX Regional Campaign 2003. Technical Report 102, ECMWF newsletter, 2004.

[12] S. J. Majumdar, S. D. Aberson, C. H. Bishop, R. Buizza, M. S. Peng, and C. A. Reynolds. A comparison of adaptive observing guidance for Atlantic tropical cyclones. Technical Report 482, ECMWF Technical Memorandum, 2005.

[13] S. J. Majumdar, C. H. Bishop, R. Buizza, and R. Gelaro. A comparison of ensemble-transform Kalman-Filter targeting guidance with ecmwf and nrl total-energy singular-vector guidance. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, 128(585):2527–2549, 2002.

[14] S. J. Majumdar, C. H. Bishop, B. J. Etherton, and Z. Toth. Adaptive sampling with the ensemble transform Kalman Filter. part II: Field program implementation. *MONTHLY WEATHER REVIEW*, 130(5):1356–1369, 2002.

[15] D. Mansfield, D. Richardson, and B. Truscott. An overview of the Atlantic THORPEX Regional Campaign. In *The First THORPEX International Science Symposium*, Montreal, Canada, December 2004.

[16] T. N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *JOURNAL OF THE ATMOSPHERIC SCIENCES*, 55(4):633–653, 1998.

[17] G. N. Petersen, S. J. Majumdar, and A. J.Thorpe. The properties of sensitive area predictions based on the ensemble transform Kalman Filter (ETKF). *Quarterly Journal of the Royal Meteorological Society*, 133(624):697–710, 2007.

[18] G. N. Petersen and A. J. Thorpe. The impact on weather forecasts of targeted observations during A-TReC. *Quarterly Journal of the Royal Meteorological Society*, 133:417–431, 2007.