Tellus (2006), 58A, 179–195
Printed in Singapore. All rights reserved

TELLUS

An examination of ensemble filter based adaptive observation methodologies

By S. P. KHARE* and J. L. ANDERSON, *Statistical and Applied Mathematical Sciences Institute, 19 T.W. Alexander Drive, RTP, NC, 27709, USA, and National Center for Atmospheric Research, Boulder, CO, 80307-3000, USA*

(Manuscript received 31 March 2005; in final form 19 September 2005)

ABSTRACT

The type of adaptive observation (AO) schemes of interest in this paper are those which make use of an ensemble forecast generated at a given initial time. The ensemble forecast can be used to quantify the influence of hypothetical observational networks on forecast error covariances. The ensemble transform kalman filter (ETKF) scheme is an example of such a scheme and is used operationally at the National Centers for Environmental Prediction (NCEP). A Bayesian framework for ETKF schemes is developed in this paper. New ETKF AO schemes that make use of covariance localization (CL) are introduced. CL is a technique used to alleviate problems due to sampling errors when estimating covariances from finite samples. No previous study has developed ETKF schemes that make use of CL. A series of observing system simulation experiments (OSSEs) in the non-linear Lorenz 1996 model are used to develop a fundamental understanding of ETKF methods. The OSSEs simulate the problem of choosing observations in a large data void region, to improve forecasts in a verification region located within the data void region. The results demonstrate the important role that techniques for alleviating problems due to sampling errors play in improving the performance of ensemble-based AO techniques.

1. Introduction

Severe storms can have devastating economic and social impacts. It is often the case that storms are not adequately observed, leading to poor forecasts by operational centres. To improve this situation, movable observing platforms can be used. Vertical/horizontal profiles of various dynamic variables can then be relayed to operational centres. This additional information can be fused with the prediction model in hopes of improving the subsequent forecast of the storm. This procedure, known as 'adaptive observing' was originally suggested by Emanuel et al. (1995). Adaptive observing holds the promise of dramatically improving Numerical Weather Prediction (NWP). Recently, the use of adaptive observations (AOs) in mid-latitude weather forecasting has been tested in several field experiments and operational programs (FASTEX, NORPEX, Winter Storm Reconnaissance Program) (Bergot, 1999; Langland et al., 1999; Montani et al., 1999; Pu and Kalnay, 1999; Gelaro et al., 2000; Szunyogh et al., 2000, 2002; Majumdar et al., 2001, 2002a,b). AOs are also used in hurricane track forecasting (Aberson, 2003). AOs are one of the key focus areas of THORPEX, a current international predictability research campaign (THORPEX International Science Plan, Version III, Shapiro and Thorpe, 2004; see Web site www.mmm.ucar.edu/uswrp/programs/thorpex.html) (hereafter ST04). While the various field programs over the last decade indicate that positive benefits are obtainable from AOs, further fundamental understanding is required before the full benefits of AOs can be realized (ST04).

The optimal placement of AOs in space and time is dependent on several factors that include the system dynamics, model errors, the data assimilation (DA) scheme, background errors and the configuration and accuracy of the present day observing network. Effective AO methodologies must be able to account for the subtle interplay between such factors. Given the complexity of the problem, it is not surprising that a number of different approaches have been developed. Singular vector techniques (Palmer et al., 1998) and model adjoint sensitivity-based techniques (Bergot et al., 1999) have been developed and tested. Baker and Daley (2000) develop and test an observation/background adjoint sensitivity technique that can be used to assess the influence of hypothetical observations. Leutbecher (2003) develops and tests a technique based on Hessian singular vectors. Monte Carlo (ensemble)-based techniques can be used in an attempt to quantify the impact of hypothetical observations on the uncertainty of the system state. A number of studies have explored the use of ensemble-based methods (Lorenz and Emanuel, 1998; Bishop and Toth, 1999; Hansen and Smith, 2000; Bishop et al., 2001; Hamill and Snyder, 2002; Khare, 2004).

*Corresponding author.
e-mail: khare@ucar.edu
DOI: 10.1111/j.1600-0870.2006.00163.x

With the growing availability of ensemble forecasting systems, further exploration of ensemble-based methods is well motivated. Ensemble-based AO methodologies are the subject of concern in this paper.

To illustrate the type of ensemble filter-based method of primary concern in this paper, consider the following example problem. Let the initial (present) time be t_i. Suppose one wants to determine the optimal location to place an AO at a future time $t_a > t_i$ ('a' for adaptive), to improve the forecast at $t_v > t_a$ ('v' for verification), for a forecast to be initialized at t_a. Geophysical systems of interest are characterized by many degrees of freedom and high-dimensional state spaces. AO methodologies that require repeated integrations of the prediction model equations may not be suitable for real-time applications due to prohibitive computational expense. Therefore, *the types of ensemble-based algorithms of interest in this study are those which make use of an ensemble forecast initialized at t_i and valid up to t_v, without requiring any additional forecast model integrations*. In Section 3.2, AO algorithms based on Bayes rule are described. These Bayesian algorithms consider the ensemble forecast generated at t_i as a sample of a prior probability density function (pdf). The prior ensemble is updated using an ensemble Kalman filter (EnKF)-based algorithm. When a deterministic ensemble square root Kalman filter (DEnSRF) (Tippett et al., 2003) update algorithm is used *without covariance localization*, the method is equivalent to the ETKF AO methodology derived and tested by Bishop et al. (2001) (hereafter B01) (shown in Section 3.2.2). For realistic applications of EnKF, the number of ensemble members is typically far less than the number of degrees of freedom in the model. As a result sample statistics may be contaminated by sampling error. A type of methodology for handling sampling error is commonly referred to as 'covariance localization' in the EnKF literature. Given this equivalence, the algorithms described in Section 3.2 will be referred to as 'ETKF-type' methodologies. The importance of the ETKF AO method is reflected by the fact that it is currently used to determine flight paths of airplanes equipped with global positioning system dropwindsondes in the Winter Storm Reconnaissance Program (WSR) at NCEP. Results from the WSR demonstrate that the AOs generally improve the forecasts of targeted weather events (Szunyogh et al., 2002). Majumdar et al. (2001, 2002a,b) describe a series of studies exploring the operational use of the ETKF. Given the operational importance of the ETKF, developing a deeper understanding of such algorithms is well motivated.

In this paper, a series of observing system simulation experiments (OSSEs) in a low-order dynamical system are used to address two broad questions: What problems might be associated with the use of ETKF-type schemes? What is the statistical/physical nature of any problems and how can they be overcome? More specifically, the following questions are addressed: How is the performance of ETKF-type schemes affected by the ensemble size? How do the impacts of ensemble size vary with the inherent time scale ($t_v - t_i$) of the problem? The OSSEs are performed in a Lorenz 1996 model (hereafter L96) (Lorenz, 1995). While the L96 model equations cannot be derived from some truncation the geophysical fluid equations, they have characteristics which are not unlike atmospheric models (Lorenz and Emanuel, 1998). The L96 model is composed of non-linear advection terms, linear dissipation terms and a constant forcing. The L96 exhibits sensitive dependence to initial conditions and small disturbances propagate eastwards. The L96 model can therefore be viewed as a reasonable low-order testing ground for atmospheric prediction experiments. This is reflected by the fact that it has been used in several investigations of AOs (Lorenz and Emanuel, 1998; Berliner et al., 1999; Hansen and Smith, 2000; Trevisan and Uboldi, 2004). The use of a low-order model affords a statistically significant examination of the problem (impractical in a GCM due to computational expense). The DA is done using the ensemble adjustment Kalman filter (EAKF) developed by Anderson (2001, 2003). For the L96 experiments in this paper, it is clearly impossible to mimic the ratio of ensemble size, to state space size encountered in realistic applications of EnKF. In this paper, the emphasis is placed on results for ensemble sizes which have been chosen to ensure that sampling error problems (among others) of similar nature to those encountered in realistic applications must be dealt with. Indeed, as will be demonstrated in this paper, proper handling of sampling errors through a space- and time-dependent covariance localization is critical to realizing the full benefits of ETKF-type AO algorithms. There are two key issues in AOs, which are beyond the scope of this paper. The first issue involves finding *efficient* strategies for locating the optimal AO locations suggested by a given method. The second issue involves coping/dealing with model errors when using ensemble-based AO algorithms. It is hoped that such issues will be adequately addressed by future workers in atmospheric data assimilation.

Section 2 provides a statement of the AO problem to be examined in this study. Section 3 describes the theory behind the AO algorithms studied in this paper applied to the problem of Section 2. Section 4 describes the OSSEs and how the data are collected. Section 5 shows the results of the OSSEs and provides a discussion. Section 6 provides a summary and conclusions.

2. The adaptive observations problem examined in this study

2.1. The model

Let the time evolution of the true discrete system state (or atmospheric state) be given by some finite difference solution to,

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = M(\mathbf{x}_t, t), \tag{1}$$

where $\mathbf{x}_t$ is an n-dimensional state at time t and M is an n-dimensional vector function. For the purposes of this study, the same finite difference version of eq. (1) also serves as the forecast model.

2.2. Problem definition

The formal definition provided here will be necessary to understand the theoretical developments of Section 3. An attempt to follow the notation suggested by B01 for the AOs problem has been made throughout this paper wherever possible. The notation of B01 is a suggested extension of the standard notation in Ide et al. (1997) to the AOs problem. Again, let the present time be t_i. In general, the observational network at a given time will consist of both routine and adaptive components. The routine component of the network corresponds to observations whose locations are *not* varied at will (e.g. a network of radiosonde observations). Let the routine component of the network at t_a be summarized by a possibly non-linear operator $H_{t_a}^{\text{routine}}$, which maps the n-dimensional true state $\mathbf{x}_{t_a}$ onto the expected routine observations. The adaptive component of the network corresponds to observations whose locations *can* be varied at will. Let a particular hypothetical spatial configuration of the adaptive component be denoted by $H_{t_a}^{*,\text{adaptive}}$ (possibly non-linear). The superscript '$*$' denotes a particular spatial configuration of the adaptive component. Let the combination of the routine and adaptive component be given by $H_{t_a}^{*}$. The superscript '$*$' on the observation operator $H_{t_a}^{*}$ emphasizes that its adaptive component has some hypothetical configuration. Assume that for every configuration of $H_{t_a}^{*}$, the statistical characteristics of the instrument errors are known.

The goal of the AOs problem can now be stated: At time t_i, one wants to determine the spatial configuration of the adaptive component of the network at $t_a > t_i$, which minimizes the forecast uncertainty $\Phi_{t_v}(H_{t_a}^{*})$ for $t_v > t_a$, for a forecast that will be initialized at t_a. For simplicity and clarity, a particularly simple version of the AOs problem has been chosen as a focus. The AOs problem can be stated with great generality to include multiple adaptive observation (among other generalities such as multiple verification times, see B01). The implications of the numerical results in this paper for the multiple AO time problem will be discussed in Section 5.

3. Theory: Ensemble filter-based methodologies

3.1. Preliminaries

Various ensemble filter (EF)-based AO methodologies, applied to the problem of Section 2, will now be discussed. At the initial time t_i, assume that a posterior ensemble of model states, $[\mathbf{x}_{t_i,k}^{u}]$, are available (superscript 'u' denotes updated or posterior, k indicates ensemble member where $k = 1, \ldots, K$, K is the ensemble size, the brackets [] are used to denote a collection of K ensemble members). The ensemble $[\mathbf{x}_{t_i,k}^{u}]$ is thought of as output from an EF DA algorithm (in principle, these methods can be applied to systems with other ensemble-generation methods). The posterior ensemble has been updated with all available observations at t_i. Given $[\mathbf{x}_{t_i,k}^{u}]$ and the forecast model (1), methodologies for estimating $\Phi_{t_v}(H_{t_a}^{*})$ will be discussed in Sections 3.2 and 3.3.

The formulation of $\Phi_{t_v}(H_{t_a}^{*})$ depends on the choice of norm. A mean squared error (MSE) norm is used in this paper. Let $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ denote a forecast of the covariance of the state at t_v *given* $H_{t_a}^{*}$ at t_a. The superscript 'f' on $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ denotes that it is a forecast quantity computed at t_i. Using this notation, $\mathbf{P}^{f}(t_v)$ (*not* conditioned on $H_{t_a}^{*}$) is the sample covariance for t_v for the forecast ensemble valid for t_v denoted by $[\mathbf{x}_{t_v,k}^{f}]$ ($[\mathbf{x}_{t_v,k}^{f}]$ is $[\mathbf{x}_{t_i,k}^{u}]$ evolved under (1) out to t_v). In this study, the forecast uncertainty is $\Phi_{t_v}(H_{t_a}^{*}) = \text{Tr}(\mathbf{P}^{f}(t_v|H_{t_a}^{*}))$ (which can be localized for localized verification regions). The goal of the problem is to determine the configuration of $H_{t_a}^{*}$ that minimizes $\Phi_{t_v}(H_{t_a}^{*}) = \text{Tr}(\mathbf{P}^{f}(t_v|H_{t_a}^{*}))$. The distinguishing characteristic of the various AO methods discussed in this Section is how $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ is computed. We stress the difference of computing $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ for the AO problem as opposed to ordinary data assimilation. The calculation of $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ is assumed to take place at t_i, as opposed to ordinary data assimilation that computes covariances at times up to and including the last available data. In general, $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ will depend on specified values of observations at the future time t_a, in contrast to ordinary data assimilation that uses observation values obtained from a realization of a true random process.

In Sections 3.2, the non-linear ensemble mean method is described. In Section 3.3, a Bayesian framework for ETKF-type methods is discussed. In Section 3.4, the equivalence between the non-linear ensemble mean method and ETKF-type methods under certain circumstances is discussed. All the developments in Sections 3.2, 3.3 and 3.4 can be easily extended to the case of multiple AO times (not shown in this paper for brevity and clarity). The extension to the multiple AO time case is discussed in Khare (2004).

3.2. The non-linear ensemble mean method (NONLIN-NOLOC, NONLIN-LOC)

To achieve our goal of understanding computationally efficient ETKF-type AO schemes, it will be helpful to make comparisons to an AO scheme that is equivalent to a sequential in-time filtering process for non-linear dynamics. This scheme, the non-linear ensemble mean method, computes $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$ in the following way. The posterior ensemble at t_i is evolved under the full nonlinear model (1) to generate $[\mathbf{x}_{t_a,k}^{f}]$. Let the observation values at the future time t_a be $\mathbf{y}_{t_a}^{o\,*} = [H_{t_a}^{*}(\overline{\mathbf{x}}_{t_a}^{f})]$ (the motivation for doing so will be discussed in Section 3.4). Given $\mathbf{y}_{t_a}^{o\,*}$, the prior ensemble $[\mathbf{x}_{t_a,k}^{f}]$ is then updated using the EF algorithm, which is then integrated out to t_v using (1). The ensemble valid for t_v is then used to compute $\mathbf{P}^{f}(t_v|H_{t_a}^{*})$.

For the remainder of this paper the terminology 'covariance localization' (hereafter CL) will be used frequently. As mentioned in Section 1, when applying EnKF methods to realistic prediction problems, computationally feasible ensemble sizes

are typically far less than the number of degrees of freedom in the prediction model. This can lead to prior estimates of covariance which are contaminated by sampling error, which can lead to observations having large impacts on physically unrelated state variables. To reduce this unwanted impact, a common approach is to pre-multiply prior covariances by some function (valued between 0 and 1) which falls off with physical distance (Hamill et al., 2001; Houtekamer and Mitchell, 2001). In Section 5, this non-linear ensemble mean method is implemented using an EAKF. The scheme that does not use localization in updating $[\mathbf{x}^f_{t_a,k}]$ is called NONLIN-NOLOC. The scheme which does use localization in updating $[\mathbf{x}^f_{t_a,k}]$ will be called NONLIN-LOC. The forecast sample covariance for t_a is given by $\mathbf{P}^f(t_a)$ (it is not necessary to compute all the elements of this matrix in the actual implementation used in this paper). To formalize the notion of CL, the notation introduced by Houtekamer and Mitchell (2001) is followed. For the NONLIN-LOC method, localization amounts to a Schur product of $\mathbf{P}^f(t_a)$ with an $n \times n$ matrix ρ denoted by $\rho \circ \mathbf{P}^f(t_a)$. Element $\rho_{l,m}$ $(l, m = 1, \ldots, n)$ is obtained from a correlation function with compact support which is a function of the distance between grid points x_l and x_m. The Schur product $\rho \circ \mathbf{P}^f(t_a)$ is an $n \times n$ matrix where $\rho \circ \mathbf{P}^f(t_a)_{l,m} = \rho_{l,m}\mathbf{P}^f(t_a)_{l,m}$. For the NONLIN-LOC method, the prior/forecast covariance at t_a is taken to be $\rho \circ \mathbf{P}^f(t_a)$. The details of how localization is actually implemented for the results in this paper is left to Section 5. Unlike the ETKF-type methods, the non-linear ensemble mean methods require repeated integrations of (1) to test various configurations of $H^*_{t_a}$. As a result, the non-linear ensemble mean method may be computationally too expensive to apply operationally.

3.3. A Bayesian framework for ETKF-type methods

In what follows, algorithms for estimating $\mathbf{P}^f(t_v|H^*_{t_a})$ which only require *one* integration of the ensemble $[\mathbf{x}^u_{t_i,k}]$ from t_i to t_v will be discussed. These ETKF-type methods are suitable for real-time applications in high-dimensional systems due to their computational efficiency.

3.3.1. Formulating $\mathbf{P}^f(t_v|H^*_{t_a})$ using Bayes rule. Assume that at t_i, $[\mathbf{x}^u_{t_i,k}]$ has been integrated out to t_v using the prediction model (1). The forecast ensembles for t_a and t_v are denoted $[\mathbf{x}^f_{t_a,k}]$ and $[\mathbf{x}^f_{t_v,k}]$, respectively. Define an augmented state $\mathbf{z} = [\mathbf{x}^T_{t_a}, \mathbf{x}^T_{t_v}]^T$ (2n-dimensional). The forecast ensemble members are taken to represent a sample of a prior pdf $\mathbf{p}(\mathbf{z}) = \mathbf{p}(\mathbf{x}_{t_a}, \mathbf{x}_{t_v})$. Bayes rule (Cohn, 1997) can be used to update the prior pdf given the values of observations at the future AO time t_a corresponding to the trial configuration $H^*_{t_a}$ (the observations values are denoted by $\mathbf{y}^{o\,*}_{t_a}$),

$$\mathbf{p}\left(\mathbf{x}_{t_a}, \mathbf{x}_{t_v} \middle| \mathbf{y}^{o\,*}_{t_a}\right) = \frac{\mathbf{p}\left(\mathbf{x}_{t_a}, \mathbf{x}_{t_v}\right)\mathbf{p}\left(\mathbf{y}^{o\,*}_{t_a} \middle| \mathbf{x}_{t_a}\right)}{\text{Normalization}}. \tag{2}$$

At the initial time t_i, the observation values at the future time t_a, of course, cannot be known. To proceed, some values for the observations must be specified in a manner consistent with $H^*_{t_a}$. Let the total number of hypothetical observations at time t_a (corresponding to $H^*_{t_a}$) be given by $[\mathbf{x}^f_{t_v,k}]$. Let the $[\mathbf{x}^f_{t_v,k}]$ vector of observations be given by $\mathbf{y}^{o\,*}_{t_a} = [H^*_{t_a}(\overline{\mathbf{x}}^f_{t_a})]$, where $\overline{\mathbf{x}}^f_{t_a}$ denotes the mean of $[\mathbf{x}^f_{t_a,k}]$.

To proceed, the distributions in (2) must be specified. Assume that $\mathbf{p}(\mathbf{y}^{o\,*}_{t_a}|\mathbf{x}_{t_a})$ is Gaussian with known $p^*_{t_a} \times p^*_{t_a}$ covariance $\mathbf{R}^*_{t_a}$. For the purposes of the update, let $\mathbf{p}(\mathbf{x}_{t_a}, \mathbf{x}_{t_v})$ be represented by a Gaussian. The covariance for the joint state $\mathbf{z}$ in (2) is then given by the Kalman filter (KF) updated (posterior) covariance equation (Cohn, 1997). Using an EnKF update methodology, the prior ensemble (the sample of $\mathbf{p}(\mathbf{z})$) can be transformed to give an updated ensemble that is consistent with the Kalman filter (KF) updated covariance equation. When applying the EnKF to this problem, the prior sample covariance of $\mathbf{p}(\mathbf{x}_{t_a}, \mathbf{x}_{t_v})$ is computed from the forecast ensembles $[\mathbf{x}^f_{t_a,k}]$ and $[\mathbf{x}^f_{t_v,k}]$. For the types of EnKFs under consideration, the normalization in (2) is not required. Note that the KF updated covariance equation is, formally, independent of the specified observation values given in $\mathbf{y}^{o\,*}_{t_a}$ (Cohn, 1997). The updated ensemble for the joint state $\mathbf{z}$ can then be used to compute $\mathbf{P}^f(t_v|H^*_{t_a})$. Note that many different configurations of $H^*_{t_a}$ can be tested without having to repeatedly integrate the model equations. A wide variety of EnKFs can be used to implement this strategy. This includes both perturbed observation EnKFs (Evensen, 1994; Burgers et al., 1998) and Deterministic Ensemble Square Root Filters (DEnSRFs) (Tippett et al., 2003) (hereafter T03). In this paper, the focus is on implementations using DEnSRFs.

3.3.2. Computing $\mathbf{P}^f(t_v|H^*_{t_a})$ using a DEnSRF without covariance localization (ETKF/NOLOC). Next, the implementation of the method discussed in Section 3.3.1 will be shown using a DEnSRF for the case where CL *is not* used in the update. The key point is that the expression developed for $\mathbf{P}^f(t_v|H^*_{t_a})$ is *equivalent* to what is obtained using the ETKF AO method introduced by B01. When CL *is* used in the update, the method is no longer equivalent to the ETKF AO method of B01 (new methods that make use of CL are discussed in Section 3.3.3). Using the forecast ensembles, $[\mathbf{x}^f_{t_a,k}]$ and $[\mathbf{x}^f_{t_v,k}]$, a sample prior covariance matrix for the augmented state, $\mathbf{z} = [\mathbf{x}_{t_a}{}^T, \mathbf{x}_{t_v}{}^T]^T$ can be computed. This is given by the outer product of the $2n \times K$ matrix of perturbations $\mathbf{Z}^f$, whose k^{th} column is given by $[\mathbf{z}^f_k - \overline{\mathbf{z}}^f]/\sqrt{K-1}$, where $[\mathbf{z}^f_k] = [\mathbf{x}^f_{t_a,k}{}^T, \mathbf{x}^f_{t_v,k}{}^T]^T$ and $\overline{\mathbf{z}}^f$ is the mean of $[\mathbf{z}^f_k]$. Thus, the *input or prior information* is the $2n \times 2n$ sample prior covariance matrix for the augmented state given by,

$$\mathbf{P}^f_{aug}(t_a, t_v) = \mathbf{Z}^f\mathbf{Z}^{f\,T}. \tag{3}$$

The subscript *aug* is used to emphasize that it is for the augmented state. Note that in the actual implementation it is not necessary to compute all the elements of the covariance matrix. $\mathbf{P}^f_{aug}(t_a, t_v)$ can now be conditioned on the trial network $H^*_{t_a}$. $H^*_{t_a}$ operates on the state at t_a. Working in the augmented state, one

must define the observation operator for the augmented state. Define H^* such that $H^*(\mathbf{z}) = H^*_{t_a}(\mathbf{x}_{t_a})$. Using a DEnSRF, the updated covariance matrix can be written as $\mathbf{P}^f_{aug}(t_a, t_v|H^*_{t_a}) = \mathbf{Z}^f \mathbf{S}^*_{t_a} \mathbf{S}^{*\,T}_{t_a} \mathbf{Z}^{f\,T}$. Formally, for a given DEnSRF, the $K \times K$ transformation matrix $\mathbf{S}^*_{t_a}$ is obtained by solving, $\mathbf{S}^*_{t_a} \mathbf{S}^{*\,T}_{t_a} = (\mathbf{I} - \mathbf{V}_{t_a} \mathbf{D}^{-1}_{t_a} \mathbf{V}_{t_a}{}^T)$, where $\mathbf{V}_{t_a} = (H^*(\mathbf{Z}^f))^T = (H^*_{t_a}(\mathbf{X}^f_{t_a}))^T$ and $\mathbf{D}_{t_a} = \mathbf{V}_{t_a}{}^T \mathbf{V}_{t_a} + \mathbf{R}^*_{t_a}$. $\mathbf{X}^f_{t_a}$ is the matrix of perturbations for time t_a generated from $[\mathbf{x}^f_{t_a,k}]$. This is an application of the formalism developed in T03 to this problem. The different versions of DEnSRFs discussed in T03 solve for different versions of $\mathbf{S}^*_{t_a}$. However, the different versions yield numerically identical values for $\mathbf{P}^f_{aug}(t_a, t_v|H^*_{t_a})$. Finally, the desired *output or posterior information*, the $n \times n$ matrix $\mathbf{P}^f(t_v|H^*_{t_a})$ can be formed,

$$\mathbf{P}^f\left(t_v\middle|H^*_{t_a}\right) = \left(\mathbf{Z}^f \mathbf{S}^*_{t_a} \mathbf{S}^{*\,T}_{t_a} \mathbf{Z}^{f\,T}\right)_{n\times n} = \mathbf{X}^f_{t_v} \mathbf{S}^*_{t_a} \mathbf{S}^{*\,T}_{t_a} \mathbf{X}^{f\,T}_{t_v}. \quad (4)$$

The subscript $n \times n$ in (4) indicates the $n \times n$ submatrix consisting of rows $n+1 \to 2n$ from columns $n+1 \to 2n$. Applying the formalism of B01 to the same problem yields an *equivalent* expression for $\mathbf{P}^f(t_v|H^*_{t_a})$. Therefore, when using a DEnSRF to implement the method described in Section 3.3.1 without covariance localization, this method is equivalent to the ETKF AO strategy of B01. For the numerical results of Section 5, the above methodology will be implemented using the EAKF (Anderson 2001, 2003), which is a particular version of a DEnSRF (T03). For the EAKF implementation with no localization, this AO method will hereafter be called ETKF/NOLOC. The name, ETKF/NOLOC. recognizes that no localization is being used and the method is equivalent to the ETKF AO strategy originally derived by B01.

3.3.3. New ETKF-type methods making use of space-time covariance localization (ETKF-SPACE, ETKF-SPACETIME). In applications of interest, a limited number of ensemble members are available, and the prediction models have many degrees of freedom. As a result, it is important to understand the impacts of sampling error when implementing the strategy of Section 3.3.1 with a DEnSRF. Here, we introduce two new strategies that do make use of CL when implementing the strategy in Section 3.3.1 These methods, which are called ETKF-SPACE and ETKF-SPACETIME, are therefore *not* equivalent to the ETKF AO strategy of B01.

The approach to CL that will be discussed in this paper views the computation of $\mathbf{P}^f(t_v|H^*_{t_a})$ as an application of an EnKF in the augmented state space $\mathbf{z} = [\mathbf{x}_{t_a}{}^T, \mathbf{x}_{t_v}{}^T]^T$. When applying the EnKF to the augmented state space, a CL can be achieved by applying the same type of distance-dependent functions used by the EF DA algorithm. Formally, this involves setting the prior/forecast covariance matrix to $\tilde{\rho} \circ \mathbf{P}^f_{aug}(t_a, t_v)$ (Schur product of eq. 3 with $\tilde{\rho}$ as defined in Section 3.2). Here, $\tilde{\rho}$ is now a $2n \times 2n$ matrix obtained from a correlation function with compact support, which is a function of 'distance' between state variables z_l and z_m. When applying distance-dependent CL functions in the augmented state space, it is crucial to recognize that distance between state variables at *different* times is not a well-defined quantity. Intuitively, proper definition of distance in the augmented state space will depend on the dynamics (1).

In Section 5, numerical results will be shown for two new ETKF-type methods that make use of CL. The first such method is called ETKF-SPACETIME, which is an implementation of the AO of Section 3.3.1. using an EAKF with a CL that uses a characterization of the dynamics in defining distances between state variables valid at different times. The second method is called ETKF-SPACE, which is an implementation of the AO of Section 3.3.1. using an EAKF with a CL that does *not* use a characterization of the dynamics in defining distances between state variables valid at different times. The details of the CL implementation are left to Section 5.

3.4. *Equivalence between the non-linear ensemble mean and ETKF-type methods*

Assume that the model dynamics are linear. For any finite K member ensemble, the expression for the verification time covariance obtained via the ETKF/NOLOC can be re-written as $\mathbf{P}^f(t_v|\mathrm{H}^*_{t_a}) = \mathbf{M}_{t_a,t_v}$, $\mathbf{M}_{t_i,t_a} \mathbf{X}_{t_i} \mathbf{S}^*_{t_a} \mathbf{S}^{*\,T}_{t_a} \mathbf{X}_{t_i}{}^T \mathbf{M}^T_{t_i,t_a}$ $\mathbf{M}^T_{t_a,t_v}$, where $\mathbf{M}_{t_i,t_a}$ is the linear dynamics propagator from t_i to t_a (same for t_a to t_v, also if $t_v - t_a = t_a - t_i$ then $\mathbf{M}_{t_i,t_a} = \mathbf{M}_{t_a,t_v}$). This is identical to what is obtained by applying the NONLIN-NOLOC algorithm to the linear dynamics case where the specified observations values at t_a are $\mathbf{y}^{o\,*}_{t_a} = [H^*_{t_a}(\overline{\mathbf{x}}^f_{t_a})]$. This equivalence motivates the choice of specified observation values for the non-linear ensemble mean methods. This equivalence was originally derived in B01. It is in this sense that the ETKF/NOLOC is equivalent to a sequential in-time filtering process for linear dynamics with no CL. The non-linear ensemble mean and ETKF-type methods may not be equivalent when implemented with CL. The reason is that there is no simple way of relating a CL for covariance between simultaneous variables in the NONLIN to a CL for covariance between variables at different time levels in the ETKF method.

How do the differences between the ETKF/NOLOC and NONLIN-NOLOC arise for non-linear dynamics? For non-linear dynamics, the ensemble perturbations in (4) (i.e. $\mathbf{X}^f_{t_v}$) are obtained via integration of the full non-linear model equations from t_i to t_v. The ensemble perturbations in (4) have been integrated from t_a to t_v without being influenced by the observations as they would in the non-linear ensemble mean implementations. As a result, the non-linear ensemble mean and ETKF-type methods cannot, in general, be equivalent for non-linear dynamics.

4. Description of the observing system simulation experiments

4.1. Model description, experimental design and data collection

The L96 model equations are given by, $dx_i/dt = -x_{i-1}(x_{i-2} - x_{i+1}) - x_i + F$, where $i = 1, \ldots, 40$ and the forcing parameter $F = 8$ (this subscript i is not to be confused with the subscript on the initial time t_i). The model is cyclic with $x_{i+40} = x_i$. The locations of the state variables can be thought of as equidistant locations on a latitude circle. The distance between x_i and x_{i+1} is defined as one zone. Similar to atmospheric systems, L96 has (i) 'energy' conservation (ii) nonlinear advection and linear dissipation (iii) sensitivity to initial conditions and (iv) external forcing. The energy $E = \sum_{i=1}^{40} x_i^2/2$ is conserved through a compensation between the constant forcing term and linear dissipation term. For $F = 8$, disturbances propagate from low-to-high indices ('west' to 'east') (Lorenz and Emanuel, 1998) (hereafter L98). Following LE98, a fourth-order Runge–Kutta scheme with a time stepping of $0.05 = \triangle t$ time units is used. One 'lead time unit' is defined as $0.05 = \triangle t$. Numerical experiments yield an error doubling time of roughly $8\ \triangle t$ (Lorenz, 1995). The times between assimilation times is set to 0.05 or roughly 1/8 the doubling time, chosen to mimic current operational settings (Bishop et al., 2003). Assuming that the doubling time in the atmosphere is roughly 2 d, one lead time unit $\triangle t$ can be thought of as equivalent to 6 h. Analysis of linear perturbations about a steady-state solution gives a group velocity of the most unstable wavenumber (8) of roughly $+1/2$ zone per $\triangle t$ (LE98). A particle moving at the group velocity of the most unstable wavenumber will travel half the distance between model grid points in time $\triangle t$. Wavenumber 8 dominates the power spectrum (LE98). The model climatology is $\overline{x_i} = 2.3$ with $\sigma_{\text{climate},i} = 3.6$ (same for all i) (LE98). The observational standard deviation is set to $\sigma_{\text{obs}} = 0.2$, giving accurate observations compared to σ_{climate}. These choices of F and σ_{obs} match LE98.

For simplicity and clarity, a particularly simple version of the problem described in Section 2 is analyzed in this paper. The problem is to determine (at t_i) the optimal location of exactly *one* AO at t_a to improve the forecast error for x_1 at t_v. *No* routine observations are assimilated at $t_a (H_{t_a}^{\text{routine}}$ is empty). The chosen problem allows for clean comparisons of the various methods to be tested (see Section 3.4). The differences between the initial (t_i), adaptive (t_a) and verification (t_v) times is a matter of choice. Operationally these are constrained by practical considerations. For simplicity, we choose to set $t_v - t_a = t_a - t_i = lead \triangle t$, where *lead* is an integer representing the number of lead time units.

Next, the collection of statistics using OSSEs is described. Assume that an EF is being used for DA. *For a given AO scheme and lead the following procedure is used*: An initial true state $\mathbf{x}^t$ is sampled from climatology. The state $\mathbf{x}^t$ is integrated 10^4 time steps. An ensemble is then obtained by perturbing around $\mathbf{x}^t$ to obtain an initial ensemble $[\mathbf{x}_k]$. The ensemble and $\mathbf{x}^t$ are then integrated for 10^4 more time steps up to $t_{i,1}$ (the second subscript indicates the first forecast cycle). Assuming that L96 is ergodic, the ensemble and $\mathbf{x}^t$ can be considered random samples of climatology. The OSSEs can now begin. At $t_{i,1}$, a decision regarding where to place the one AO at $t_{a,1}$ is made. The AO can be taken at any of the 40 grid points (the chosen location is the one that minimizes the forecast variance for x_1 at $t_{v,1}$, according to the AO scheme being tested). Since there are no routine observations, only the AO is assimilated at $t_{a,1}$. The ensemble and truth for time $t_{i,1}$ is then integrated out to $t_{a,1}$. The observation is simulated by adding a random error to the truth at the observation location specified by the AO scheme. The random error is a sample of a normal distribution with standard deviation σ_{obs} and mean equal to the value of the truth at the observed location. The prior ensemble for $t_{a,1}$ (denoted $[\mathbf{x}^f_{t_{a,1},k}]$) is then updated using the EF to generate $[\mathbf{x}^u_{t_{a,1},k}]$. $[\mathbf{x}^u_{t_{a,1},k}]$ is then integrated out to the verification time ($t_{v,1}$) along with $\mathbf{x}^t$. The absolute value of the difference between the ensemble mean and truth at $t_{v,1}$ for state variable x_1 is taken as the measurement of forecast error. The absolute difference between the ensemble mean and truth is commonly used to evaluate forecast/analysis errors in EnKF DA studies (Houtekamer and Mitchell, 1998, 2001; Anderson, 2001; Whitaker and Hamill, 2002; Snyder and Zhang, 2003). The procedure is repeated cyclically, where the output of the last OSSE is taken as the input to the next OSSE. For OSSE number $j + 1$, the new initial time is given by $t_{i,j+1} = t_{a,j}$. The analysis ensemble for $t_{i,j+1}$ is given by $[\mathbf{x}^u_{t_{a,j},k}]$. The above procedure is repeated $J = 10\,200$ times where $j = 1, \ldots, J$. The first 200 error statistics are discarded to ensure that statistics are only collected when the model/assimilation system is in statistical equilibrium (200 deemed sufficient via empirical testing). For a given lead time and AO scheme, the mean absolute forecast error (over the 10 000 cases) for x_1 is taken as the measure of performance.

4.1.1. Details regarding the data assimilation. The EF DA scheme used in the experiments is the EAKF developed by Anderson (2001, 2003). This choice of EF helps to achieve consistency between the DA scheme and the methodology used to select locations of AOs, an important consideration in the design of AO networks (Majumdar et al., 2002a). The majority of the discussion in Section 5 will focus on the case where the ensemble size is $K = 20$. When generating results for $K = 20$, heuristic adjustments to the EAKF implementation were required to achieve stable assimilations, namely CL and covariance inflation (CI). Here, the CL works in the following way. Suppose, at a given time, the EAKF is updating state variable x_α with an observation of x_β. Suppose that x_α is physically unrelated to x_β due to the physical distance between x_α and x_β. For smaller ensemble sizes, the observation of x_β may be strongly correlated to x_α at some instances. Consequently, updating the ensemble of state variable x_α with an observation x_β may degrade the quality

of the analysis (Hamill et al., 2001). To alleviate this problem, covariances between the prior state variables and observation variables are multiplied by a correlation function with local support (originally proposed/tested by Houtekamer and Mitchell, 1998). The correlation function is a fifth-order piecewise rational function with compact support characterized by a parameter c that is the half-width of the correlation function (Gaspari and Cohn, 1999). For the L96 system, the total distance of one cycle from grid point x_i to x_i is defined to be one unit. For $c = 0.3$, observations influence state variables which are up to 2×0.3 units away from the observation location in both the east and west direction. In this setup, a given state variable can be at most be 0.5 units distance away from an observation. This choice of half-width parameter, therefore, implies that observations have global impact. Note that for $c < 0.25$ observations would not have global impact. CI is used to circumvent filter divergence (Jazwinski, 1970). CI increases the uncertainty in the prior covariance estimates so that the filtering results do not diverge from the observations (Anderson and Anderson, 1999). For the $K = 20$ results, CI of $\gamma = \sqrt{1.04}$ has been used for all the experiments. For prior state variable x_α, covariance inflation is implemented by setting the k^{th} ensemble member to $x_{\alpha,k} = \gamma(x_{\alpha,k} - \overline{x_\alpha}) + \overline{x_\alpha}$ where $\overline{x_\alpha}$ is the prior ensemble mean. The fixed values of c and γ were chosen to yield stable assimilations (no filter divergence) while still maintaining small prior covariances. It was found that small alterations to the prescribed values of c and γ do not change the qualitative conclusions of the results for $K = 20$.

In present day applications of EnKF to realistic prediction problems, the number of ensemble members that can be run is typically far less than the degrees of freedom in the prediction model (due to computational expense). As a result, heuristic adjustments of similar nature to those discussed here are required for application of the EnKF to realistic problems. This is the reason for placing the emphasis on $K = 20$ in this paper. To demonstrate certain principles, results will also be shown for $K = 128$ (much larger than the model size). For the $K = 128$ results, CL or CI *was not* required to achieve stable assimilations and, therefore, no CL or CI was used.

4.2. Summary of the methods to be examined and experimental parameters

For all OSSE results with $K = 20$ both CL and CI are used in the actual data assimilation. For all results with $K = 128$ CL and CI are not used in the actual data assimilation. The performance of various AO algorithms will be examined for the remainder of this paper. The methods that make use of CL when selecting the location of the AO are ETKF-SPACE, ETKF-SPACETIME and NONLIN-LOC. The methods that do not make use of CL when selecting the AO locations are ETKF/NOLOC and NONLIN-NOLOC.

5. Observing system simulation experiments: Results and discussion

5.1. The non-linear ensemble mean scheme

The objective of the numerical experiments is to develop an understanding of the various computationally efficient ETKF-type AO schemes. It will be helpful to first consider results for the non-linear ensemble mean schemes. Recall that the non-linear ensemble mean schemes require repeated integrations of the model equations in testing various hypothetical observing networks for t_a, and, therefore, may not be suitable for operational applications. Results for a random strategy (RANDOM) will also be shown. The RANDOM strategy selects the observed grid point at t_a randomly. The RANDOM strategy is useful as a measure of a reasonable lower bound in the sense that a given AO scheme is deemed to be of no use if it does not outperform the RANDOM scheme.

We begin by presenting a detailed analysis of the NONLIN-NOLOC method's performance (which does not use CL when updating the forecast ensemble at t_a). Such a detailed analysis is presented so that the impacts of sampling errors and the performance of the scheme with localization (NONLIN-LOC) can be understood. Fig. 1 depicts the mean absolute forecast errors in the verification region x_1 over the 10^4 experiments for the NONLIN-NOLOC and RANDOM schemes (among others, note that the ensemble size is $K = 20$). Note that Fig. 1 depicts results from 8 independent experiments for each scheme (one for each *lead* $\triangle t = t_v - t_a = t_a - t_i$). The mean absolute forecast errors are joined by a straight line for aesthetic purposes only. To set yet another benchmark for comparison, results for a FIXED strategy are also shown in Fig. 1. For each lead time, the fixed observation location that yielded the smallest mean absolute forecast errors over the entire 10^4 experiments was determined. For lead times 1 through 4, the optimal fixed location was grid point 1. For lead times 5 and 6, location 40 was optimal. Locations 39 and 38 were optimal for lead times 7 and 8, respectively. The values of forecast errors in Fig. 1 are reasonable given $\sigma_{\text{climate}} = 3.6$. Results are shown up to *lead* $= 8$ since all the AO schemes tested in this study did not outperform the RANDOM scheme for longer *lead* experiments. Assuming that $\triangle t$ represents 6 h, the experiments cover *lead* $\triangle t$ from 6 h to 2 d. For all the mean absolute forecast errors in Fig. 1, standard errors of the mean have been computed. The standard errors of the mean could have been used to put error bars of Fig. 1. However, given the large number of experiments over which averages have been taken, the size of the error bars are of the order of the thickness of the lines drawn in Fig. 1 and will not be shown. The magnitudes for the standard errors of the mean imply that the results of Fig. 1 are statistically significant.

Results were also obtained for the ensemble spread (ES) technique (LE98; Morss et al., 2001) (not shown on Fig. 1). For a given t_i, the sample forecast standard deviation for all model grid points for t_a can be computed. The ES technique directs the

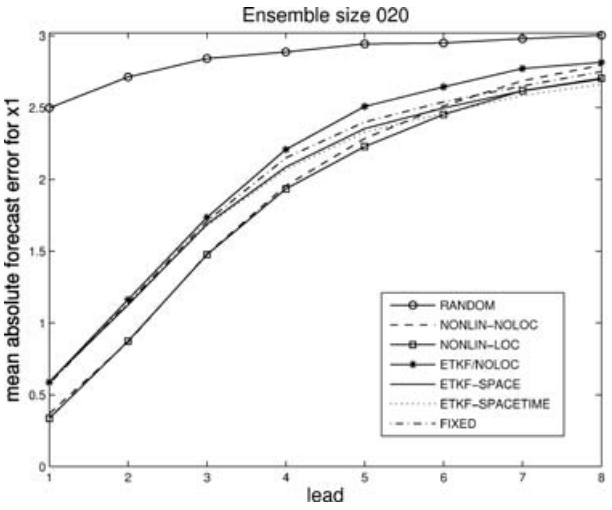

Fig. 1. Results for eight independent experiments where the lead time $lead\triangle t = t_v - t_a = t_a - t_i$ has been varied from $lead = 1, \ldots, 8$. For each method, results for the eight independent experiments have been joined with a line for aesthetic purposes only. The y-axis is the mean absolute forecast error in $x_1(t_v)$ averaged over 10^4 consecutive experiments. The errors are computed by taking the absolute difference in the ensemble mean forecast for $x_1(t_v)$ and the value of the control run (truth) for x_1 at t_v. The ensemble forecast is initialized time t_a. Results are shown for the RANDOM scheme, the non-linear ensemble mean scheme with (NONLIN-LOC) and without (NONLIN-LOC) localization, the ETKF/NOLOC, the ETKF-type scheme with spatial (ETKF-SPACE) and space/time (ETKF-SPACETIME) localization and the fixed scheme (FIXED).

observation to the location with the largest ES. For $lead = 1$, the ES roughly yields a 4 percent improvement in forecast errors compared to the RANDOM strategy, and less than 1 percent for the longer lead times.

To develop an understanding of the statistical behaviour of the NONLIN-NOLOC scheme, histograms of the observed location are useful. The left-hand column of Fig. 2 plots the histogram of observed locations for NONLIN-NOLOC $lead = 1, 4, 8$ experiments with $K = 20$. Recall that analysis of linear perturbations about a steady-state solution in L96 reveals that a particle moving at the group velocity of the most unstable wavenumber travels $1/2$ the distance between grid points per unit $\triangle t = 0.05$. Looking at the left-hand column of Fig. 2, the upstream shifts in the peak of the distributions are physically reasonable. However, for the $lead = 4, 8$ results, the AO is sometimes directed to locations far from the verification region (the upper left-hand panel for $lead = 1$ has a few observations directed far from the verification region, although it is hard to make out on the plot). If one assumed that disturbances propagate at the group velocity of the most unstable wavenumber, one could regard this behaviour as 'physically surprising'.

For the $lead = 8$ results in the lower left panel of Fig. 2, the NONLIN-NOLOC scheme observes one of the locations $[x_{12}, \ldots, x_{29}]$ 1863 times out of the 10^4 experiments. The grid points $[x_{12}, \ldots, x_{29}]$ were chosen to represent locations far from the verification region x_1. Given the group velocity of the most unstable wavenumber, one may speculate that these 1863 observations led to very little improvement in the forecast errors. To quantify whether or not the 1863 observations actually improved the forecast errors in $x_1(t_v)$, the following calculations were done. The average forecast error in $x_1(t_v)$ over the 1863 cases was 2.903. If the observation was not assimilated at t_t, the mean forecast error in $x_1(t_v)$ would have been 2.905. Therefore, over the 10^4 experiments, when the observation was directed to one of the grid points $[x_{12}, \ldots, x_{29}]$, little improvement in forecast error was achieved. The lower left panel of Fig. 2 shows that the most frequently observed grid points were x_{38}, x_{39} and x_{40}. For the 1863 cases being discussed, had the observation been directed to x_{38} instead of one of the grid points $[x_{12}, \ldots, x_{29}]$, the mean forecast error in $x_1(t_v)$ over the 1863 cases would have been 2.743. This calculation was done retrospectively (i.e. after the sequence of 10^4 experiments were completed) by re-running the 1863 forecasts using observations of x_{38} instead of one of the grid points $[x_{12}, \ldots, x_{29}]$. Had the observations been directed to x_{39} or x_{40}, the mean forecast errors would have been 2.582 and 2.623, respectively.

The NONLIN-NOLOC ensemble size $K = 20$ results in Fig. 1 and left column Fig. 2, do not make use of CL when deciding where to place the AO. Is the NONLIN-NOLOC scheme directing observations to locations far from the verification region due to sampling errors? In other words, for the forecast ensemble valid at t_a, are sampling errors causing spurious correlations to arise between state variables nearby and far away from the verification region, ultimately directing AOs to locations far from the verification region? This can be determined empirically by running the analogous experiment for a relatively large ensemble size of $K = 128$. For the $K = 128$ case, the analogous histograms are in the middle column of Fig. 2 showing that the AO is rarely directed to far away locations.

The results in the first two columns of Fig. 2 suggest that the so-called 'physically surprising' behaviour of the NONLIN-NOLOC scheme for the $K = 20$ case is due to sampling error. We further test this notion by implementing the non-linear ensemble mean scheme using CL (the NONLIN-LOC scheme). For consistency, the NONLIN-LOC scheme updates the forecast ensemble at t_a ($[\mathbf{x}^f_{t_a,k}]$) using the same CL ($c = 0.3$) as the DA scheme described in Section 4.1.1 The histograms for the NONLIN-LOC results are shown in the right-hand column of Fig. 2. Comparing the left and right columns of Fig. 2 for $lead = 4, 8$ allows one to see the effect of the CL. For the NONLIN-LOC scheme, the CL has the effect of diminishing the expected benefit of far away observations. Fig. 1 shows the results for the NONLIN-LOC scheme. For the longer lead time experiments, there is a statistically significant improvement in the errors (albeit not dramatic). This is consistent with the empirical observation that the sampling error problem is more severe at longer lead times. Note that unlike the NONLIN-NOLOC

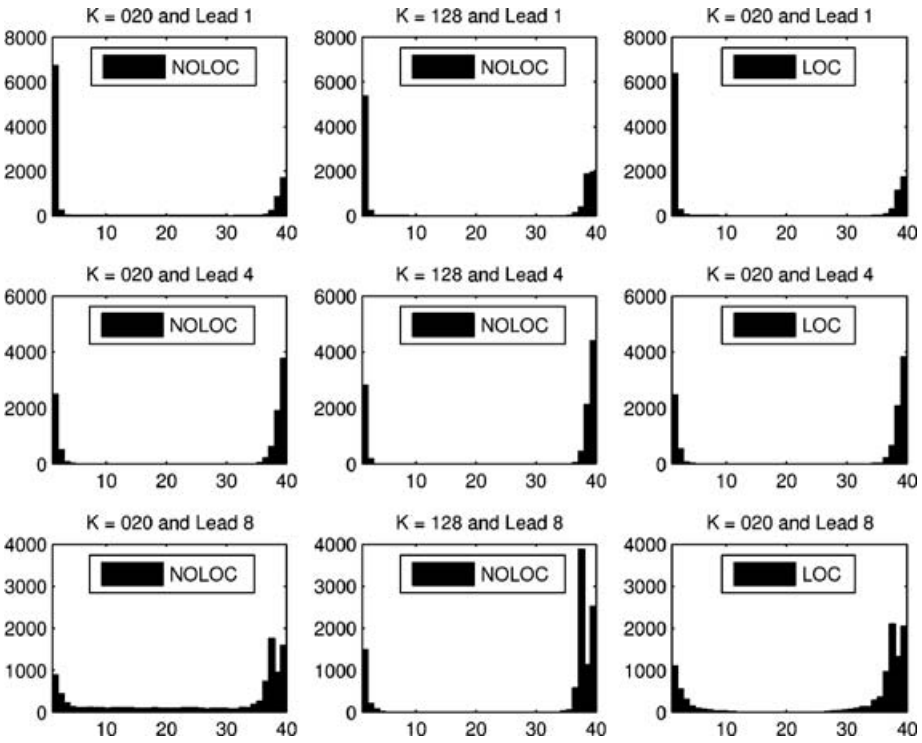

Fig. 2. Histograms of locations selected for the adaptive observation (AO) over the 10^4 experiments. For the first two columns, the NONLIN-NOLOC scheme has been used to select the location of the AO at t_a. The left-hand column shows results from experiments with ensemble size $K = 20$, the middle column is for $K = 128$. The right-hand column shows results for the NONLIN-LOC scheme for the ensemble size $K = 20$ results. For all three methods, results from three independent experiments are shown, where the $lead = 1, 4, 8$ (recall that $lead \triangle t = t_v - t_a = t_a - t_i$).

scheme, the NONLIN-LOC scheme yields smaller mean forecast errors compared to the FIXED results for the lead time 6, 7 and 8 results. The key point is that when working with smaller ensemble sizes, including a CL in the non-linear ensemble mean scheme improves the performance of the algorithm. Notice in the context of the non-linear ensemble mean scheme, *the sampling error that is being handled by the CL is for statistical estimates between state variables valid at the same time* (t_a).

Figure 2 suggests that the impacts of ensemble size for the non-linear ensemble mean scheme are less severe for the shorter *lead* experiments. What is the relationship between the *lead* and the impacts of ensemble size? For the $lead = 1$ results, the dynamics from $t_a \rightarrow t_v$ can be thought of as a linear advection process that nudges uncertainty downstream less than the distance between grid points (given a group velocity of $1/2$ zone per unit *lead*). Since observations are assimilated frequently (every $\triangle t$), we expect locations nearby x_1 to be strongly correlated with x_1 at each t_a. Therefore, observations of state variables nearby x_1 will strongly reduce the uncertainty in the verification region x_1. For $lead = 1$, the dynamics is simply advecting this uncertainty for a very brief span of time ($t_v - t_a = \triangle t$), therefore it is not surprising that locations far from the verification are chosen infrequently. For longer lead times, the dynamics from $t_a \rightarrow t_v$ can no longer be thought of as a simple linear advection process. For the longer *lead* experiments, the frequency of data assimilation cycles is less. Therefore, we expect the absolute correlations (at t_a) of state variables nearby x_1 to be diminished, increasing the likelihood that spuriously correlated state variables far from x_1 are chosen by the NONLIN-NOLOC scheme.

5.2. *The ETKF/NOLOC scheme*

Next, results for the ETKF/NOLOC scheme discussed in Section 3.2.2 will be analysed. Recall that this method is equivalent to the ETKF AO method introduced by B01 and is used operationally in the WSR at NCEP. The goal is to identify any problems when using the ETKF/NOLOC method, and if possible, suggest ways of overcoming them. Figure 1 depicts the forecast errors for the ETKF/NOLOC scheme. The ETKF/NOLOC method is suboptimal compared to the non-linear ensemble mean results. For the $K = 20$ results, the histograms of the observed locations are depicted in the left-hand column of Fig. 3. Compared to the NONLIN-NOLOC results in the left-hand column of Fig. 2, the ETKF/NOLOC scheme exhibits different behaviour. For $lead = 1$, the ETKF/NOLOC method directs the AO to locations that are more localized around the verification region x_1. For longer lead times, the ETKF/NOLOC selects locations far from the verification region slightly more frequently than

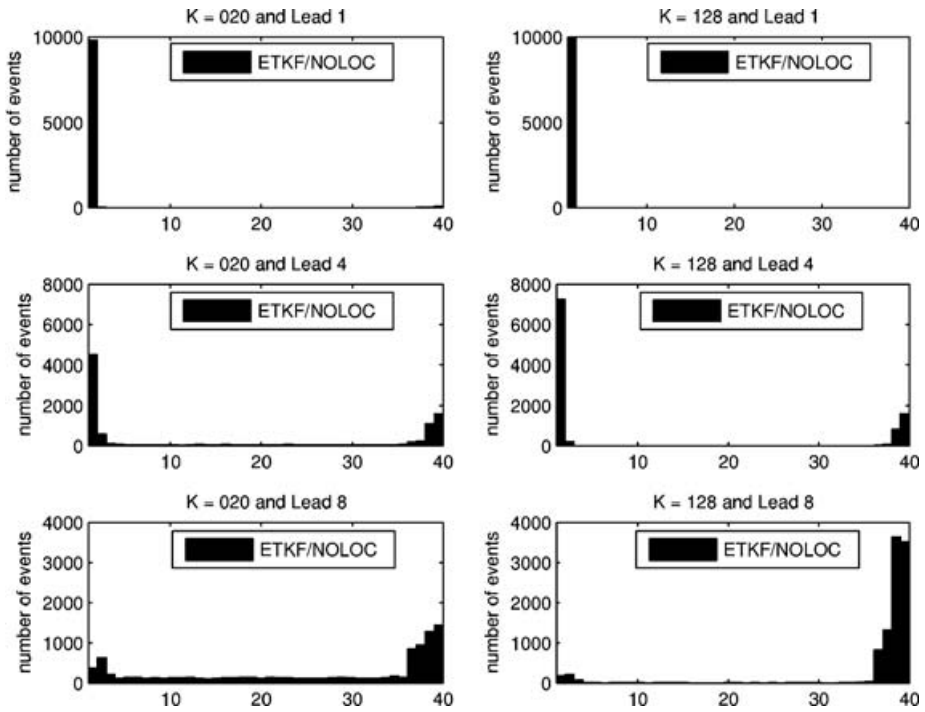

Fig. 3. Histograms of locations selected for the adaptive observation (AO) over the 10^4 experiments. The ETKF/NOLOC scheme has been used to select the location of the AO. The left-hand column shows results for an ensemble size $K = 20$, the right-hand column is for $K = 128$. For both ensemble sizes, results from three independent experiments are shown, where $lead = 1, 4, 8$ (recall that $lead \triangle t = t_v - t_a = t_a - t_i$).

the NONLIN-NOLOC. This contrasting statistical behaviour is reflected in higher mean forecast errors for the ETKF/NOLOC method (compared to NONLIN-NOLOC and NONLIN-LOC).

To understand the ETKF/NOLOC results, the criterion that the ETKF/NOLOC scheme uses to select the AO location can be examined. Using the analytical expression for the updated variance for x_1 at t_v given a hypothetical observation at location α at t_a, the ETKF/NOLOC scheme chooses the location that *maximizes*,

$$\frac{r^2_{x_{1,t_v},x_{\alpha,t_a}}}{1+\sigma^2_{\alpha,obs}\big/\sigma^{f\,2}_{x_\alpha}(t_a)}, \tag{5}$$

where $r^2_{x_{1,t_v},x_{\alpha,t_a}}$ is the squared sample correlation coefficient between $x_1(t_v)$ and $x_\alpha(t_a)$ computed using $[\mathbf{x}^f_{t_a,k}]$ and $[\mathbf{x}^f_{t_v,k}]$, $\sigma^{f\,2}_{x_\alpha}(t_a)$ is the forecast standard deviation for x_α and $\sigma_{\alpha,obs} = 0.2$ is the observational error standard deviation. A derivation of this maximization criterion is given in Appendix.

To understand the behaviour of the ETKF/NOLOC, it is instructive to examine the role of $r^2_{x_{1,t_v},x_{\alpha,t_a}}$ in (5) for short and long *lead* experiments. For the $lead = 1$ results, $r^2_{x_{1,t_v},x_{\alpha,t_a}}$ is plotted for several different state variables in Figs. 4 and 5. Results for ensemble sizes $K = 20$ and $K = 128$ are shown. For both ensemble sizes, the value of $r^2_{x_{1,t_v},x_{1,t_a}}$ appears to dominate over all others. The empirical results suggest that for both the $K = 20$ and $K = 128$ experiments, $r^2_{x_{1,t_v},x_{1,t_a}}$ is so dominant in (5), that ETKF/NOLOC scheme rarely directs observations away from x_1. For the $K = 20$ results, the differences between the locations chosen by the NONLIN-NOLOC, NONLIN-LOC and ETKF/NOLOC scheme are subtle for $lead = 1$. However, the more localized distribution of AOs given by the ETKF/NOLOC method results in significantly higher mean forecast errors. It is not clear how this problem can be overcome. As shown on the right-hand side of Fig. 3, increasing the ensemble size from $K = 20$ to $K = 128$ causes an even greater localization of the AO, since the correlation between $x_1(t_a)$ and $x_1(t_v)$ becomes *slightly* less noisy (this can be seen visually by comparing the $K = 20$ and $K = 128$ results in the upper left-hand panel of Fig. 4) and therefore even more dominant in (5).

For linear dynamics, the ETKF/NOLOC and NONLIN-NOLOC are equivalent (as discussed in Section 3.4). The large discrepancy in performance for the $lead = 1$ results in Fig. 1 is surprising, since one might expect linear dynamics to apply as the error doubling time of L96 is roughly 8 $\triangle t$. However, for linear dynamics to apply, the average size of the ensemble perturbations about the ensemble mean must also be small. One way of quantifying this is to compute the mean forecast standard deviation of $x_1(t_a)$ over all the ETKF/NOLOC experiments. The mean forecast standard deviation for $x1(t_a)$ was found to be roughly 0.8. This value is large given that the climatological standard deviation of each individual state variable is roughly 3.6. The mean size of the perturbations about the ensemble mean, suggests that the ensemble perturbations evolve non-linearly from t_i to t_v. The discrepancy in the results for $lead = 1$ is a reflection

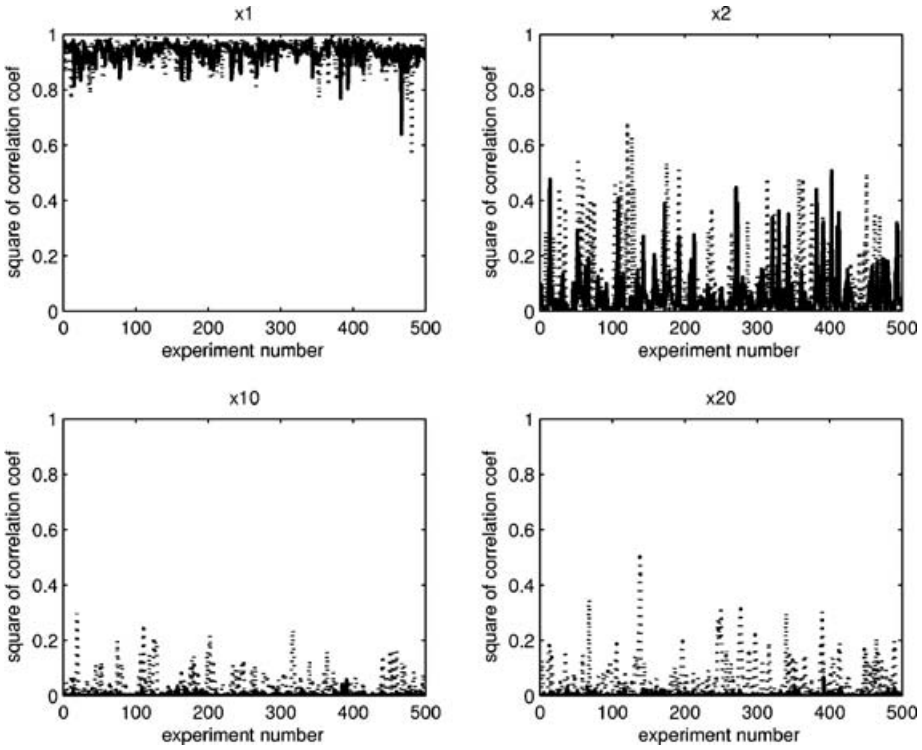

Fig. 4. For the ETKF/NOLOC experiment where the lead time is $1\triangle t$. The square of the correlation coefficient between the state variable indicated on the top of each panel at the adaptive observation time t_a, and the state variable of interest x_1 at the verification time t_v. Statistics are plotted for a sample of 500 consecutive experiments from the total of 10^4 (all samples are qualitatively similar). The correlation coefficients are computed from the ensemble forecasts for t_a and t_v initialized at t_i. The dashed line is for ensemble size $K = 20$, the solid line is for $K = 128$.

of this non-linearity. To confirm that the ETKF/NOLOC and NONLIN-NOLOC do indeed converge, 10^6 experiments were run for $\triangle t = t_v - t_a = t_a - t_i = 0.0001$ and $lead = 1$. The mean forecast standard deviation for $x_1(t_t)$ was found to be roughly 0.06 over all the ETKF/NOLOC experiments. The combined effect of small $\triangle t = 0.0001$ and perturbation size was apparently enough for linearity to apply since the ETKF/NOLOC and NONLIN-NOLOC were found to give identical results.

What happens for the longer *lead* experiments? $r^2_{x_{1,t_v},x_{\alpha,t_a}}$ for the $lead = 8$ experiments is plotted in Figs. 6 and 7. $r^2_{x_{1,t_v},x_{\alpha,t_a}}$ for grid points immediately upstream of x_1 are generally higher than grid points far away (leading to more frequent observations of locations nearby the verification region). Figures 6 and 7 show that occasionally, there are high correlations associated with grid points x_{10}, x_{20}, x_{30} for the small ensemble size $K = 20$, but they do not appear for the large ensemble size $K = 128$. It appears that these 'spurious' correlations are arising due to the smaller ensemble size. For the $K = 20$ and $lead = 8$ results, the ETKF/NOLOC methods frequently direct observations to locations far from the verification region (surprising given the group velocity of the most unstable wavenumber). For the $K = 128$ experiments, the lower right panel of Fig. 3 demonstrates that the ETKF/NOLOC rarely directs the AO to far away locations ('physically surprising' locations). It appears that the behaviour of the ETKF/NOLOC for the $K = 20$ and $lead = 8$ results is attributable to the ensemble size. *The nature of the sampling problem in the NONLIN-NOLOC method is not quite the same as the sampling problem in the ETKF/NOLOC method.* As discussed in Section 5.1, the sampling problem encountered in the NONLIN-NOLOC scheme is for statistical estimates of quantities relating state variables valid for the *same* time, t_a. For the ETKF/NOLOC scheme, the sampling problem is for statistical estimates of quantities relating state variables valid for *different* times, the AO time t_a and the verification time t_v.

By comparing the lower left panels of Figs. 2 and 3, note that the ETKF/NOLOC scheme directs the observations to locations far from the verification region slightly more often. Intuitively, one expects correlations for state variables valid at different times (relevant to the ETKF/NOLOC scheme) to be less than the expected correlations for state variables valid at the same time (relevant to the NONLIN-NOLOC scheme). As the expected correlations are diminished, the effects of small ensemble size are more severe. As a result, it is not surprising that the ETKF/NOLOC directs the AO to far away locations more often.

For the $lead = 8$ results in the lower left panel of Fig. 3, the ETKF/NOLOC scheme observes one of the locations $[x_{12}, \ldots, x_{29}]$ 2333 times out of the 10^4 experiments (compared to 1863 times for the NONLIN-NOLOC scheme). Again, given

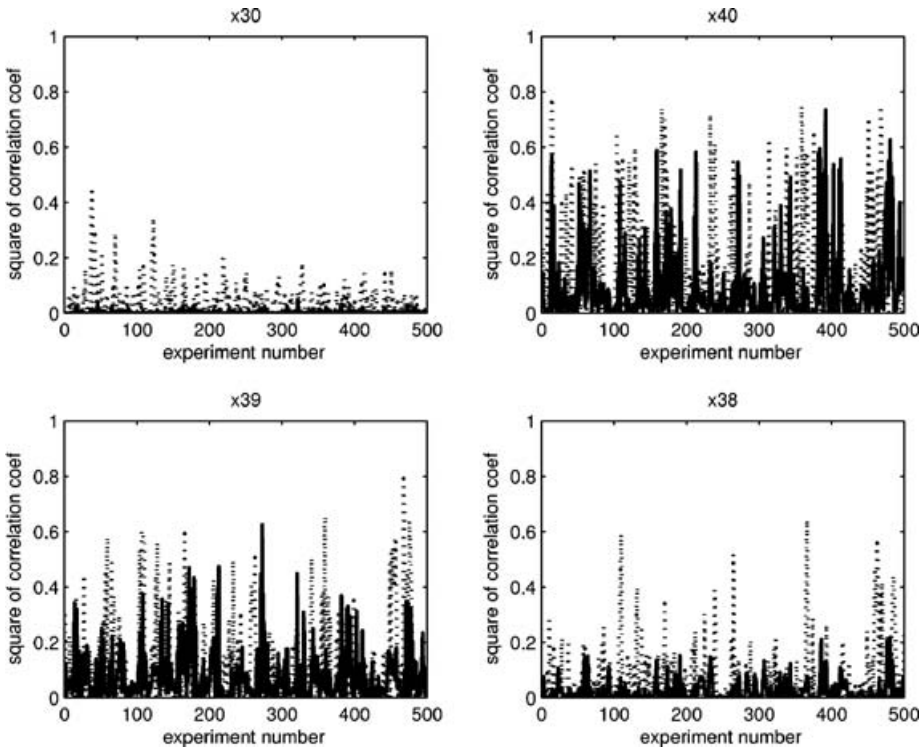

Fig. 5. For the ETKF/NOLOC experiment where the lead time is 1 Δt. Same as Fig. 4 except for different state variables indicated at the top of each panel. The dashed line is for $K = 20$ and solid line for $K = 128$.

the group velocity of the most unstable wavenumber, one may speculate that these 2333 observations led to very little improvement in the forecast errors. Similar calculations to the NONLIN-NOLOC cases were done. The average forecast error in $x_1(t_v)$ over the 2333 cases was 2.892. If the observation was not assimilated at t_t, the mean forecast error in $x_1(t_v)$ would have been 2.893. Therefore, over the 10^4 experiments, when the observation was directed to one of the grid points $x_{12}, \ldots, x_{29}$, nearly zero improvement in forecast error was achieved. The lower left panel of Fig. 3 shows that the most frequently observed grid points were x_{38}, x_{39} and x_{40}. Had the observations been directed to x_{38}, x_{39} or x_{40}, the mean forecast errors would have been 2.766, 2.612 and 2.710 respectively (again, these calculations were done retrospectively as for the NONLIN-NOLOC results).

5.3. *The impacts of space/time covariance localization on ETKF-type schemes: ETKF-SPACETIME and ETKF-SPACE*

Given that operational systems can only run a limited number of ensemble members, and the results of Section 5.2, one is led to ask if the performance of ETKF-type schemes can be improved through the use of CL? Section 5.1 demonstrates that for the case of $K = 20$, the performance of the non-linear ensemble mean scheme is improved through the use of CL. In what follows, the benefits of using CL in ETKF-type schemes are demonstrated.

First, consider the forecast variance for x_1 at t_v given a hypothetical observation of location α at t_a ($\sigma^{f\,2}_{x_1}(t_v|\mathbf{H}_{t_a,\alpha})$) computed by the ETKF-type methods examined in this paper,

$$\sigma^{f\,2}_{x_1}(t_v|\mathbf{H}_{t_a,\alpha}) = \sigma^{f\,2}_{x_1}(t_v) - \frac{f(|d|_{1,\alpha})^2 \sigma^{f\,4}_{x_{1,t_v},x_{\alpha,t_a}}}{\sigma^{f\,2}_{x_\alpha}(t_a) + \sigma^2_{\alpha,\mathrm{obs}}}, \tag{6}$$

where $\sigma^{f\,2}_{x_1}(t_v)$ is the forecast variance for $x_1(t_v)$, $\sigma^{f\,4}_{x_{1,t_v},x_{\alpha,t_a}}$ is the squared forecast covariance between $x_1(t_v)$ and $x_\alpha(t_a)$, $\sigma^{f\,2}_{x_\alpha}(t_a)$ is the forecast variance for grid point $x_\alpha(t_a)$ and $\sigma^2_{\alpha,\mathrm{obs}}$ is the observational variance of the observation at grid point α (the forecast variances/covariance are computed from the ensemble forecast generated at t_i). $f(|d|_{1,\alpha})$ is some function of the 'distance' $|d|_{1,\alpha}$ between $x_1(t_v)$ and $x_\alpha(t_a)$. See Appendix for a derivation of (6). The location α which *minimizes* (6) is chosen by the ETKF-type scheme as the location of the AO.

For the implementation with no localization (ETKF/NOLOC), $f(|d|_{1,\alpha}) = 1$ in (6). Note that in B01, eq. (6) is called the 'signal variance' for $f(|d|_{1,\alpha}) = 1$. The type of CL considered here multiplies $\sigma^{f\,4}_{x_{1,t_v},x_{\alpha,t_a}}$ by a factor $f(|d|_{1,\alpha})^2$, where $|d|_{1,\alpha}$ is the absolute 'distance' between state variable $x_1(t_v)$ and $x_\alpha(t_a)$, and f is a function with compact support that falls off with increasing 'distance'. We take f to be the same fifth-order piecewise rational function characterized by a parameter $c = 0.3$ (Gaspari and Cohn, 1999) used in the DA scheme (see Section 4.1.1.). This multiplication has the potential to damp 'spurious' correlations

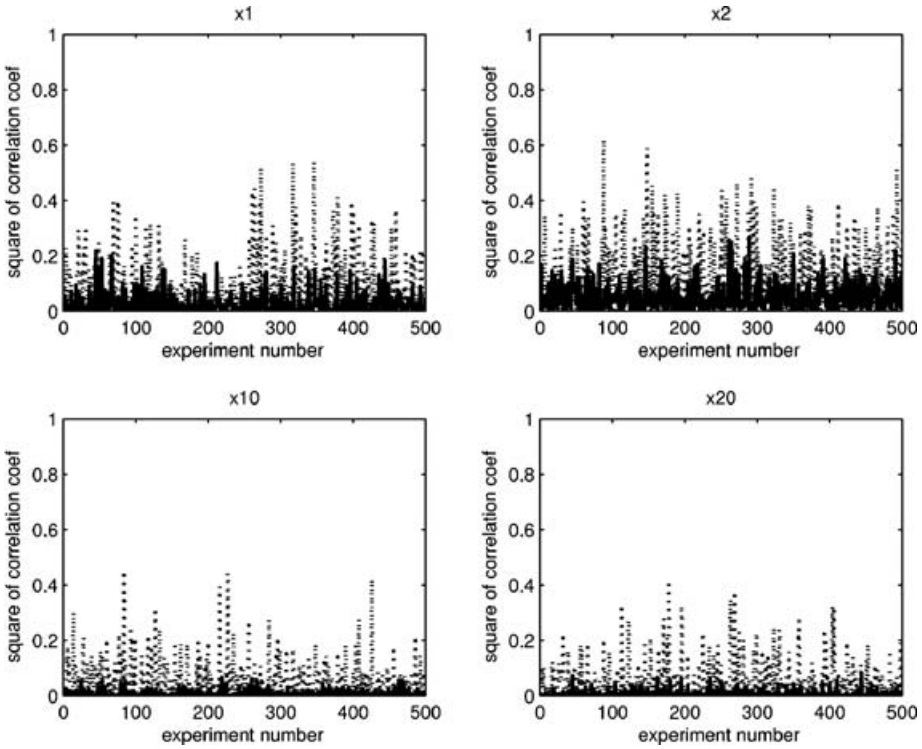

Fig. 6. For the ETKF/NOLOC experiment where the lead time is 8 $\triangle t$. The square of the correlation coefficient between the state variable indicated on the top of each panel at the adaptive observation time t_a, and the state variable of interest x_1 at the verification time t_v. Statistics are plotted for a sample of 500 consecutive experiments from the total of 10^4 (all samples are qualitatively similar). The correlation coefficients are computed from the ensemble forecasts for t_a and t_v initialized at t_i. The dashed line is for ensemble size $K = 20$, the solid line is for $K = 128$.

between $(x_1(t_v))$ and observed variable $(x_\alpha(t_a))$ which may be degrading the performance of the ETKF/NOLOC scheme. Given that the covariance $\sigma^{f^2}_{x_{1,t_v},x_{\alpha,t_a}}$ is computed using information at different times, it seems reasonable to assume that the distance $|d|_{1,\alpha}$ should somehow incorporate flow information for the time interval $t_a \rightarrow t_v$. Properly characterizing the 'distance' $|d|_{1,\alpha}$ and using $f(|d|_{1,\alpha})$, will hopefully lead to an effective space- and time-dependent localization. For the scheme labelled ETKF-SPACE, the distance $|d|_{1,\alpha}$ is simply taken as the physical distance between grid points $x_1(t_v)$ and $x_\alpha(t_a)$, effectively treating the grid points as though they are at the same time. Only the simplest of flow-dependent definitions of $|d|_{1,\alpha}$ are offered here. Suppose that the L96 dynamics is well characterized by the group velocity of the most unstable wavenumber. The technique that is tried here is to translate the location of $x_\alpha(t_a)$ by $+c_{gu}lead$ $\triangle t$ zones downstream (for a given *lead* experiment). Here, c_{gu} is defined as the group velocity of the most unstable wavenumber. This definition effectively sets the 'distance' between $x_1(t_v)$ and $x_\alpha(t_a)$ to zero if the grid point $x_\alpha(t_a)$ happens to be $-c_{gu}lead\ \triangle t$ upstream of x_1. Once $x_\alpha(t_a)$ has been translated downstream, the distance between the translated grid point and grid point x_1 is computed and set equal to $|d|_{1,\alpha}$. The implementation that uses this method to compute $|d|_{1,\alpha}$ is labelled ETKF-SPACETIME. In experiments with GCMs, Szunyogh et al. (2000, 2002) demonstrate that the velocity of data impacts is associated with baroclinic Rossby wave group velocity, providing motivation for using the group velocity of the most unstable wavenumber in this context.

The histograms for the ETKF-SPACE and ETKF-SPACETIME schemes are plotted in Fig. 8. Not surprisingly, for $lead = 1$ the AO remains very localized around the verification region x_1. For the $lead = 8$ experiments, the ETKF-SPACE and ETKF-SPACETIME schemes no longer direct the AO to the so-called 'physically surprising' locations. The impact of the CL is evident by comparing the lower left and right panels of Fig. 8 and the lower left panel of Fig. 3. The CL schemes result in improved (compared to the ETKF/NOLOC) mean forecast errors for larger lead times as shown in Fig. 1. As well, the ETKF-SPACE and ETKF-SPACETIME outperform the FIXED scheme for all the 8 lead times, unlike the ETKF/NOLOC. The differences between the histograms for the ETKF-SPACE and ETKF-SPACETIME schemes are subtle, yet the space and time localization scheme picks upstream locations slightly more often. One likely reason for the subtle difference between the ETKF-SPACE and ETKF-SPACETIME results is the small group velocity of the most unstable wave. For the $K = 20$ results, the flow-dependent definition of distance yields a small improvement in forecast errors for longer lead times. For *lead* =

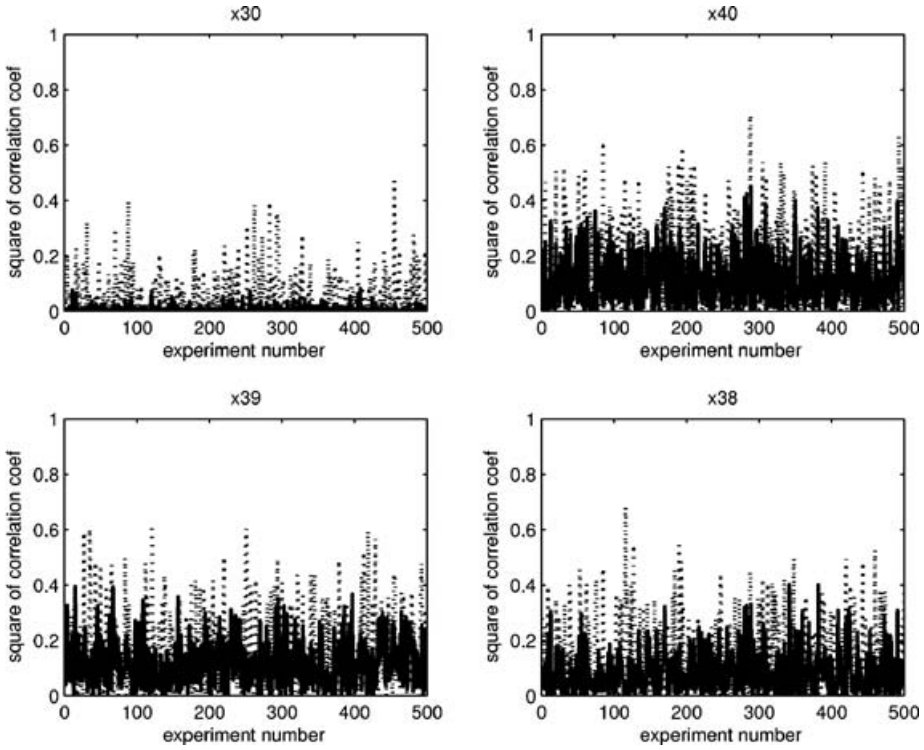

Fig. 7. For the ETKF/NOLOC experiment where the lead time is $8\triangle t$. Same as Fig. 6 except for different state variables indicated at the top of each panel. The dashed line is for $K = 20$ and solid line for $K = 128$.

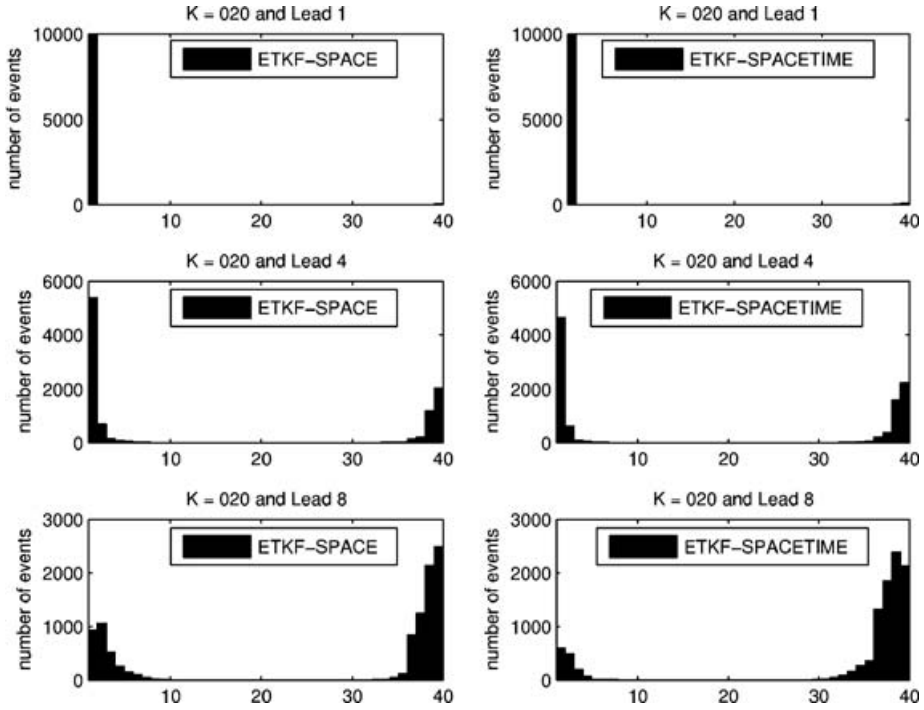

Fig. 8. Histograms of locations selected for the adaptive observation (AO) over the 10^4 experiments. The left-hand column shows results for the ETKF-SPACE scheme, the right-hand column for the ETKF-SPACETIME scheme. For both ensemble sizes, results from three independent experiments are shown, where $lead = 1, 4, 8$ (recall that $lead\triangle t = t_v - t_a = t_a - t_i$).

7, 8, ETKF-SPACETIME yields slightly better forecast errors for long *lead* experiments compared to the NONLIN-LOC scheme. We view these results optimistically, since statistically significant improvement over the ETKF-SPACE was achievable using only a crude representation of the L96 dynamics.

Given these results, it is interesting to note that in a comparison between total energy singular vectors and the ETKF AO techniques for 10 cases in the NORPEX experiment, Majumdar et al. (2002a) found that the ETKF generally identifies broader regions for AOs. Majumdar et al. (2002a) mention that a possible explanation of their results is the lack of a CL scheme appropriate for ensemble-based targeting. Given the results of this paper, it would be interesting to see if the performance of present day ETKF schemes could be improved through the use of CL. One simple yet potentially effective scheme would be to only calculate $\mathbf{P}^f(t_v|H^*_{t_d})$ for physically reasonable locations suggested by the group velocity of propagating disturbances.

6. Summary and conclusions

The type of ensemble-based AO methods of primary concern in this paper are so-called ETKF-type schemes. For a given initial time t_i, ETKF-type schemes make use of the ensemble forecast generated at t_i in testing various hypothetical AO networks. ETKF-type schemes do not require repeated integrations of the prediction model, making them suitable for real-time applications.

The underlying theory behind the ETKF-type methods is discussed in Section 3.3. This theory regards the ensemble forecast generated at t_i as a sample of a prior distribution for a joint state. The joint state combines system states at the relevant times beyond t_i. To test a given hypothetical AO network, the ensemble forecast can be updated in a manner consistent with the Kalman filter update equations (seen in Section 3.3.1 starting from Bayes rule). The forecast covariance of the system state at the verification time, conditioned on a hypothetical AO network can then be quantified. When a deterministic ensemble square root Kalman filter (Tippett et al., 2003) is used for the update *without* CL, this methodology is equivalent to the ETKF AO methodology developed by Bishop et al. (2001) as shown in Section 3.3.2 (this method has been called ETKF/NOLOC throughout this study). The ETKF/NOLOC is equivalent to a sequential in-time filtering process when the forecast ensemble evolves linearly (Section 3.4). In Section 3.3.3, new ETKF-type AO methodologies that make use of CL are introduced.

To develop an understanding of the various ETKF-type schemes discussed in Section 3.3, OSSEs in the non-linear L96 model have been analysed in Section 5. The OSSEs simulate the problem of choosing an AO in a large data void to improve forecasts of an important physical location. An ensemble adjustment Kalman filter (EAKF) has been used for the data assimilation. When implementing ensemble Kalman filters (EnKF) in realistic prediction models, a limited number of ensembles can be run. As a result, realistic implementations of EnKF often require heuristics such as covariance inflation/localization to handle filter divergence/sampling errors to achieve stable assimilations. The emphasis has been placed on analysis of ensemble size $K = 20$ results, since covariance inflation/localization were necessary to achieve stable assimilations, mimicking the situation often encountered in realistic applications.

Results have been obtained for a variety of lead times (varying from 6 h to 2 d). The short lead time experiments demonstrate that locations selected by the ETKF/NOLOC method are more localized when compared to the non-linear ensemble mean scheme (equivalent to a sequential in-time filtering process for non-linear dynamics described in Section 3.2). This is reflected in higher mean forecast errors for the ETKF/NOLOC compared to the non-linear ensemble mean method. The problem is not alleviated by increased ensemble size and is therefore not necessarily attributable to sampling error. It is not clear how this can be overcome in a systematic way. The results for longer lead times and $K = 20$ show that the performance of the ETKF/NOLOC method can be improved upon by the use of an ETKF-type scheme that makes use of space/time CL. Space/time CL is used to handle sampling error for relationships between state variables at *different* times. While several methods for CL have been suggested for ensemble-based data assimilation, no previous study has developed a method for ETKF-type AO schemes (Majumdar et al., 2002a). Two approaches to space/time CL for ETKF-type schemes are proposed in Section 3.3.3 (ETKF-SPACE, ETKF-SPACETIME). The results in Section 5.3 for longer lead times and $K = 20$ show that the ETKF-SPACE and ETKF-SPACETIME methods yield improved forecast errors compared to the ETKF/NOLOC. The first approach (ETKF-SPACE) simply treats state variables at different times as though they are at the same time to define a distance, which can be plugged into the localization function f. While this approach seems reasonable for small lead times, it is questionable for long lead times since evolution of the flow is ignored. The second approach (ETKF-SPACETIME), which yields a slight improvement over the first, uses a simple characterization of the L96 dynamics to define distances between state variables at different times. Operational systems typically use ensemble sizes much less than the size of the state vector. This fact, combined with the results in this paper suggests that an appropriate space/time CL scheme is crucial to realizing the full benefits of ETKF-type schemes when applied to realistic prediction problems. Development and testing of a generally applicable space/time CL scheme is left to a future study.

7. Acknowledgments

This work was completed while the first author was a PhD student in the Atmospheric and Oceanic sciences program at Princeton University. The authors would like to acknowledge members of GFDL and the AOS program at Princeton University for helpful

discussions leading up to this work. The authors would like to thank Sharan Majumdar for helpful discussions leading up to this work. The authors would also like to thank two anonymous reviewers for helpful comments that led to significant improvements in this paper.

8. Appendix
Derivation of $\sigma^{f\,2}_{x_1}(t_v|\mathbf{H}_{\alpha,t_a})$ for ETKF-type methods

The following derivation amounts to an application of the EAKF to the AO problem examined in Section 5 for ETKF-type methods. For a given t_i, assume that an ensemble forecast has been generated out to t_v. The augmented state $\mathbf{z} = [\mathbf{x}_{t_a}, \mathbf{x}_{t_v}]$ is a $2n$-dimensional vector (n is the model size). The ensemble forecast is considered a random sample of the prior distribution, $\mathbf{p}(\mathbf{x}_{t_a}, \mathbf{x}_{t_v})$. As discussed in Section 3.2, the prior ensemble is updated in a manner consistent with the Kalman Filter update equations. Going through a derivation of the Kalman Filter (Cohn, 1997) demonstrates that the expression for the updated covariance for the augmented state $\mathbf{z}$ can be written as,

$$\mathbf{P}^f_{aug}\left(t_v|\mathbf{H}_{aug_{\alpha,t_a}}\right) = \left(\mathbf{I} - \mathbf{K}\mathbf{H}_{aug_{\alpha,t_a}}\right)\mathbf{P}^f_{aug}(t_a, t_v), \tag{A1}$$

where α is the hypothetical AO location at t_a, $\mathbf{P}^f_{\text{aug}}(t_a, t_v|\mathbf{H}_{aug_{\alpha,\,t_a}})$ ($\mathbf{P}^f_{\text{aug}}(t_a, t_v)$) is the $2n \times 2n$ posterior (prior) covariance, $\mathbf{I}$ is a $2n \times 2n$ identity matrix, $\mathbf{K}$ is the $2n \times 1$ 'Kalman gain matrix' and $\mathbf{H}_{aug_{\alpha,\,t_a}}$ is the $1 \times 2n$ observation operator where all the columns are *zero* except for a 1 in the α^{th} column (corresponding to an observation of grid point α at t_a, subscript *aug* is used to indicate the augmented state space). The Kalman gain matrix is given by $\mathbf{K} = \mathbf{P}^f_{aug}(t_a, t_v)\mathbf{H}^T_{\text{aug}_{\alpha,t_a}}(\mathbf{H}_{\,aug_{\alpha,\,t_a}}\mathbf{P}^f_{aug}(t_a, t_v)\mathbf{H}^T_{aug\,\alpha,t_a} + \mathbf{R})^{-1}$, where $\mathbf{R} = \sigma^2_{\alpha,\text{obs}}$ is the observational variance at grid point α. For the MSE norm and problem of interest, we are only concerned with the updated variance for state variable x_1 at t_v. This is given by the $(n + 1, n + 1)$ element of $\mathbf{P}^f_{aug}(t_v|\mathbf{H}_{aug\,t_a,\alpha})$. Using eq. (A1), the updated variance is (after some simple algebra),

$$\sigma^{f\,2}_{x_1}(t_v|\mathbf{H}_{t_a,\alpha}) = \sigma^{f\,2}_{x_1}(t_v) - \frac{f(|d|_{1,\alpha})^2\sigma^{f\,4}_{x_{1,t_v},x_{\alpha,t_a}}}{\sigma^{f\,2}_{x_\alpha}(t_a) + \sigma^2_{\alpha,\text{obs}}}, \tag{A2}$$

where $\mathbf{H}_{t_a,\alpha}$ is the $1 \times n$ observation operator and $\sigma^{f\,4}_{x_{1,t_v},x_{\alpha,t_a}}$ is the squared covariance between $x_1(t_v)$ and $x_\alpha(t_a)$ computed from the ensemble forecast generated at t_i. For a finite size ensemble, the EAKF will generate an updated sample variance that fits eq. (A2) exactly. For the implementation *without* localization, $f(|d|_{1,\alpha}) = 1$ and is equivalent to what is computed by the ETKF/NOLOC. As discussed in Section 5.3, ETKF-SPACE and ETKF-SPACETIME methods specify values of f to diminish the potentially negative impacts of sampling error. Finally, the maximization criterion given in eq. (5) is simple to derive from eq. (A2). Since $\sigma^{f\,4}_{x_{1,t_v},x_\alpha,t_a} = r^2_{x_{1,t_v},x_\alpha,t_a}\sigma^{f\,2}_{x_\alpha}(t_a)\sigma^{f\,2}_{x_1}(t_v)$, one wants to maximize the second term in (A2) to minimize $\sigma^{f\,2}_{x_1}(t_v|H_\alpha, t_a)$.

References

Aberson, S. D. 2003. Targeted observations to improve operational tropical cyclone track forecast guidance. *Mon. Wea. Rev.* **131**, 1613–1628.

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.

Anderson, J. L. 2003. A local least squares framework for ensemble filtering. *Mon. Wea. Rev.* **131**, 634–642.

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.

Baker, N. L. and Daley, R. 2000. Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.* **126**, 1431–1454.

Bergot, T. 1999. Adaptive observations during FASTEX: a systematic survey of upstream flights. *Quart. J. Roy. Meteor. Soc.* **127**, 635–655.

Bergot, T. 2001. Influence of the assimilation scheme on the efficiency of adaptive observations. *Quart. J. Roy. Meteor. Soc.* **127**, 635–660.

Bergot, T., Hello, G., Joly, A. and Malardel, S. 1999. Adaptive observations: a feasibility study. *Mon. Wea. Rev.* **127**, 743–765.

Berliner, L. M., Lu, Z. and Snyder, C. 1999. Statistical design for adaptive weather observations. *J. Atmos. Sci.* **56**, 2536–2552.

Bishop, C. H. and Toth, Z. 1999. Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.

Bishop, C. H., Etherton, B. J. and Majumdar, S. 2001. Adaptive sampling with the ensemble transform Kalman filter, part I. *Mon. Wea. Rev.* **129**, 420–436.

Bishop, C. H., Reynolds, C. A. and Tippett, M. K. 2003. Optimization of the fixed global observing network in a simple model. *J. Atmos. Sci.* **60**, 1471–1489.

Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteor. Soc. Japan* **75**, 257–288.

Emanuel, K. A., Raymond, D., Betts, A., Bosart, L., Bretherton, C. and co-authors 1995. Report of the first prospectus development team of the U.S. Weather Research Program to NOAA and NSF. *Bull. Amer. Meteor. Soc.* **76**, 1194–1208.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10 143–10 162.

Gaspari, G. and Cohn, S. E. 1998. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.* **125**, 723–757.

Gelaro, R., Reynolds, C. A., Langland, R. H. and Rohaly, G. D. 2000. A predictability study using geostationary satellite wind observations during NORPEX. *Mon. Wea. Rev.* **128**, 3789–3807.

Hamill, T. and Snyder, C. 2002. Using improved background error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.* **130**, 1552–1572.

Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790.

Hansen, J. A. and Smith, L. 2000. The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.* **57**, 2859–2871.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.

Ide, K., Courtier, P., Ghil, M. and Lorenc, A. 1997. Unified notation for data assimilation: Operational sequential and variational. *J. Meteor. Soc. Japan* **75**, 181–189.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 376 pp.

Khare, S. P. 2004. Observing network design for improved prediction of geophysical fluid flows—analysis of ensemble methods. PhD thesis, Princeton University, 195 pp.

Langland, R. H., Toth, Z., Gelaro, R., Szunyogh, I., Shapiro, M. A., and co-authors 1999. The North Pacific Experiment (NORPEX-98): targeted observations for improved north American weather forecasts. *Bull. Amer. Meteor. Soc.* **80**, 1363–1384.

Leutbecher, M. 2003. A reduced rank estimate of forecast error variance changes due to intermittent modification of the observing network. *J. Atmos. Sci.* **60**, 729–742.

Lorenz, E. N. 1995. Predictability—a problem partly solved. *ECMWF Seminar Proceedings I*, **1**, 1–18.

Lorenz, E. N. and Emanuel, K. A. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414.

Majumdar, S. J., Bishop, C. H., Etherton, B. J., Szunyogh, I. and Toth, Z. 2001. Can an ensemble transform Kalman filter predict reduction in forecast error variance produced by targeted observations. *Quart. J. Roy. Meteor. Soc.* **127**, 2803–2820.

Majumdar, S. J., Bishop, C. H., Buizza, R. and Gelaro, R. 2002a. A comparison of ensemble transform Kalman filter targeting guidance with ECMWF and NRL total energy singular vector guidance. *Quart. J. Roy. Meteor. Soc.* **128**, 1–23.

Majumdar, S. J., Bishop, C. H. and Etherton, B. J. 2002b. Adaptive sampling with the ensemble transform Kalman filter. Part II: field program implementation. *Mon. Wea. Rev.* **130**, 1356–1369.

Montani, A., Thorpe, A. J., Buizza, R. and Unden, P. 1999. Forecast skill of the ECMWF model using targeted observations during FASTEX. *Quart. J. Roy. Meteor. Soc.* **125**, 3219–3240.

Morss, R. E., Emanuel, K. A. and Snyder, C. 2001. Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.* **58**, 210–232.

Palmer, T. N., Gelaro, R., Barkmeijer, J. and Buizza, R. 1998. Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.* **55**, 633–653.

Pu, Z.-X. and Kalnay, E. 1999. Targeting observations with the quasi-linear and adjoint NCEP global models: performance during FASTEX. *Quart. J. Roy. Meteor. Soc.* **125**, 3329–3337.

Snyder, C. and Zhang, F. 2003. Assimilation of simulated doppler radar observation with an ensemble Kalman filter. *Mon. Wea. Rev.* **131**, 1663–1677.

Szunyogh, I., Toth, Z., Morss, R. E., Majumdar, S. J., Etherton, B. J. and co-authors. 2000. The effect of targeted dropsonde observations during the 1999 winter storm reconnaissance program. *Bull. Amer. Meteor. Soc.* **120**, 3520–3537.

Szunyogh, I., Toth, Z., Zimin, A. V., Majumdar, S. J. and Persson, A. 2002. Propagation of the effect of targeted observations: the 2000 Winter Storm Reconnaissance Program. *Mon. Wea. Rev.* **130**, 1144–1165.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Trevisan, A. and Uboldi, F. 2004. Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle system. *J. Atmos. Sci.* **61**, 103–113.

Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913–1924.