A Transformed Lagged Ensemble Forecasting Technique for Increasing Ensemble Size

ANDREW R. LAWRENCE* AND JAMES A. HANSEN+

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

An ensemble-based data assimilation approach is used to transform old ensemble forecast perturbations with more recent observations for the purpose of inexpensively increasing ensemble size. The impact of the transformations are propagated forward in time over the ensemble's forecast period without rerunning any models, and these transformed ensemble forecast perturbations can be combined with the most recent ensemble forecast to sensibly increase forecast ensemble sizes. Because the transform takes place in perturbation space, the transformed perturbations must be centered on the ensemble mean from the most recent forecasts. Thus, the benefit of the approach is in terms of improved ensemble statistics rather than improvements in the mean. Larger ensemble forecasts can be used for numerous purposes, including probabilistic forecasting, targeted observations, and to provide boundary conditions to limited-area models. This transformed lagged ensemble forecasting approach is explored and is shown to give positive results in the context of a simple chaotic model. By incorporating a suitable perturbation inflation factor, the technique was found to generate forecast ensembles whose skill were statistically comparable to those produced by adding nonlinear model integrations. Implications for ensemble forecasts generated by numerical weather prediction models are briefly discussed, including multimodel ensemble forecasting.

1. Introduction

A popular method used to account for initial condition uncertainty in Numerical Weather Prediction (NWP) is the ensemble approach to forecasting (Palmer et al. 1992; Toth and Kalnay 1993). The atmosphere's initial condition probability distribution function (PDF) is discretely sampled, and each sample is propagated forward by the NWP model. The resulting collection of forecasts is treated as a discrete sample of the forecast PDF. One factor that can limit an ensemble forecast's utility is its relatively small size. This paper presents an approach to increase forecast ensemble size using lagged ensemble forecasts that have been transformed to account for all observations that have be-

come available since the forecasts were launched. This transformed lagged ensemble forecasting (TLEF) technique is equally valid for single-model and multimodel ensembles, but the technique introduced here places emphasis on the single-model case in the context of an idealized model. Issues associated with operational NWP ensembles and multimodel scenarios will appear in a companion paper.

The TLEF proves an extremely inexpensive way to increase ensemble size, but as will be shown below, because the methodology is applied to ensemble perturbations it impacts only the higher-order moments of the distribution, not the mean. Thus, the traditional benefit of improved ensemble mean skill for increased ensemble size (Buizza 1997; Houtekamer and Mitchell 1998) is not realized for TLEF ensembles. Instead, benefits are derived by other ensemble-based methods such as targeting or limited-area modeling where increased numbers of ensemble forecast perturbations can be put to good use.

Increasing the size of NWP ensemble forecasts is a computationally demanding endeavor. The method described here increases ensemble size inexpensively by utilizing old ensemble forecasts. The idea of using old forecasts to augment the most recent forecast is not

* Current affiliation: European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom.

+ Current affiliation: Marine Meteorology Division, Naval Research Laboratory, Monterey, California.

Corresponding author address: Andrew R. Lawrence, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, United Kingdom.
E-mail: a.lawrence@ecmwf.int

new. Lagged average forecasting was introduced by Hoffmann and Kalnay (1983) as a precursor to ensemble forecasting. Instead of defining an ensemble of initial conditions at the initial time and propagating them forward to some forecast time of interest, one generates a forecast ensemble by collecting the single forecasts launched at different times but all verifying at the same time. Because the different forecasts are conditioned upon different amounts of observational information, the authors proposed weighting the older forecasts by a measure of their expected skill. In the majority of applications, each of the forecasts are given equal weight (Kalnay 2003). A more sophisticated weighting for lagged average forecasting was proposed by Ebisuzaki and Kalnay (1991). The method proposed in the current work differs from the above approaches in two important ways: 1) it combines ensemble forecasts rather than a single forecast, and 2) it uses the combination of observations and ensemble-based uncertainty information to provide a statistically rigorous, state-dependent "weighting."

The TLEF, as presented here, is based on the ensemble transform Kalman filter (ETKF) of Bishop et al. (2001). It is important to note that the technique is not limited to the ETKF or even to deterministic ensemble filters; Evensen (2003) provides a formulation of the ensemble Kalman filter (EnKF) that is amenable to application with the TLEF. In our implementation of the TLEF, a transformation based on the ETKF is performed on old ensemble forecasts each time new observations become available. The information contained in the transformation is then propagated forward in the ensemble subspace under the assumption of linear error evolution to alter the ensemble forecasts *without rerunning models*. The ensemble forecast itself provides all the information necessary to propagate the impact of the observations forward in time: rerunning the numerical model is not necessary. The transformation process results in a collection of ensemble forecasts that are directly comparable at any verification time; all ensemble members have been conditioned upon the same number of observations. This approach effectively increases the forecast ensemble size without adding any new model runs.

The most recent ensemble forecast will provide an estimate of $p[\mathbf{x}^t(\tau_{\mathrm{ver}})|\mathbf{Y}(0)]$: the probability of the true state of the system at $t = \tau_{\mathrm{ver}}$ given all observations up to and including those at $t = 0$. The ensemble forecast launched one observation time prior to the most recent provides an estimate of $p[\mathbf{x}^t(\tau_{\mathrm{ver}})|\mathbf{Y}(-\tau_{\mathrm{obs}})]$, a different distribution from that which is desired. The aim of the TLEF is to transform the samples from $p[\mathbf{x}^t(\tau_{\mathrm{ver}})|\mathbf{Y}(-\tau_{\mathrm{obs}})]$, so that they become samples from $p[\mathbf{x}^t(\tau_{\mathrm{ver}})|\mathbf{Y}(0)]$ and can be used to augment the most recent ensemble forecast. Ensembles can be transformed multiple times, and each time new observations become available previously transformed ensembles can undergo another transformation. The number of times an ensemble forecast can be transformed is controlled by the validity of the underlying linearity assumption, and the level of model error. It is important to point out that the approach only provides new information when the model dynamics are nonlinear and/or when the ensemble prediction system's ensemble construction methodology is different from that used by the TLEF (the ETKF in this case). This point is discussed in section 2c.

This manuscript describes a proof-of-concept study of the TLEF technique by applying it to an idealized model. In section 2 the ETKF is briefly described, followed by a description of the TLEF methodology. Section 3 explores the implementation of this framework in an idealized model environment where the sensitivity of the results can be extensively studied. Section 4 concludes with a brief discussion of the results and some suggestions for implementation into an NWP framework and possible improvements to the approach.

2. Formulation of the TLEF

a. The ETKF

The ETKF is conceptually similar to both the ensemble square root filter (Whitaker and Hamill 2002) and the ensemble adjustment filter (Anderson 2001). The ETKF has been applied to the targeted observation problem (Bishop et al. 2001), ensemble generation (Wang and Bishop 2003), and data assimilation (Majumdar et al. 2002). In this work we use the ETKF to propagate observational influence forward in time. The TLEF can use the ETKF machinery to update both the forecast mean and ensemble perturbations when observations become available, and then propagate those changes forward in the ensemble subspace under the assumptions that they evolve linearly. *This is achieved by simply applying the transformation matrix derived at the observation time to all forecast perturbations.*

The ETKF is an approximation of the extended Kalman filter (Jazwinski 1969). The extended Kalman filter is a near-optimal algorithm that provides a minimum error variance estimate of the state $\mathbf{x}^a$, and its uncertainty $\mathbf{P}^a$ given a prior estimate of the state $\mathbf{x}^f$, available observational data $\mathbf{y}$, and the uncertainty associated with each, $\mathbf{P}^f$ and $\mathbf{R}$, respectively.[1] The first

---

[1] For notational clarity the time indices have been removed from these expressions.

guess is typically in the form of a short-term forecast, $\mathbf{x}^f = \mathbf{F}(\mathbf{x}^a)$, where $\mathbf{F}$ can be a nonlinear operator. The uncertainty associated with the first guess is obtained by linearly propagating the uncertainty associated with the previous minimum error variance estimate forward to the new observation time, $\mathbf{P}^f = \mathbf{M}\mathbf{P}^a\mathbf{M}^T$, where $\mathbf{M}$ is the linear uncertainty propagator defined by integrating the Jacobian of $\mathbf{F}$. The ensemble approximations to the extended Kalman filter take a Monte Carlo, or ensemble, approach to producing the uncertainty associated with the first guess, $\mathbf{P}^f = \mathbf{X}^f\mathbf{X}^{fT}$. Here $\mathbf{X}^f$ is a matrix whose columns consist of the vectors $(\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(N_{\text{ens}} - 1)^{-1/2}$ for $i = 1, \ldots, N_{\text{ens}}$, where $\bar{\mathbf{x}}$ is the ensemble mean, $\mathbf{x}_i^f$ is an individual ensemble member, and N_{ens} is the number of ensemble members: $\mathbf{X}^f$ is a matrix of normalized ensemble perturbations.

The minimum error variance state estimate is given by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f), \quad (1)$$

where $\mathbf{H}$ is a (linear) operator that maps from model space to observation space,[2] and $\mathbf{y}$ is an observation. To obtain analysis ensemble perturbations, forecast ensemble perturbations $\mathbf{X}^f$ are subjected to a *linear transformation* (via a $\mathbf{T}$ matrix) so that the distribution of transformed ensemble perturbations exactly matches the Kalman filter expression for the uncertainty associated with the analysis estimate:

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}^f = \mathbf{X}^f\mathbf{T}\mathbf{T}^T\mathbf{X}^{fT}.$$
$$(2)$$

Bishop et al. (2001) described a computationally efficient approach to solving Eqs. (1) and (2) when the ensemble size N_{ens} is smaller than the number of observations N_{obs}. Note that as is the case for all ensemble filters, if the ensemble size is smaller than the rank of the true forecast error covariance matrix in observation space, the mean ETKF analysis increments and error covariance updates are far from optimal. The ETKF minimum error variance estimate is obtained through the following expression:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}^f[\mathbf{C}\mathbf{\Gamma}^{1/2}(\mathbf{\Gamma} + \mathbf{I})^{-1}\mathbf{E}^T](\mathbf{R}^{-1/2}\mathbf{y} - \mathbf{R}^{-1/2}\mathbf{H}\bar{\mathbf{x}}^f),$$
$$(3)$$

where $\mathbf{C}$ and $\mathbf{\Gamma}$ contain the eigenvectors and eigenvalues of $\mathbf{X}^{fT}\mathbf{H}^T\mathbf{R}^{-T/2}\mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^f$, respectively, and $\mathbf{E}$ contains the eigenvectors of $\mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^f\mathbf{X}^{fT}\mathbf{H}^T\mathbf{R}^{-T/2}$ that are

---

[2] Note that because the ensemble-based filters can be configured so that the observation operator operates on $\mathbf{X}^f$ rather than $\mathbf{P}^f$: $\mathbf{H}$ need not be linear.

not in the null space of $\mathbf{X}^f\mathbf{X}^{fT}$. Only the highlights of the method are given here, for a complete treatment the interested reader should see Bishop et al. (2001).

The transformation matrix can then be shown to take the form

$$\mathbf{T} = \mathbf{C}(\mathbf{\Gamma} + \mathbf{I})^{-1/2}\mathbf{C}^T, \quad (4)$$

so that the normalized analysis ensemble perturbations are given by

$$\mathbf{X}^a = \mathbf{X}^f\mathbf{T}. \quad (5)$$

To produce the new analysis ensemble, the normalization is removed, and the columns of the unnormalized $\mathbf{X}^a$ are added to the minimum error variance of the state vector $\bar{\mathbf{x}}^a$. Note that the transformation $\mathbf{T}$ only depends on the forecast and observation error *statistics*, not on the forecasts or observations themselves. Equations (2)–(5) provide the tools necessary to perform ensemble-based data assimilation using the ETKF.

The expression for the $\mathbf{T}$ matrix in Eq. (4) includes a matrix $\mathbf{C}^T$, which, although not included in the original formulation by Bishop et al. (2001), provides an extra transformation to ensure the new analysis perturbations are centered about a zero mean without changing their sample covariance. If this matrix is omitted, one of the analysis ensemble members is nearly indistinguishable from the mean. This is the "spherical simplex" ETKF formulation of Wang et al. (2004).

We reiterate that the TLEF is not dependent upon the use of the ETKF. Other ensemble data assimilation schemes can be similarly framed (see Tippett et al. 2003). As will be discussed later, the TLEF is derived under the assumption that the dynamics is linear, but will actually give no new information if the dynamics really is linear. The current work relies on the nonlinearity of the model and the mismatch between the scheme used to generate the forecast ensembles (the EnKF) and the scheme used in the TLEF (the ETKF). Other researchers may consider utilizing the EnKF in the TLEF framework; the perturbed observations utilized by the EnKF would allow transformed ensembles to be different from untransformed ensembles in the case of linear dynamics.

b. Ensemble inflation

If the ensemble size is small in relation to the number of directions in which the forecast error projects, the expression for $\mathbf{P}^f$ underestimates the total analysis error variance both because it lacks contributions from important parts of the error space and because of sample size limitations. A simple approach to partially address this deficiency (although not to entirely ameliorate the problem) uses an inflation factor to boost

the perturbation amplitude. From the basic methodology of Wang and Bishop (2003), this inflation factor, denoted by Π_r, is calculated to ensure that the forecast ensemble variance at some chosen lead time t_r is made consistent with the control forecast error variance at the observation locations. Therefore, assuming that the statistics of the next globally averaged forecast will be similar to those of the previous forecast, the value of Π_r can be obtained by

$$\Pi_r = \Pi_{r-1} \sqrt{\alpha_r}, \tag{6}$$

where $r = 1, \ldots, N_r$ (N_r is the total number of available forecasts or "realizations") and α_r is defined as

$$\alpha_r = \frac{\tilde{\mathbf{d}}_r^{\mathrm{T}} \tilde{\mathbf{d}}_r - N_{\mathrm{obs}}}{\mathrm{trace}(\tilde{\mathbf{H}} \mathbf{P}_r^e \tilde{\mathbf{H}}^{\mathrm{T}})}, \tag{7}$$

where $\tilde{\mathbf{d}}_r$ is the innovation vector at t_r normalized by the square root of the observation error covariance matrix. Similarly, $\tilde{\mathbf{H}}$ is the observation operator normalized by the square root of the observation error covariance matrix. Here $\mathbf{P}_r^e$ is the ensemble covariance at t_r and N_{obs} is the number of observations.

We slightly modify the definition of the inflation factor given in Eq. (6). A running average of the last 200 values of α is calculated for each realization, and Π_r is obtained directly from the square root of this running average of α values (previous values of the inflation factor Π_{r-1} were not used in the calculation). The reason for this modification can be seen in Eq. (7). When the dimension of the observation space N_{obs} is small it is possible to have innovation vectors that are so small that the numerator of Eq. (7) becomes negative. The running average allows us to avoid square roots of negative numbers, but the memory in the running average can result in a long string of $\sqrt{\alpha_r} > 1$, which, if multiplied by past Π_r cause the inflation factor to diverge. In the following study, inflation is only applied to the initial ensemble forecast (not to the transformed forecasts).

Section 2c describes the simple extensions to the ETKF that are necessary to implement the TLEF.

c. The TLEF

The crux of the TLEF is that for linear systems the order of estimation and propagation is arbitrary; a forecast followed by the assimilation of old observations (the TLEF) is equivalent to assimilating the observations and then running the forecast (the traditional approach). This is easily shown by considering an analysis step bracketed by two forecasts where ETKF data assimilation is employed. Define $\mathbf{M}_t$ to be a linear uncertainty propagator, then the traditional approach is given by

$$\mathbf{X}_1^f = \mathbf{M}_1 \mathbf{X}_0^a, \tag{8}$$

$$\mathbf{X}_1^a = (\mathbf{M}_1 \mathbf{X}_0^a) \mathbf{T}_1, \quad \text{and} \tag{9}$$

$$\mathbf{X}_2^f = \mathbf{M}_2 [(\mathbf{M}_1 \mathbf{X}_0^a) \mathbf{T}_1]. \tag{10}$$

Alternatively, it is possible to jump straight from the analysis at time $t = 0$ to the forecast at time $t = 2$ that is conditioned by the observations at time $t = 1$ by first linearly propagating to time $t = 2$ and then applying the impact of the observations at time $t = 1$:

$$\mathbf{X}_2^f = (\mathbf{M}_2 \mathbf{M}_1 \mathbf{X}_0^a) \mathbf{T}_1. \tag{11}$$

This is trivially extended to longer forecast leads and more observation periods. A more detailed explanation is given below. The treatment and nomenclature will appear confusing to the casual reader, but are necessary for those who wish to reproduce our results. If desired, the remainder of the section can be skipped by the casual reader without loss of comprehension.

The time indices were removed from the equations in section 3a for the sake of notational simplicity. A coherent discussion of the TLEF requires that the time indices now be reintroduced in an unusual fashion. Define $\mathbf{X}^f(\tau_1, \tau_2, \tau_3)$ to be the normalized ensemble forecast perturbations from a forecast that was originally launched at $t = \tau_1$, that were last transformed at $t = \tau_2$, and are being assessed at $t = \tau_3$. For example, $\mathbf{X}^f(-6, 0, 6)$ would be the normalized ensemble forecasts verified at $t = 6$ h that were launched at $t = -6$ h, but were transformed by the observations at $t = 0$ h. An analyzed state will always have $\tau_1 = \tau_2 = \tau_3$; for example, $\bar{\mathbf{x}}^a(0, 0, 0)$. Using this nomenclature an analysis state can also be expressed in terms of a forecast state as $\bar{\mathbf{x}}^a(0, 0, 0) = \bar{\mathbf{x}}^f(-\tau_{\mathrm{obs}}, 0, 0)$.

To understand how the ETKF framework can update future states, consider Eq. (3) written in analysis increment form:

$$\bar{\mathbf{x}}^a(0, 0, 0) - \bar{\mathbf{x}}^f(-1, -1, 0)$$

$$= \mathbf{X}^f(-1, -1, 0) [\mathbf{C} \boldsymbol{\Gamma}^{1/2} (\boldsymbol{\Gamma} + \mathbf{I})^{-1} \mathbf{E}^{\mathrm{T}}]$$

$$\times [\mathbf{R}^{-1/2} \mathbf{y} - \mathbf{R}^{-1/2} \mathbf{H} \bar{\mathbf{x}}^f(-1, -1, 0)], \tag{12}$$

where the analysis time has been set to $t = 0$, and observation times have been normalized to integer values. The analysis increment gives the perturbation that must be added to the first-guess state in order to obtain the analyzed state. Assuming the analysis increment is small, the impact of the increment at some future time can be obtained by linearly propagating the analysis increment forward in time. Define $\mathbf{M}(\tau_1, \tau_3)$ to be the

linear propagator that takes a perturbation at $t = \tau_1$ and propagates it forward to $t = \tau_3$ (in this example $\tau_3 = 4$). Then

$$\bar{\mathbf{x}}^f(-1, 0, 4) - \bar{\mathbf{x}}^f(-1, -1, 4)$$
$$= \mathbf{M}(0, 4)[\bar{\mathbf{x}}^a(0, 0, 0) - \bar{\mathbf{x}}^f(-1, -1, 0)], \qquad (13)$$

or equivalently, Eqs. (12) and (13) can be combined to obtain

$$\bar{\mathbf{x}}^f(-1, 0, 4) - \bar{\mathbf{x}}^f(-1, -1, 4)$$
$$= \mathbf{M}(0, 4)\{\mathbf{X}^f(-1, -1, 0)[\mathbf{C}\boldsymbol{\Gamma}^{1/2}(\boldsymbol{\Gamma} + \mathbf{I})^{-1}\mathbf{E}^{\mathrm{T}}]$$
$$\times [\mathbf{R}^{-1/2}\mathbf{y} - \mathbf{R}^{-1/2}\mathbf{H}\bar{\mathbf{x}}^f(-1, -1, 0)]\}. \qquad (14)$$

If one assumes that the ensemble subspace captures all dynamically important directions, and that forecast ensemble perturbations evolve linearly, then

$$\mathbf{X}^f(-1, -1, 4) = \mathbf{M}(0, 4)\mathbf{X}^f(-1, -1, 0). \qquad (15)$$

In words, the linearly propagated perturbations are nothing more than the original forecast perturbations produced by the ensemble forecasting system. Equation (15) can be substituted into Eq. (14) to give

$$\bar{\mathbf{x}}^f(-1, 0, 4) - \bar{\mathbf{x}}^f(-1, -1, 4)$$
$$= \mathbf{X}^f(-1, -1, 4)[\mathbf{C}\boldsymbol{\Gamma}^{1/2}(\boldsymbol{\Gamma} + \mathbf{I})^{-1}\mathbf{E}^{\mathrm{T}}]$$
$$\times [\mathbf{R}^{-1/2}\mathbf{y} - \mathbf{R}^{-1/2}\mathbf{H}\bar{\mathbf{x}}^f(-1, -1, 0)]. \qquad (16)$$

Thus, mean forecast corrections can be made simply by replacing the $\mathbf{X}^f(-1, -1, 0)$ of Eq. (12) with the normalized forecast perturbations at the desired forecast time, $t = 4$ in this example.

A similar argument can be made for the ensemble perturbations. If $\mathbf{X}^a(0, 0, 0) = \mathbf{X}^f(-1, -1, 0)\mathbf{T}(0)$, and if we assume

$$\mathbf{X}^f(-1, 0, 4) = \mathbf{M}(0, 4)\mathbf{X}^a(0, 0, 0)$$
$$= \mathbf{M}(0, 4)\mathbf{X}^f(-1, -1, 0)\mathbf{T}(0), \qquad (17)$$

then $\mathbf{X}^f(-1, 0, 4) = \mathbf{X}^f(-1, -1, -4)\mathbf{T}(0)$. Here $\mathbf{T}(0)$ indicates that the transformation matrix was calculated at $t = 0$. *All that is necessary to propagate the influence of the new observations forward in time over the forecast period is to transform the normalized forecast ensemble perturbations with the* same *transformation matrix that was used at the observation time.*

Continuing with this example, the transformed forecast ensemble perturbations at $t = 4$ can be transformed again once the next set of observations become available at $t = 1$: $\mathbf{X}^f(-1, 1, 4) = \mathbf{X}^f(-1, 0, 4)\mathbf{T}(1)$. This process can be repeated until $\tau_2 = \tau_3 [\mathbf{X}^f(-1, 4, 4)]$ or until the assumptions that go into the transformation (linear-

ity, validity of ensemble subspace, etc.) are violated. The reader is reminded that the implications of this treatment are that *if* the system is linear and *if* the data assimilation scheme employed is the ETKF, then the TLEF as presented above provides no new information. The TLEF aims to provide value by exploiting the atmosphere's nonlinearity and the current disconnect between operational data assimilation and ensemble construction.

In an effort to clarify the different time indices, in the narrative of the text we will refer to *lags* and *leads*. The lag will be defined as the number of observation times between the time at which a forecast was launched and the time at which the last observations were incorporated into the forecast: lag $= \tau_2 - \tau_1$. An untransformed ensemble forecast launched from the most recent observations would be a lag-0 forecast.[3] The lead will be defined as the number of observation times between the time at which the last observations were incorporated and the forecast time of interest: lead $= \tau_3 - \tau_2$. For example, $\mathbf{X}^f(-1, 1, 4)$ is a lag-2, lead-3 forecast: it has been transformed twice, and we are interested in a forecast time 3 observation times from the most recent observation.

Incorporating observations into old ensemble forecasts using the TLEF machinery allows for increased forecast ensemble size. Following the above examples where the most recent observation was at time $t = 1$, the ensemble forecast at $t = 4$ can be made up of the ensembles given by the most recent update $\mathbf{X}^f(l, 1, 4)$ (lag 0, lead 3), and a collection of transformed ensemble forecast perturbations [$\mathbf{X}^f(0, 1, 4)$ (lag 1, lead 3), $\mathbf{X}^f(-1, 1, 4)$ (lag 2, lead 3), and any additional lags that one wishes to include]. A simplified cartoon of this process is shown in Fig. 1. In Fig. 1a an ensemble forecast (represented by dark circles), based on observations (denoted by the small white box) is generated by an ensemble prediction system (EPS), and is launched from $t = -1$ to $t = 4$. The ensemble perturbations at $t = 4$ are described by $\mathbf{X}^f(-1, -1, 4)$. In Fig. 1b a new observation becomes available at $t = 0$ and a transformation of $\mathbf{X}^f(-1, -1, 0)$ perturbations is performed. In addition, a new ensemble is produced by the EPS at $t = 0$ and propagated forward to $t = 4$. The $\mathbf{X}^f(0, 0, 4)$ forecast perturbations at $t = 4$ are given by the dark triangles, and the transformed $\mathbf{X}^f(-1, 0, 4) = \mathbf{X}^f(-1, -1, 4)\mathbf{T}(0)$ ensemble perturbations at $t = 4$ are given by the new position of the dark circles. Notice that the

---

[3] This is not strictly true. The lag-0 forecast ensemble perturbations are multiplied by an inflation factor in a calibration aimed at reducing ensemble variance deficiencies arising from limited sample sizes. Details are given in section 2b.
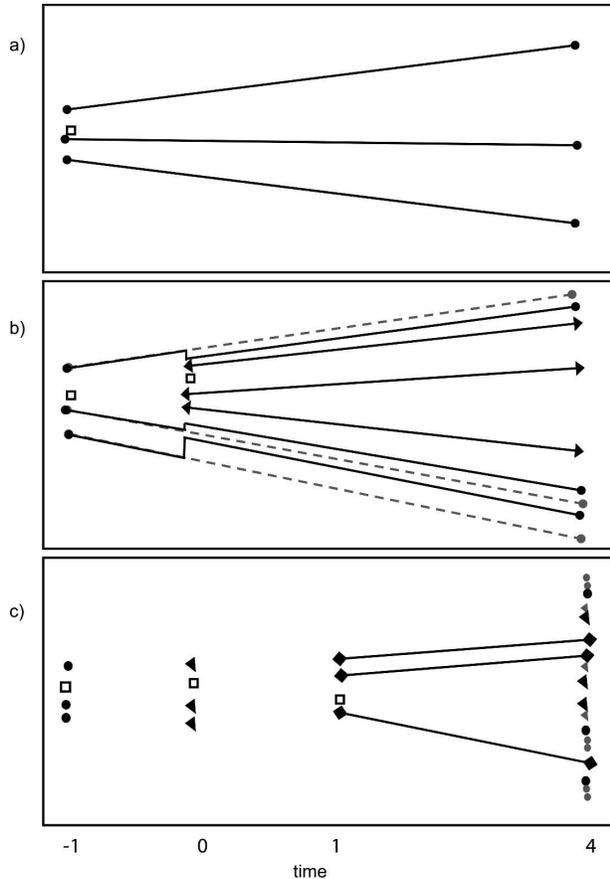
FIG. 1. Schematic of TLEF methodology. The white boxes represent observations. At verification time $t = 4$, gray symbols indicate pretransformed ensembles and the dark symbols represent ensembles updated with the same observation. Dashed lines indicate redundant forecast trajectories. Using notation introduced in the text (a) circles denote $\mathbf{X}^f(-1, -1, 4)$; (b) circles denote $\mathbf{X}^f(-1, 0, 4)$ and triangles denote $\mathbf{X}^f(0, 0, 4)$; (c) circles denote $\mathbf{X}^f(-1, 1, 4)$, triangles denote $\mathbf{X}^f(0, 1, 4)$, and diamonds denote $\mathbf{X}^f(1, 1, 4)$.

original forecasts launched from $t = -1$ are now shown as gray circles to denote they have been superseded by the updated forecasts. Another set of new observations become available at $t = 1$ in Fig. 1c. A transformation of both $\mathbf{X}^f(-1, 0, 1)$ and $\mathbf{X}^f(0, 0, 1)$ takes place,[4] and a new ensemble is generated at $t = 1$ by the EPS. The ensemble is propagated forward and the $\mathbf{X}^f(l, 1, 4)$ forecast perturbations at $t = 4$ are given by the dark

squares, the singly transformed $\mathbf{X}^f(0, 1, 4) = \mathbf{X}^f(0, 0, 4)\mathbf{T}(1)$ forecast perturbations are given by the dark triangles, and the doubly transformed forecast perturbations,

$$\mathbf{X}^f(-1, 1, 4) = \mathbf{X}^f(-1, 0, 4)\mathbf{T}(1) = \mathbf{X}^f(-1, -1, 4)\mathbf{T}(0)\mathbf{T}(1),$$
(18)

are given by the dark circles. In Fig. 1c, the previous "forecast trajectory" lines have been omitted for clarity and only the new forecast trajectory and the transformed states remain. The forecast states prior to transformation are always shown as gray symbols. So in this example, the size of the lead-3 ensemble forecast valid at $t = 4$ has been tripled, as shown by the collection of dark circles, triangles, and squares in Fig. 1c.

It is important to point out that the transformed perturbations do not rely upon the transformed mean [as seen in Eq. (4)]. This implies that the perturbations need not be added back onto the transformed mean at the desired forecast time. Instead, they can be added to the mean of the untransformed ensemble (lag 0) launched from the most recent observations. The implementation of the ETKF considered here is unable to account for the spurious long-distance correlations that can arise from small sample sizes,[5] and Wang and Bishop (2003) suggest that these incorrect correlations can make obtaining a good transformed mean more difficult than obtaining good transformed ensemble perturbations. In this work we choose to use the latest lag-0 forecasts to center the transformed ensemble perturbations; we do not explore the impact of using the transformed mean. Also note that the method does not require that the ETKF be used to construct the initial ensembles, it only assumes that the ensembles accurately reflect the underlying forecast uncertainty.

3. Application of the TLEF to an idealized model

a. Experimental setup

To demonstrate the capabilities of the TLEF, a series of experiments are carried out using the Lorenz 96 system (L96; Lorenz 1996). Well documented in Lorenz and Emanuel (1998) and Hansen and Smith (2000), the equations are

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F, \quad i = 1, \ldots, n,$$
(19)

---

[4] One can choose to construct a $\mathbf{T}$ from the most recent lag-0 forecast and apply it to all previously transformed ensembles, or construct a new $\mathbf{T}$ for each transformed ensemble using their ensemble perturbations. In the limit of all assumptions being satisfied the two are equivalent. In this work, a $\mathbf{T}$ is computed for each lagged ensemble. Experiments were run using only the lag-0 $\mathbf{T}$ and results were qualitatively unchanged.

[5] The ETKF can be modified to include localization, but complicating issues associated with propagating localization information forward in the forecast ensemble subspace motivated neglecting such an implementation.

where for the current study $n = 100$, and $F = 8$. The equations represent a forced dissipative system containing a nonlinear advection-like term, a damping term, and an external forcing. A fourth-order Runge–Kutta integration scheme is employed with a fixed time step of 0.05. This corresponds to approximately 6 h in the atmosphere based on error-doubling times. The boundary conditions are cyclic, so if one insists upon an atmospheric analog, the variables can be thought of as representing some atmospheric quantity distributed along a latitude circle. The important aspects of the system are that it is a chaotic, spatially extended system with a reasonable number of degrees of freedom.

The results reported in this work are for the case where a perfect model of the system exists: the model equations are identical to the system equations. The impact of model inadequacies will be explored in a companion paper. Although the TLEF is based on the ETKF framework, it does not require that the ensemble forecasts be generated using ETKF data assimilation. In the majority of experiments described below, data assimilation and ensemble generation is done using the EnKF (Evensen 1994) with $N_{obs} = 100$ and $N_{ens} = 1024$. A subset of these 1024 ensemble members and a reduced number of observations are utilized by the TLEF. The motivation for this experimental setup was to provide ensemble forecasts that are as close to random draws from the correct forecast PDF as possible, and to eliminate a potential source of misinterpretation. Assessment performed on the large EnKF ensembles (not shown) demonstrate that this aim was achieved. The sensitivity of results to the quality of the ensemble forecasts is explored and discussed below. The ensemble members used in the TLEF experiments are randomly chosen from the large EnKF analysis ensemble each iteration. Experiments were repeated where the same subsample of ensemble members were always used, and there was no qualitative impact on the results.

While not strictly necessary, the TLEF is best suited for situations where the ensemble size N_{ens} is less than the number of observations N_{obs}. In the experiments described below, the TLEF ensemble size varies from $N_{ens} = 10$ to $N_{ens} = 80$, and the number of observations varies from $N_{obs} = 20$ to $N_{obs} = 90$. The baseline configuration sets $N_{ens} = 80$ and $N_{obs} = 90$. Observations were generated by randomly perturbing the true system state and are taken every six model hours. The observational error is assumed to be Gaussian and uncorrelated in both space and time. While these assumptions are unlikely to be satisfied in NWP applications, they allow us to eliminate the impact of correlated observational errors and simplify the interpretation of results.

The standard deviation of the observational noise was set to $\epsilon = 0.2$, which is roughly 3% of the range of the state variables. All statistics are based on the average of 10 independent experiments, each using 3300 successive observation times.

The implementation of the TLEF used in this work does not support localization, but it does support covariance inflation. As described in section 2b, the inflation factor is applied *only* to the lag-0 forecast perturbations (not to the transformed forecast perturbations), and is applied prior to any transformations. This means that lag-0 forecast perturbations are calibrated on the fly and any variance deficiency associated with ensemble size will be reduced. This removes any advantage the TLEF transformed ensembles might have that is simply due to calibration.

The final aspects of the experimental setup are to define the number of lagged ensembles to be included in a combined (or "augmented") ensemble, and the distance into the future that transformations are performed (the definitions of lag and lead are given in section 3c). Note that the lead is defined with respect to the most recent observation. The ensemble forecast with lag 4 (launched one model day in the past) and lead 4 (verifying one model day in the future) will have been subjected to four transformations from the four observations encountered over the lag period; the lag gives the number of transformations performed on the ensemble, and the lead gives the distance into the future that transformations have been performed. For all of the experiments described below, there are 4 lags (one model day) and 24 leads (six model days) available at any time.

b. Ensemble mean error and ensemble spread

To demonstrate the characteristics of the TLEF, Figs. 2a–c displays a plot of the average root-mean-square (RMS) distance of the *perturbations* from the ensemble mean (here referred to as the average ensemble spread) as a function of lag. This method of ensemble evaluation gives an indication of the effect of the transformations on the variance of the ensemble. Negative *x*-axis values in Fig. 2a indicate times before the most recent observation (observations are available every six model hours over the negative interval) and positive values along the *x* axis represent times beyond the most recent observation.

Figure 2a shows a plot of average ensemble spread of the lag 0 and *transformed* perturbations for all available lags and leads up to six model days. For this combination of ensemble size, observational uncertainty level, and number of observations considered for the baseline experiment, the transformed ensembles do not degrade
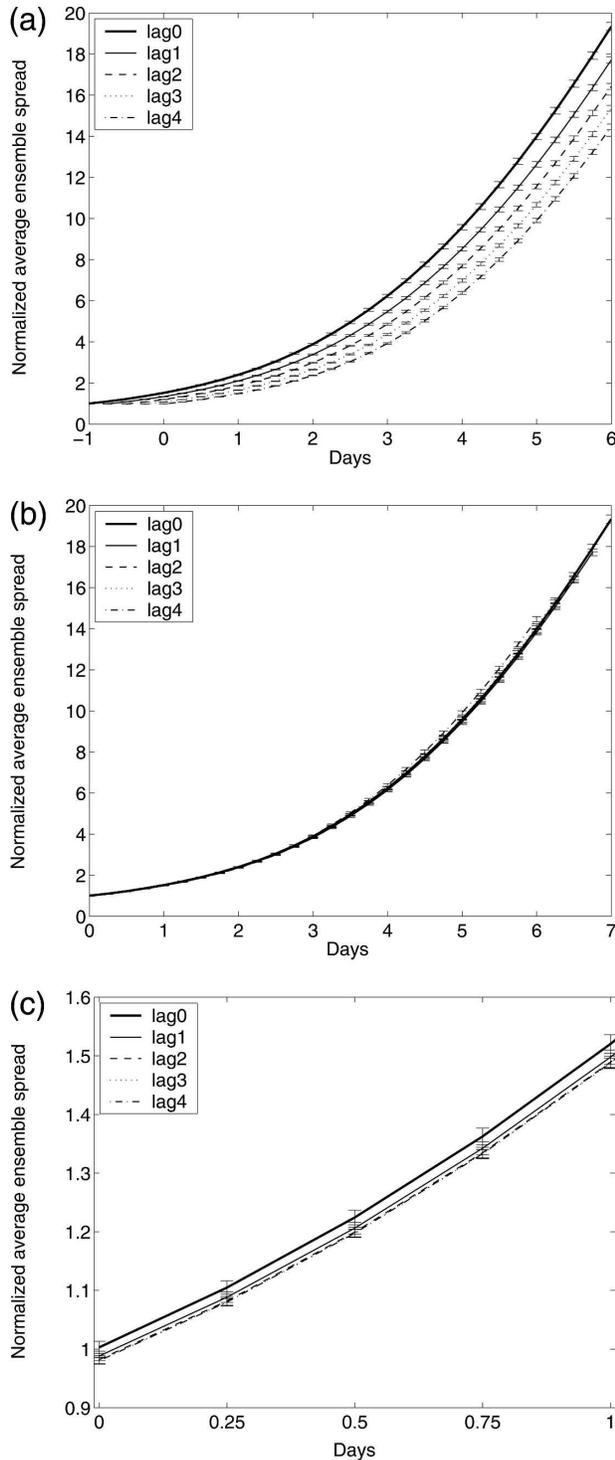
FIG. 2. (a) Average spread of ensemble perturbations; (b) average spread of ensemble perturbations (with respect to common observation time); (c) same plot as in (b), but enlarged to display 0–1-day lead time (*x* axis). Note that the spread of each of the lagged ensembles is almost identical. Error bars represent standard deviation values from 10 experiments.

significantly before 6 days. The thick black solid curve in Fig. 2a is for lag 0, and gives the spread for the untransformed (initial) ensemble perturbations. Remember that the lag-0 forecast perturbations have been calibrated using the state-dependent inflation factor. As a result the lag-0 spread curve is consistent with the average RMSE of the ensemble mean forecasts (not shown). The thin solid curve gives the same information for lag-1 perturbations; this is the average spread of the ensemble that results from applying the TLEF to the lag-0 ensemble perturbations once an observation becomes available. The legend in Fig. 2a indicates the line types for lag 0–lag 4 where the same process of successive transformations continues. Note that the lag-4 forecasts have undergone four transformations and that each 6-h transformation slightly reduces the average ensemble spread of the new forecast perturbations. The *y* axis has been normalized with respect to the lag-0 value at $t = 0$ to aid comparison and error bars reflecting the standard deviation of the RMS values over 10 experiments have been added to all curves. Figure 2b shows the same information as Fig. 2a but plotted relative to the time of the most recent observation seen by the ensembles. Therefore, all curves are lined up as if launched from the same time. Figure 2c shows the information from Fig. 2b in more detail between 0 and 1 day. Note that the rate of increase in spread of the transformed ensemble curves is almost identical to the rate of increase of spread in the lag-0 curve in Fig. 2c. It is also clear that 1-day forecasts exhibit relatively linear error characteristics. Assuming this behavior is characteristic of the TLEF in this configuration, 1-day forecasts are assessed for clarity and ease of comparison in the rest of this study.

It is important to emphasize that one should not expect the TLEF to produce ensemble perturbations that are an improvement upon the calibrated lag-0 ensemble perturbations.[6] This implies that while it would be unprofitable use a transformed ensemble in isolation (unless it is in the period between the time that new observations have arrived and the time that the latest ensemble forecast is completed), it can be used to augment the most recent untransformed ensemble forecast. Using this approach, forecast ensemble sizes can be increased by combining ensemble forecasts that have been corrected by all available observational information; the untransformed and all transformed en-

---

[6] A pathological counterexample is the case where the lag-0 ensemble is constructed without regard to dynamics, observational density, or level of observational uncertainty. In such a case the ensemble data assimilation roots of the TLEF would result in improved ensembles relative to the lag 0.

TABLE 1. Average second-moment statistics (over 10 experiments), where second-moment statistics are averaged over 100 components in each. All means are close to 0 indicating correctly sampled distributions. Standard deviations above 1.0 indicate slightly overdispersive ensembles. The lag 0 is closest to 1.0 and the augmented ensemble ($N_{ens} = 400$) is closer to 1.0 than any of the lagged forecasts.

	Lag 0	Lag 1	Lag 2	Lag 3	Lag 4	Augmented	Uncertainty
Mean	0.004	0.004	0.004	0.004	0.004	0.004	±0.007
Std dev	1.01	1.02	1.02	1.03	1.03	1.02	±0.02

sembles are draws from the same distribution, $p[\mathbf{x}(t)|\mathbf{Y}(0)]$.

The transformed ensemble perturbations are added onto the most recent lag-0 forecast mean to produce a new ensemble forecast. This newly transformed forecast can then be combined with the lag-0 forecast to produce an ensemble with a larger number of members. As all ensembles have been updated using the same number of observations, it does not matter when the original perturbations were "launched" or how many ensembles are combined. As all perturbations are centered around the most recent forecast, it is clear that an ensemble augmented in this way would have an ensemble mean forecast error identical to the lag-0 ensemble mean forecast error. This negates the most practical benefit of increased ensemble size (improved ensemble mean), but benefits can still be gained for techniques such as targeting (where a large ensemble would reduce spurious long-distance correlations) and regional ensemble forecasting (where larger ensembles of boundary conditions can be generated).

To assess whether the error variances forecast by an augmented forecast ensemble are correct, ensemble mean forecast error (with respect to the true system state) for a single component is divided by that component's associated ensemble standard deviation.[7] If the ensembles have correct second moment statistics, then the distribution of this ratio will have a mean of 0 and a standard deviation of 1. Means different than 0 indicate a bias, standard deviations different than 1 imply that the ensemble variance is systematically over- or underestimated. The covariances are not considered here. Results for the baseline case for the lag 0 (80 members), transformed (lag 1, lag 2, lag 3, and lag 4 with 80 members each), and augmented (400 member) 1-day forecasts are given in Table 1 and represent an average of second-order moments over all components.

The mean values of the distributions (displayed in the first row) are indistinguishable from 0, indicating a relatively unbiased ensemble. Standard deviation val-

ues greater than 1.0 imply that the ensembles are all slightly underdispersive. The mean standard deviation values in Table 1 indicate slightly underdispersive forecast ensembles, although the associated uncertainty indicates that the standard deviations are indistinguishable from 1.0. Note that in the mean values of the standard deviations, the lag-0 value is always closest to 1.0 and the value increases with increasing number of lags (lag 1, lag 2, lag 3, and lag 4). These values reflect the characteristics of the ensemble spread curves in Fig. 2c. When all ensembles are augmented into a 400-member ensemble (the lag-0 ensemble, combined with all transformed ensembles lag 1–lag 4), the standard deviation value of the ratio is equal or closer to 1 than all the individual 80-member lagged ensembles, but not that of the 80-member lag-0 ensemble. Perhaps one should not be too surprised by these results; after all, the lag-0 forecasts includes an inflation factor to help overcome variance deficiency due to finite ensemble size, and that inflation propagates through the TLEF. The important thing to take away is that the first two moments of the transformed and augmented ensemble forecasts indicate that the TLEF is behaving as expected, and that the ensemble forecasts seem consistent with draws from an approximately correct distribution; the transformation process does not significantly degrade the second-moment statistics of the forecast ensemble relative to the lag-0 results.

c. Rank histograms

Rank histograms were constructed to test whether these second-moment results hold for higher-order moments as well. One-day forecast ensembles for 10 specific components were used in the calculation of rank histograms. These components were spaced equally throughout the 100-component domain to minimize correlation between histogram increments, and the results were qualitatively similar to rank histograms produced for one single component. Truth is used as the verification, and in each case, the ensemble is subsampled to $N_{ens} = 9$ in order to adequately populate the rank histograms and fairly compare ensembles of different size. Rank histograms for each of the 10 compo-

---

[7] The L96 equations are symmetric, making the result independent of the choice of component.

nents and each of the 10 realizations were averaged to produce a single rank histogram for each of the lag-0, transformed, and augmented ensembles. The variability in the 10 independent realizations permitted the production of standard deviation error bars for the population in each bin. Results are shown in Fig. 3. Figure 3a plots the 1-day rank histogram for the lag-0 ensemble. The solid horizontal line indicates the expected value for each bin (based on the number of bins), and the dashed horizontal lines represent plus and minus the expected one standard deviation (based on the number of bins and number of realizations making up the rank histogram; Smith and Hansen 2004). The error bars at the top of each histogram bar indicate the *realized* standard deviations of the distributions over the 10 experiments. Also note that the *y* axis has been enlarged to range between 0.8 and 1.1 so that the structure of the distribution can be more clearly seen. Upon close inspection, the rank histogram for lag-0 ensemble forecasts display end bins that are slightly underpopulated in the mean, although the error bars for all bins are consistent with the expected mean (i.e., for some realizations the end bins are overpopulated); statistically, the lag-0 rank histogram is indistinguishable from a uniform distribution.

Figure 3b plots the 1-day rank histogram from the augmented 400-member ensemble. To ease comparison with the lag-0 ensemble forecast results, this combined ensemble is also subsampled to nine random members so that all rank histograms have the same number of bins. The augmented ensemble's rank histogram appears underdispersive in the mean, but again the error bars indicate that the distribution is indistinguishable from a uniform distribution. Figures 3c–f plots the rank histograms from lag 1 to lag 4, respectively. Each rank histogram illustrates that the transformed ensembles exhibit slightly underdispersive in the mean, but are statistically indistinguishable from a uniform distribution. A rank histogram produced using a 400-member lag-0 ensemble (not shown), was indistinguishable from the collection of rank histograms displayed in Fig. 3. Consistent with the second-moment statistics of section 3b, the transformation process does not significantly degrade the lag-0 performance, but does significantly increase the ensemble size. The rank histogram generated by a 400-member lag-0 forecast is statistically identical to the rank histograms of Fig. 3.

d. Brier skill scores

Our final form of assessment is the Brier score (Wilks 1995). The Brier score measures the mean-square error of probabilistic forecasts and ranges from 0 (for a perfect forecast) to 1 (for an unskillful forecast).

The Brier score is

$$BS = \frac{1}{n} \sum_{k=1}^{n} (y_k - o_k)^2, \qquad (20)$$

where y_k is the forecast probability and o_k is the observation probability such that $o_k = 1$ if the event occurs and $o_k = 0$ if the event does not occur. The index k denotes a numbering of the forecast/event pairs. Therefore the score averages the squared difference between the forecast probabilities and the subsequent binary observations.

For the simple model experiment, values of truth are readily available and so we define an event occurrence as when truth exceeds a climatological threshold value. The climatology is defined as the average value of the truth over all realizations and all components over 3300 realizations. The threshold values are based on the standard deviation (σ) of the climatology and is set at the mean, $\pm\sigma$, $\pm 0.5\sigma$, and $\pm 0.25\sigma$. We consider an event to have occurred when verification exceeds the threshold value. Forecast probabilities are calculated based on the number of ensemble members exceeding each of the threshold values. The Brier score was calculated for each threshold at all model components and then averaged over all components and all realizations.

The Brier skill score (BSS) is used to measure the fractional improvement in the Brier score of an EPS relative to a reference forecast (usually climatology). The BSS is defined as

$$BSS = \frac{BS_{ref} - BS}{BS_{ref}} \times 100, \qquad (21)$$

where BS_{ref} is the Brier score of the reference forecast. This expresses the *percentage improvement* of the Brier score of the new EPS over the reference EPS. We define our reference EPS to be the lag-0 forecast. The BSS will provide a measure of the improvement of the transformed and augmented ensembles over the initial lag-0 ensemble, if any. The BSS was calculated from the average Brier scores (not by averaging all components' BSS results) and 15 independent experiments with 5000 realizations each were run to assess the variability of results. The individually transformed ensemble BSS results were assessed along with augmented ensemble results containing one, two, three, or four transformed ensembles. The mean results over the 15 independent experiments are shown in Table 2, along with the results from a 400-member fully nonlinear ensemble. The results of this large nonlinear ensemble should be compared with the results of the 400-member, augmented ensemble consisting of 4 transformations. Positive values of BSS (improvements over the lag-0 ensemble) are
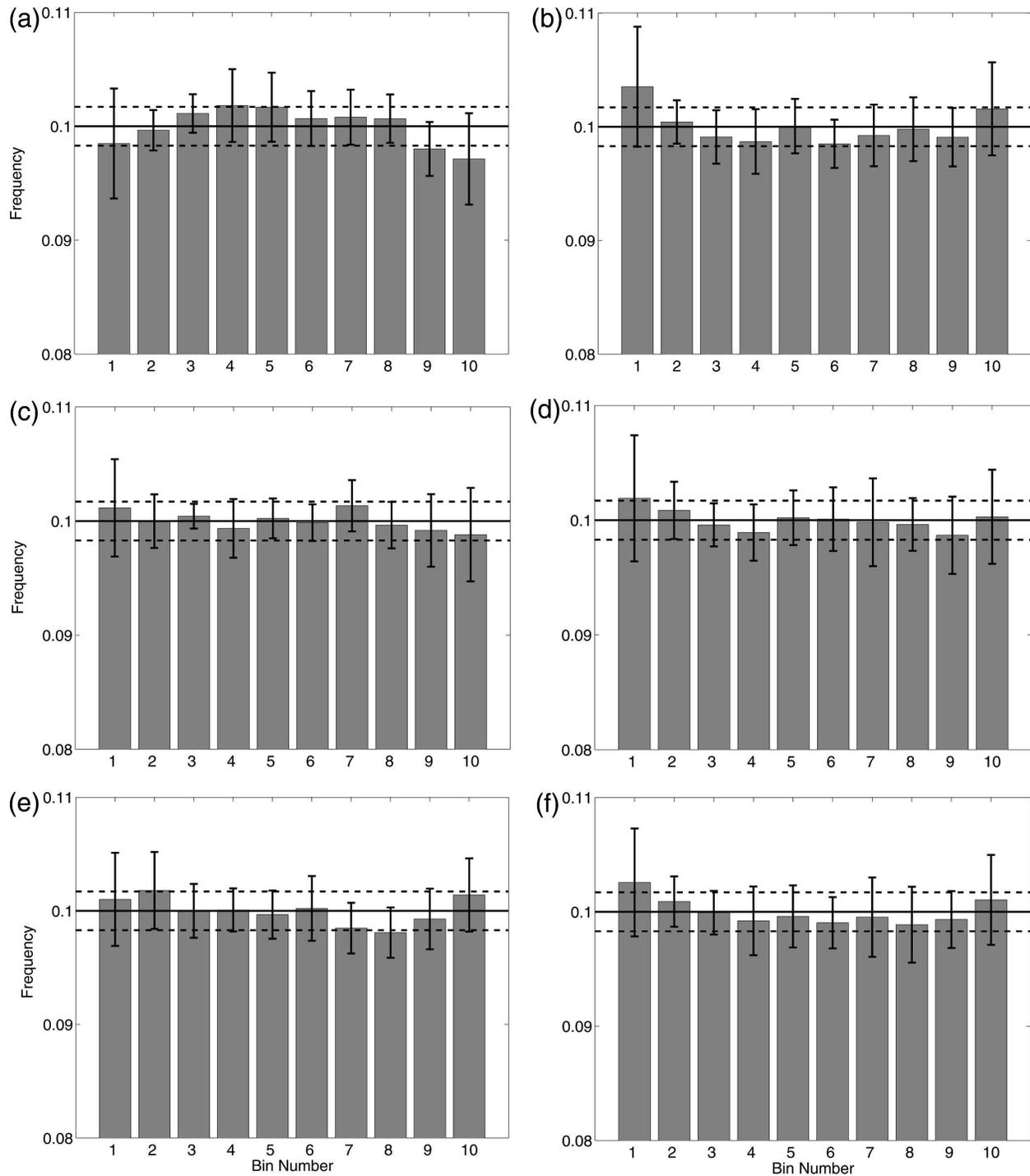
Fig. 3. Rank histograms for subsampled ($N_{ens} = 9$) ensembles (at 10 locations): (a) lag 0 (untransformed) ensemble; (b) augmented (0 + 1 + 2 + 3 + 4) ensemble; (c) lag-1 ensemble; (d) lag-2 ensemble; (e) lag-3 ensemble; and (f) lag-4 ensemble. Error bars represent standard deviation values from 10 experiments. All ensembles display excellent ensemble characteristics (indicating good ensemble distribution about truth).

TABLE 2. BSS (for individual and augmented ensembles with respect to lag 0) over all components. All values are an average over 15 experiments. Analysis of the variability over the 15 experiments indicates that the BSS of the individually transformed ensembles are indistinguishable from zero, while those of the augmented ensembles are robustly positive. Positive values of BSS (improvements over the lag-0 ensemble) are shown in boldface type for clarity. The last column contains BSS from a large nonlinear ensemble that is the same size as the 0, 1, 2, 3, and 4 augmented ensembles. Error bars indicate the statistical similarity between the augmented and the nonlinear ensemble.

	Individual				Augmented				
	Lag 1	Lag 2	Lag 3	Lag 4	0, 1	0, 1, 2	0, 1, 2, 3	0, 1, 2, 3, 4	Nonlinear 400 member
$+\sigma$	−0.01	−0.02	**0.13**	**0.05**	**0.40**	**0.54**	**0.63**	**0.67** ± 0.32	**0.86** ± 0.37
$+0.5\sigma$	**0.05**	**0.012**	**0.05**	**0.049**	**0.36**	**0.50**	**0.58**	**0.64** ± 0.27	**0.90** ± 0.43
$+0.25\sigma$	−0.02	−0.04	**0.15**	**0.01**	**0.40**	**0.58**	**0.75**	**0.79** ± 0.40	**0.59** ± 1.0
-0.25σ	**0.04**	**0.01**	−0.06	−0.18	**0.23**	**0.43**	**0.50**	**0.55** ± 0.39	**0.97** ± 1.12
-0.5σ	−0.09	−0.10	**0.21**	**0.01**	**0.40**	**0.57**	**0.67**	**0.74** ± 0.23	**0.85** ± 0.35
$-\sigma$	**0.12**	**0.10**	**0.13**	**0.13**	**0.42**	**0.58**	**0.66**	**0.71** ± 0.19	**0.80** ± 0.25

shown in boldface type for clarity. The values reveal that a large majority of individually transformed ensembles are worse than lag-0 ensembles (i.e., they have negative BSS). However, the BSS of the augmented ensembles are positive in all cases with mean scores ranging between +0.23% and +0.79%. Analysis of the distribution of results reveals that the BSS of the individually transformed ensembles are indistinguishable from zero: they are statistically the same as the lag-0 forecast. The results also show that the BSS of the augmented ensembles are robustly positive, indicating the augmented ensembles are an improvement over the lag-0 ensembles. Error bars were provided for the largest of the augmented ensembles (lag 1–lag 4) and for the 400-member nonlinear ensemble. The BSS of the nonlinear ensemble are systematically larger than the augmented ensemble, but the two are statistically consistent. The augmented ensemble forecast produced by the TLEF provides the nearly the same information as the 400-member lag-0 ensemble forecast, but at a significant reduction in cost. Because the calibration of the lag-0 forecast strives to remove problems associated with sample size, the positive BSS values are indicative of in increase in the resolution component of the Brier score, not the reliability component (Wilks 1995).

The BSS results of the 400-member ensemble indicate that there is little marginal gain in a Brier score sense for increasing the ensemble size for this experiment, and it is very likely that other techniques would produce BSS superior to those produced by the TLEF (Wilks 2002; Kharin and Zwiers 2003). But the benefit of the TLEF is that increased numbers of dynamically consistent ensemble forecasts are made available for use by other ensemble-based techniques.

e. Sensitivity of results

In nonlinear systems, the linearity criteria assumed by the TLEF is violated for various combinations of initial error magnitude, location in state space, and time scales over which the linearity assumption is assumed. To investigate the sensitivity of TLEF within this idealized model framework, we alter some of the defining parameters used in the baseline case, such as ensemble size, number of realizations, number of observations, and method of ensemble generation.

1) ENSEMBLE SIZE

Because one of the aims of the TLEF is to characterize the benefits of increased ensemble size, a number of experiments were performed with subsampled ensemble sizes both smaller and larger than the baseline case. Statistical assessment of a 10-member ensemble was performed with an identical experimental setup, and second-moment statistics revealed that the lag-0 and the subsequent transformed ensembles are underdispersive to a greater degree than the baseline case in section 3b. The ratio standard deviations of a single experiment increased from 1.19 to 1.50 for lag 0 through lag 4, respectively. Subsequent augmentation of all ensembles was consistent with the baseline case where combining the ensembles produced improved statistics over individual ensembles (with exception of the lag-0 ensemble). Rank histograms produced for this example were consistent with these results. BSSs suggest that when any combination of the lagged forecasts are augmented with the most recent lag 0 the ensemble is an improvement over the individual lag-0 ensemble (given by positive BSS). This may suggest given the small ensemble size, the lag-0 ensemble is largely degraded by sampling error and limitations imposed by the TLEF are not as prominent. This may also demonstrate that for these smaller ensemble sizes the effect of the TLEF is greatly amplified.

It has already been shown that the transformation process can make ensemble sizes larger without a significant degradation of statistics. This was be verified in

section 3b by comparing the statistics of the augmented 400-member ensemble (made up of a lag 0 and four 80-member lagged ensembles) with a lag-0, 400-member ensemble. Second-moment statistics of the (calibrated) lag-0, 400-member ensemble produced a ratio mean of 0.005 and a ratio standard deviation of 1.05. The increased computational effort of running a 400-member ensemble does not provide any noticeable benefit in this case for these measures.

2) OBSERVATIONAL DENSITY AND UNCERTAINTY MAGNITUDE

The ensembles used for the baseline experiment were a random subsample from a large ($N_{\text{ens}} = 1024$) ensemble generated by the EnKF. This experimental setup was chosen to ensure that the ensemble forecasts transformed by the TLEF were a random draw from the correct distribution. Reducing the number of observed components used in the EnKF generation of the analyses and lag-0 forecasts has the same impact on lag-0 forecasts as reducing the ensemble size; the forecast ensemble becomes a worse representation of the correct forecast distribution. Decreases in the observational density also cause the transformed ensembles to become worse than the pretransformed forecasts at shorter lag–lead combinations than in the baseline experiment; the linearity assumption become more tenuous as the size of the initial error increases. The same effect is generated by increasing the magnitude of observational uncertainty. As the quality of the analysis degrades, the TLEF is less and less able to provide quality transformed forecasts.

3) ENSEMBLE CONSTRUCTION METHODOLOGY

Experiments were carried out where the ETKF was employed as the data assimilation scheme, and the TLEF used the same ensemble size and number of observations as the assimilation scheme. Because of the smaller ensemble size ($N_{\text{ens}} = 80$) and the lack of covariance localization in our implementation of the ETKF, the quality of analysis ensembles were degraded, and forecast errors were 1.5–2 times larger than the baseline case. In this case, these increased error levels were not large enough to have a large impact on the quality of the linearity assumption, and the augmented ensemble results were qualitatively similar to the baseline case. The only major difference was that the lag-1 ensemble for lead 0 was identical to the most recent lag-0 ensemble because the TLEF employs the ETKF for its updating. This again highlights the importance of nonlinearity to the successful implementation of the TLEF; nonlinear error evolution results in a di-

vergence of the lag-1 and lag-0 forecasts for greater leads.

To further investigate the sensitivity of results to ensemble construction methodology, ensemble forecasts were also generated using singular vectors. Maintaining an ensemble size of $N_{\text{ens}} = 80$, ensembles were generated by adding perturbations to the EnKF ensemble mean analysis in plus and minus directions of the first 40 singular vectors. Both an isotropic uncertainty norm and an analysis error covariance norm (based on the EnKF analysis ensemble) were utilized. A range of perturbations magnitudes were explored. Because singular vectors do not represent a random draw from the correct analysis distribution, resulting ensemble forecasts do not represent a random draw from the correct forecast distribution. This does not mean that good second-order moments are not obtainable, but it depends upon the combination of initial perturbation size and TLEF boost factor. Initial perturbations sizes that are too small result in transformed ensembles that are variance deficient, while initial perturbation sizes that are too large result in transformed ensembles that have an excess of variance. This strong dependence on exactly how singular vectors are constructed and applied makes us hesitate to draw any general conclusions about what to expect for singular vector ensembles.

4) STATE-DEPENDENT TRANSFORMATION MATRIX

The transformation matrix **T** utilized by the TLEF both rotates and rescales ensemble perturbations in a state-dependent manner. To assess the importance of the state dependence and the rotation, experiments were performed utilizing a static, scalar "transformation." Lead-specific rescaling factors were calculated over a long control run in an effort to produce the best augmented 1-day forecast ensembles possible. By construction, good second-moment statistics were produced by these climatologically rescaled, augmented ensembles. However, the rank histograms and BSSs indicated that the simple rescaling significantly degraded the forecast statistics relative to the lag-0 ensemble forecast. The state-dependent rotation and rescaling performed by the TLEF cannot be efficiently reproduced by using a simple climatological scaling factor.

The sensitivity of the TLEF depends on many factors, and understanding how the results might break down in different situations may help us understand future results obtained using output from operational NWP models.

4. Summary and conclusions

We have presented a technique based on ensemble data assimilation ideas to increase ensemble size with-

out additional model runs. This work was motivated by a desire to sensibly and inexpensively increase forecast ensemble sizes by reusing old ensemble forecasts. We have employed the ETKF machinery to provide state-dependent transformations of old ensemble forecasts using new observations. In the idealized case considered, the resulting augmented (larger) ensemble forecasts did not degrade the statistics of the most recent ensemble forecast, and produced results that were statistically consistent with large lag-0 ensembles. Because transformations are performed in perturbation space, transformed ensembles must be centered about a mean state. While it is theoretically possible to transform the ensemble mean given observations that have become available since the forecast was launched, we chose to center the transformed forecast perturbations about the mean of the most recent ensemble forecasts. The implication is that the TLEF does not alter ensemble mean forecast statistics.

We applied the TLEF to the Lorenz 96 system of equations using a baseline case of $n = 100$, $N_{ens} = 80$, and $N_{obs} = 90$. Ensemble inflation factors were determined using a version of the approach described in Wang et al. (2004) modified to account for the small size of our observation space. Transformed and augmented ensembles were assessed by examining ensemble spread, ensemble second-moment statistics, rank histograms, and BSSs. We found that the spread, second-moment statistics, and rank histograms of the transformed and augmented ensembles were statistically indistinguishable from the lag-0 ensemble forecasts. The BSSs of the augmented ensembles indicated an improvement over the lag-0 ensemble forecasts due to an increase in resolution (ensemble size), not reliability. Extensive sensitivity studies were carried out that highlighted the dependence of the TLEF on the linearity assumption of perturbation error growth and on the relevance of the ensemble subspace.

This work has introduced the formulation of the TLEF and provided proof of concept for the ability of the TLEF to increase ensemble sizes (without having to increase the number of model integrations). Although the method presented here does not represent the definitive solution to small sample sizes, preliminary results of experiments using output from NWP ensemble prediction systems are promising. Results from ensembles from a single NWP EPS are broadly consistent with the results shown here, and application to multi-model NWP ensembles is under way.

REFERENCES

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

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

Buizza, R., 1997: Potential forecast skill of ensemble prediction, and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.,* **125,** 99–119.

Ebisuzaki, W., and E. Kalnay, 1991: Ensemble experiments with a new lagged average forecasting scheme. WMO research activities in atmospheric and oceanic modelling, WMO Rep. 15, Geneva, Switzerland, 6.31–6.32.

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.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

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

Hoffmann, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35A,** 100–118.

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

Jazwinski, A. H., 1969: Adaptive filtering. *Automatica,* **5,** 475–485.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability.* Cambridge University Press, 341 pp.

Kharin, V., and F. W. Zwiers, 2003: Improved seasonal probability forecasts. *J. Climate,* **16,** 1684–1701.

Lorenz, E. N., 1996: Predictability-a problem partly solved. *Proc. Seminar on Predictability,* Vol. 1, Reading, United Kingdom, ECMWF, 1–18.

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

Majumdar, S., C. Bishop, R. Buizza, and R. Gelaro, 2002: A comparison of ensemble transform Kalman filter targeting guidance with ECMWF and NRL total energy singular vector guidance. *Quart. J. Roy. Meteor. Soc.,* **128,** 1–24.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1992: Ensemble prediction. *Proc. Seminar on Validation of Models over Europe,* Reading, United Kingdom, ECMWF, 21–66.

Smith, L., and J. Hansen, 2004: Extending the limits of forecast verification with the minimum spanning tree. *Mon. Wea. Rev.,* **132,** 1522–1528.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and

J. S. Whitaker, 2003: Ensemble square-root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.,* **60,** 1140–1158.

——, ——, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.,* **132,** 1590–1605.

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

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 467 pp.

——, 2002: Smoothing forecast ensembles with fitted probability distributions. *Quart. J. Roy. Meteor. Soc.,* **128,** 2821–2836.