On the Orthogonalization of Bred Vectors

Article *in* Weather and Forecasting · August 2010

DOI: 10.1175/2010WAF2222334.1 · Source: OAI

4 authors, including:

Jan Keller
Deutscher Wetterdienst
34 PUBLICATIONS **607** CITATIONS

SEE PROFILE

Andreas Hense
University of Bonn
226 PUBLICATIONS **5,290** CITATIONS

SEE PROFILE

Andreas Rhodin
University of Hamburg
31 PUBLICATIONS **3,564** CITATIONS

SEE PROFILE

On the Orthogonalization of Bred Vectors

JAN D. KELLER AND ANDREAS HENSE

Meteorological Institute, University of Bonn, Bonn, Germany

LUIS KORNBLUEH

Max Planck Institute for Meteorology, Hamburg, Germany

ANDREAS RHODIN

German Meteorological Service, Offenbach, Germany

(Manuscript received 15 July 2009, in final form 6 February 2010)

ABSTRACT

The key to the improvement of the quality of ensemble forecasts assessing the inherent flow uncertainties is the choice of the initial ensemble perturbations. To generate such perturbations, the breeding of growing modes approach has been used for the past two decades. Here, the fastest-growing error modes of the initial model state are estimated. However, the resulting bred vectors (BVs) mainly point in the phase space direction of the leading Lyapunov vector and therefore favor one direction of growing errors. To overcome this characteristic and obtain growing modes pointing to Lyapunov vectors different from the leading one, an orthogonalization implemented as a singular value decomposition based on the similarity between the BVs is applied. This transformation is similar to that used in the ensemble transform technique currently in operational use at NCEP but with certain differences in the metric used and in the implementation. In this study, results of this approach using BVs generated in the Ensemble Forecasting System (EFS) based on the global numerical weather prediction model GME of the German Meteorological Service are presented. The gain in forecast performance achieved with the orthogonalized BV initialization is shown by using different probabilistic forecast scores evaluating ensemble reliability, variance, and resolution. For a 3-month period in summer 2007, the results are compared to forecasts generated with simple BV initializations of the same ensemble prediction system as well as operational ensemble forecasts from ECMWF and NCEP. The orthogonalization vastly improves the GME–EFS scores and makes them competitive with the two other centers.

1. Introduction

A basic assumption in modern numerical weather prediction (NWP) is that a forecast of a possible future atmospheric state is to some extent uncertain. The magnitude of this uncertainty strongly increases with forecast lead time and is dependent on the space–time structure of the three-dimensional atmospheric flow. During the last 15 years, ensemble weather prediction has become the dominant forecasting tool on the global scale. The cause for this development is not only the fact that the ensemble mean generally provides a better prediction of a future atmospheric state than a single deterministic forecast, but also that a posteriori probabilistic assessments of the forecasts allow an estimation of the uncertainty of the forecast and communicate it to potential users.

The sources of uncertainties are (i) the lack of knowledge about the true state of the atmosphere at the initialization time and (ii) the inherent deficiencies in the forecast model's description of the atmospheric processes. Hence, ensemble forecasts strongly depend on the initialization of the distinct ensemble members and, therefore, the use of an appropriate ensemble generation method proves crucial to the quality of the probabilistic forecast. Corresponding to the aforementioned reasons for forecast uncertainties, two different ensemble generation techniques have become prevalent in numerical weather prediction:

Corresponding author address: Jan D. Keller, Meteorological Institute, University of Bonn, Auf dem Huegel 20, 53121 Bonn, Germany.
E-mail: jkeller@uni-bonn.de

DOI: 10.1175/2010WAF2222334.1

- the perturbation of the initial atmospheric state, which aims at the errors in the initial state, be it from observational or model errors, and
- the perturbation of parameterized processes (such as turbulence or convection) to account for the uncertainties that arise from the estimation of values originating from the subgrid scale.

A general overview of the different ensemble forecasting techniques is given by, for example, Kalnay (2003).

In this study we will focus on the uncertainties estimated by the former method—the perturbation of the initial conditions. We concentrate on estimating the perturbed initial states $\mathbf{x}'_a(t_0)$, which evolve in the strongest growing error modes. Thus, the perturbed initial states can be described by

$$\mathbf{x}'_a(t_0) = \mathbf{x}_a(t_0) + \mathbf{z}_g(t_0), \quad (1)$$

with $\mathbf{x}_a(t_0)$ being the original analysis state and $\mathbf{z}_g(t_0)$ the perturbations leading to the strongest growing modes of the ensemble forecast.

Mainly, three different techniques have been developed during the past 15–20 years to find such initial perturbations. The *singular vector method* and the *bred vector method* focus on the model errors or, more precisely, on the direct estimation of the strongest growing error modes, while the *perturbed observations method* focuses on the uncertainties in the initial analysis field caused by deficiencies in the observational network, either from the lack of observations in a certain region or the measurement errors. However, the analysis error depends on both the dynamical and observational errors.

This study mainly focuses on the ensemble initialization method called *breeding of growing modes* or the bred vector (BV) technique described by Toth and Kalnay (1993, 1997). As it is quite easy to implement, this method has been used in the Global Ensemble Forecasting System (GEFS) of the National Centers for Environmental Prediction (NCEP). The breeding technique is based on a recursive rescaling of perturbed short-range forecasts in relation to an unperturbed control forecast. Hence, such a system ''breeds'' the fastest error-growing modes of the current model's representation of the atmospheric state.

BV perturbations are closely related to (local) atmospheric Lyapunov vectors (see Trevisan and Legnani 1995), which can be calculated together with the Lyapunov exponents that describe the temporal growth rate of the respective spatial modes. If at least one of the global Lyapunov exponents is positive, the behavior of the system is chaotic; that is, formerly nearby points on the model's attractor will separate into unrelated points (Wolf et al. 1985). The idea of the BV method is to estimate the leading Lyapunov vectors for a complex system such as a forecast model with an acceptable effort, for example, without the explicit calculation of the Tangent Linear Model (TLM; Eckmann and Ruelle 1985).

Through the Lyapunov vectors, the BV method is related to another technique called *singular vectors* initialization (SVs; see Buizza and Palmer 1995), which explicitly uses the TLM. Both methods are used to determine the fastest-growing error modes for the first guess of the analysis state. Buizza and Palmer (1995) point out that if their TLM approach to identifying singular vectors is not conducted for a small optimization time period around a single stationary point, but for a sufficiently long time with the evolving trajectory covering the full attractor, the singular values will have an exponential dependence on the optimization time. This dependence is controlled by the Lyapunov exponents, and the corresponding set of singular vectors is then referred to as local Lyapunov vectors.

Both aforementioned methods are still subject to ongoing improvement. A major modification of the BV implementation at NCEP has recently been made by applying a transformation to the perturbations given by the latest ensemble forecasts replacing the BVs (see Wei et al. 2008). This modification, called *ensemble transform* (ET), aims at the orthogonalization of initial perturbations by constraining them with the analysis variance of the data assimilation (DA) system and the recentering of the perturbations around the analysis by using a simplex transform (ST).

ET was originally formulated by Bishop and Toth (1999) for targeting observations and is adopted for the generation of initial ensemble perturbations in the NCEP ensemble forecasting system as described by Wei et al. (2008). For further details on analysis error covariance and the ET method, the reader is referred to Wang and Bishop (2003) and Wei et al. (2006, 2008).

As the ET method is an extension to the classical BV approach, the perturbations are based on flow-dependent spatial structures, but the recentering of the perturbations (while preserving the analysis covariance) ensures that the initial perturbations have the maximum number of effective degrees of freedom. However, the perturbations are not orthogonal for a finite number of ensemble members.

In this study we present another improvement to the BV approach by generating orthogonal initial perturbations for ensemble forecasting created from BVs. The BVs are generated with an ensemble forecasting system based on the German Meteorological Service's global NWP model—the GME–EFS (described in Keller et al. 2008, hereafter K08).

The experiment setup with simple BV perturbations presented in K08 indicates that the breeding cycle

implemented in the GME–EFS is not able to generate well-developed breeding patterns with the given lead time prior to the beginning of the ensemble forecasts. The spatial structure of the single BVs is very similar and therefore leads to an insufficient ensemble spread. The implemented ad hoc solution is to inflate the initial ensemble perturbations by a constant factor, which results in a reasonable ensemble spread averaged over a 3-month simulation period. As the BVs are significantly enhanced during this time, the initial inflation factor is too large for the subsequent simulations. Though the performance in terms of the ensemble spread score is quite reasonable, the results of a verification analysis with respect to the resolution component of the ensemble forecasts depict a poor level of performance by the simple BV approach.

2. Methodology

In the following sections, we summarize the generation of BVs in K08, describe the implementation of the BV orthogonalization in the GME–EFS, and discuss the theory of ET-based orthogonalization techniques in general. Finally, we give some concluding remarks on the differences between the NCEP ET procedure and our orthogonalization approach.

a. Generation of BVs

The BVs that we want to orthogonalize are generated by the GME–EFS. K08 illustrates how the breeding cycle is implemented and how the BVs are used to perturb the initial state. The GME–EFS makes use of a weighted total energy norm to measure the degree of perturbation and thus to calculate the rescaling factor in each breeding cycle. The weighted total energy norm is given by

$$\|\mathbf{Z}_f\|^2 = w_{\mathbf{u}} \int \frac{1}{2}\mathbf{u}'^2\, dV + w_{\mathbf{v}} \int \frac{1}{2}\mathbf{v}'^2\, dV + w_{\mathbf{T}} \int \frac{1}{2}\frac{c_p}{\overline{\mathbf{T}}}\mathbf{T}'^2\, dV + w_{\mathbf{p}_s}\frac{1}{g}\int \frac{1}{\overline{\mathbf{p}}_s}\mathbf{p}_s'^2\Phi_s\, dF, \tag{2}$$

with $\overline{\mathbf{T}}$ the mean temperature at the corresponding level, $\overline{\mathbf{p}}_s$ the mean surface pressure, and Φ_s the orographic height expressed by the surface geopotential. The weighting factors $w_{\mathbf{u}}$, $w_{\mathbf{v}}$, $w_{\mathbf{T}}$, and $w_{\mathbf{p}_s}$ are calculated as the respective energy contribution at the initial time of the breeding cycle. Each of the four energy contributions equals 0.25 for the initial time, allowing the norm to measure the relative change of the energy perturbation for all four contributions. The symbols $\mathbf{u}'$, $\mathbf{v}'$, $\mathbf{T}'$, and $\mathbf{p}_s'$ denote the u and v wind components, the temperature, and the surface pressure component of the BVs. The weighted total energy norm accounts for the respective perturbation growth of each component relative to its initial value. For further details, the reader is referred to K08.

First, the GME–EFS simulations revealed considerable similarities in the structures of the single BVs (see section 4 for an example). Therefore, the spread in the ensemble simulations initialized with these BVs was far too small and had to be amplified artificially to result in a reasonable spread–error growth relation, giving rise to the idea of orthogonalization, for example, as in Wolf et al. (1985) or Bergemann et al. (2009).

The similarities between the BVs may be a consequence of the fact that BVs are all qualitatively similar to the leading Lyapunov vector (Kalnay et al. 2002). Therefore, all BVs are also similar to each other, representing the strongest growing error mode of the given atmospheric state and forecast model. Kalnay et al. (2002) argue that BVs are never orthogonalized as this would shift them toward the subsequent Lyapunov vectors, thus changing the error growth characteristic to the subsequent smaller one. However, it may be necessary to reproduce not only the strongest growing errors but also different growing error modes to obtain larger ensemble spreads. Therefore, we decided to implement an orthogonalization in the GME–EFS ensemble generation process.

b. Implementation of the SVD-based orthogonalization

In contrast to Wolf et al. (1985), who applied a Gram–Schmidt orthogonalization, our method is based on a singular value decomposition (SVD). The orthogonalization is accomplished by transforming the set of BVs based on an estimation of the BVs' similarities. These similarities are expressed by the matrix of spatial covariance–correlation.

In our approach, we start with the $k \times m$ forecast perturbation matrix:

$$\mathbf{Z}_f = (\mathbf{z}_{f,1}, \mathbf{z}_{f,2}, \ldots, \mathbf{z}_{f,k}), \tag{3}$$

which consists of the k single BVs $\mathbf{z}_{f,i}$ of dimension m as columns. To account for the total energy scalar product, we introduce the $m \times m$ weight matrix $\mathbf{M}$ with

$$M_{ij} = \mu_j \tag{4}$$

for every column i. The single weighting factors μ_j are the coefficients for the transformation of the jth component of a BV to its energy contribution to the total energy of the perturbation as given in the definition of the weighted total energy norm [cf. Eq. (2); that is,

$$\boldsymbol{\mu} = \left(\underbrace{\frac{1}{2}, \ldots, \frac{1}{2}}_{\dim(\mathbf{u})+\dim(\mathbf{v})}, \underbrace{\frac{c_p}{\overline{T}}, \ldots, \frac{c_p}{\overline{T}}}_{\dim(\mathbf{T})}, \underbrace{\frac{1}{\overline{p}_s g}, \ldots, \frac{1}{\overline{p}_s g}}_{\dim(\mathbf{p}_s)} \right)^{\mathrm{T}}. \tag{5}$$

We are seeking an expansion of the $\mathbf{Z}_f$ matrix by a set of as yet unknown basis vectors $\mathbf{e}_i$ with $i = 1, \ldots, k$, which we arrange as columns in the matrix $\mathbf{E}$. We require these basis vectors to form an orthonormal basis under the metric $\mathbf{M}$; that is,

$$\mathbf{Z}_a = \mathbf{E}\mathbf{A} \quad \text{with} \quad \mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{E} = \mathbf{I}_m, \tag{6}$$

where $\mathbf{A}$ is the matrix of the expansion coefficients,

$$\mathbf{A} = \mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{Z}_a. \tag{7}$$

With the definition of this metric, the SVD-based orthogonalization can be seen as an ET; that is, the set of ensemble perturbations is altered by a transformation matrix such that the resulting perturbations are an optimal estimate of the analysis error. This approach is similar to the ET approach described by Bishop and Toth (1999) for adaptive observations.

If the number of basis vectors is smaller than m, the expansion is not exact but leaves a residual $\mathbf{R}$. Minimizing this residual in a least squares sense under the orthogonality constraint leads to the standard formulation for the derivation of empirical orthogonal functions (von Storch and Zwiers 1999) that the basis vectors for the expansion can be computed as eigenvectors of the matrix $\mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}$ under the metric $\mathbf{M}$:

$$\mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{E} = \mathbf{E}\boldsymbol{\Lambda}, \tag{8}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the m positive eigenvalues λ_i with $i = 1, \ldots, m$. In the case where the number of BV k is smaller than their dimension m, the matrix $\mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}$ has the rank k. Then, only k eigenvalues are nonzero, while $m - k$ are equal to zero. To account for this, we instead solve the eigenproblem (8) premultiplied with $\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}$:

$$\begin{aligned} \mathbf{Z}_f^{\mathrm{T}}\mathbf{M}(\mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{E}) &= \mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{E}\boldsymbol{\Lambda} \\ (\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{Z}_f)\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{E} &= \mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{E}\boldsymbol{\Lambda} \\ (\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{Z}_f)\widetilde{\mathbf{E}} &= \widetilde{\mathbf{E}}\boldsymbol{\Lambda}. \end{aligned} \tag{9}$$

The symmetric and positive definite matrix in brackets on the left-hand side is of dimension k and measures the similarity between the BVs under the total energy norm. The resulting diagonal matrix $\boldsymbol{\Lambda}$ has k positive eigenvalues as required. To recover the corresponding k original eigenvectors from the modified eigenvectors $\widetilde{\mathbf{E}}$, we multiply the latter by $\mathbf{Z}_f$, use their definition indicated in (9), and obtain

$$\mathbf{Z}_f\widetilde{\mathbf{E}} = \mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{E} = \mathbf{E}\boldsymbol{\Lambda} \Rightarrow \mathbf{E} = \mathbf{Z}_f\widetilde{\mathbf{E}}\boldsymbol{\Lambda}^{-1},$$

with the convention that only the k nonzero eigenvalues λ_i define the inverse $\boldsymbol{\Lambda}^{-1}$. To obtain the proper normalization, it is better to define

$$\mathbf{E} = \mathbf{Z}_f\widetilde{\mathbf{E}}\boldsymbol{\Lambda}^{-1/2}. \tag{10}$$

Then, it is straightforward to show that the constraint $\mathbf{E}^{\mathrm{T}}\mathbf{M}\mathbf{E} = \mathbf{I}_k$ is fulfilled.

The covariance matrix $\mathbf{Z}_f^{\mathrm{T}}\mathbf{M}\mathbf{Z}_f$ is a quadratic symmetric $k \times k$ matrix whose eigenvalues and eigenvectors can be calculated easily and very quickly. Within the GME–EFS, the procedure is implemented by using the QR matrix factorization of the Lapack package. The orthogonalized BV perturbations $\mathbf{Z}_a$ are now defined by taking the expansion amplitude matrix as the identity matrix $\mathbf{I}_k$:

$$\mathbf{Z}_a = \mathbf{E}\mathbf{A} = \mathbf{Z}_f\widetilde{\mathbf{E}}\boldsymbol{\Lambda}^{-1/2}\mathbf{I}_k. \tag{11}$$

From a linear point of view, our method maintains the wind–mass balance in the initial perturbation patterns, which is a key factor to fast ensemble spread growth because each orthogonalized pattern is a linear combination of balanced flow patterns (namely the original BVs).

c. Theory on ET-based orthogonalization

In the following, we discuss the theoretical background of ET-based orthogonalization techniques. Thereby, we clarify the relationship between our implementation and other approaches (e.g., the operational NCEP ET) by showing that such techniques all originate from the same general orthogonalization approach.

The ET of a set of $i = 1, \ldots, k$ linear independent but otherwise arbitrary perturbation vectors $\mathbf{z}_{f,i}$ of dimension m into the new set $\mathbf{Z}_a$ formed by $\mathbf{z}_{a,i}$ as column vectors can be written as a right multiplication with a yet unknown transform matrix $\mathbf{T}$:

$$\mathbf{Z}_a = \mathbf{Z}_f\mathbf{T}, \tag{12}$$

with $\mathbf{T}$ acting on the matrix $\mathbf{Z}_f$ formed by $\mathbf{z}_{f,i}$ (McLay and Reynolds 2009).

We will now show that a general ET can be written as an optimization problem with respect to an arbitrary metric $\mathbf{W}$. The derivation follows closely that of empirical

orthogonal functions (EOFs), well-known in multivariate statistics (von Storch and Zwiers 1999).

We would like to find a vector $\mathbf{Z}_{a,1}$ from all perturbations $\mathbf{Z}_{f,i}$ that picks up as much spatial variability as possible; that is, the vector should represent that spatial pattern that has, in a quadratic sense, the largest spatial similarity with all of the perturbations $\mathbf{Z}_{f,i}$, subject to a normalization constraint $\mathbf{Z}_{a,1}^{\mathrm{T}}\mathbf{W}\mathbf{Z}_{a,1} = 1$:

$$\frac{1}{k}\sum_{i=1}^{k}\mathbf{Z}_{a,1}^{\mathrm{T}}\mathbf{W}(\mathbf{Z}_{f,i}\mathbf{Z}_{f,i}^{\mathrm{T}})\mathbf{W}\mathbf{Z}_{a,1} + \lambda_1(1 - \mathbf{Z}_{a,1}^{\mathrm{T}}\mathbf{W}\mathbf{Z}_{a,1}) \stackrel{!}{=} \max. \tag{13}$$

The solution is found by solving the eigen problem:

$$\frac{1}{k}\sum_{i=1}^{k}(\mathbf{Z}_{f,i}\mathbf{Z}_{f,i}^{\mathrm{T}})\mathbf{W}\mathbf{Z}_{a,1} = \lambda_1\mathbf{Z}_{a,1}. \tag{14}$$

The additional vectors $\mathbf{Z}_{a,i;i\geq 2}$ are found with the same ansatz subject to the additional constraints $\mathbf{Z}_{a,i}^{\mathrm{T}}\mathbf{W}\mathbf{Z}_{a,i} = \delta_{k,i'}$, where $\delta_{k,i'}$ is Kronecker's delta (von Storch and Zwiers 1999). Arranging all eigenvectors in the matrix $\mathbf{Z}_a$ and all eigenvalues in the diagonal matrix $\mathbf{\Lambda}$, we have to solve the eigen problem:

$$\frac{1}{k}\sum_{i=1}^{k}(\mathbf{Z}_{f,i}\mathbf{Z}_{f,i}^{\mathrm{T}})\mathbf{W}\mathbf{Z}_a = \frac{1}{k}\mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a = \mathbf{Z}_a\mathbf{\Lambda}. \tag{15}$$

The matrix $(1/k)\mathbf{Z}_f\mathbf{Z}_f^{\mathrm{T}}$ is not of full rank because $k \ll m$. Instead, one has to solve the eigenvalue problem by left multiplying with $\mathbf{Z}_f^{\mathrm{T}}\mathbf{W}$:

$$\frac{1}{k}(\mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_f)\underbrace{\mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a}_{=\widetilde{\mathbf{Z}}_a} = \mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a\mathbf{\Lambda}. \tag{16}$$

This assures that only k nonzero eigenvalues and the corresponding eigenvectors $\widetilde{\mathbf{Z}}_a$ are determined.

The derivation of the desired transform matrix $\mathbf{T}$ is the formal inversion of the equation

$$\widetilde{\mathbf{Z}}_a = \mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a. \tag{17}$$

This is an underdetermined equation, which can be solved by an optimization, again requiring $(\frac{1}{2})\mathbf{Z}_a^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a$ to be small, subject to the constraint $\widetilde{\mathbf{Z}}_a - \mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a = 0$:

$$\frac{1}{2}\mathbf{Z}_a^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a + \boldsymbol{\gamma}^{\mathrm{T}}(\widetilde{\mathbf{Z}}_a - \mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a) \stackrel{!}{=} \min. \tag{18}$$

The resulting Euler–Lagrange equations read

$$\mathbf{Z}_a - \mathbf{Z}_f\boldsymbol{\gamma} = 0 \quad \text{and} \tag{19}$$

$$\widetilde{\mathbf{Z}}_a - \mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_a = 0. \tag{20}$$

The solution for the matrix of the Lagrange multiplier reads

$$\mathbf{Z}_f^{\mathrm{T}}\mathbf{W}\mathbf{Z}_f\boldsymbol{\gamma} = \widetilde{\mathbf{Z}}_a. \tag{21}$$

With the help of the eigen equation (16), it is easy to show that

$$\boldsymbol{\gamma} = \frac{1}{k}\widetilde{\mathbf{Z}}_a\mathbf{\Lambda}^{-1} \tag{22}$$

is a solution for Eq. (21). Inserting this into the first Euler–Lagrange equation results in

$$Z_a = \frac{1}{k}\mathbf{Z}_f\widetilde{\mathbf{Z}}_a\mathbf{\Lambda}^{-1}. \tag{23}$$

This is exactly the form as our orthogonalized BV ET or the formulation by Wei et al. (2008) for the current NCEP ET version.

d. Relationship between our approach and the NCEP ET

Now, we show that both methods, our approach and the NCEP ET, originate from the same general formulation for ET-based orthogonalization. However, there are certain differences between the methods that will be described in the following.

First of all, we have learned in the previous section that ET-based orthogonalization procedures differ in the metrics used to set up the eigen problem. Therefore, the well-known ET in use, for example, at NCEP and our method are different specifications of the general approach. Making use of the weighted total energy matrix $\mathbf{W} = \mathbf{M}$ will lead to our implementation while choosing the inverse analysis covariance matrix $\mathbf{W} = \mathbf{P}_a^{-1}$ will result in the NCEP approach. The $\mathbf{P}_a^{-1}$ metric does in general not depend on the inherent flow uncertainties, which may on the one hand be beneficial by taking into account errors from both the first guess and the observations. On the other hand, it could turn out to be a restraint for the perturbation growth as the two sources of error may not need to be correlated in time and space. The resulting perturbations do not necessarily properly represent either source of uncertainty and, therefore, a metric purely based on the errors arising from the dynamics might be advantageous. Therefore, we choose a metric purely based on the inherent flow uncertainties given by the breeding cycle. Anyway, the two implementations are likely to produce different results.

A second difference is in the way the two methods are connected to the breeding cycle. The NCEP ET is carried out at every analysis step, thus replacing the breeding cycle with the ET, whereas, in our approach, the orthogonalization is applied to the BVs just prior to the initialization of the ensemble forecasts, leaving the breeding cycle unchanged.

3. Experiment setup and verification

a. Experiment setup

The basis for this study is given by the experiment described in K08. Therein, ensemble simulations have been performed using the GME–EFS with a simple BV ensemble initialization for a period from 1 June to 31 August 2007 (hereafter referred to as JJA07). Ensemble simulations have been started each day at 0000 UTC and run for 144 h. The breeding cycle was initiated at 0000 UTC 28 May 2007 with a recycling interval of 12 h. The initial perturbations for the ensemble simulations in this study are generated from this set of BVs.

In the current experiment, the BV perturbations are orthogonalized prior to the ensemble initialization (but not within the breeding cycle) with the method described in section 2b. As in K08, 20 ensemble simulations with a forecast lead time of 144 h are performed for JJA07 initiated every day at 0000 UTC. The resolution of all the runs is set to a resolution such that the horizontal grid size is approximately equivalent to 120 km.

As references, two operational ensemble forecasting systems are used for comparison: the ECMWF EPS and the NCEP GEFS. To ensure the comparability of the verification results presented in the following section, data from all three ensemble prediction systems are interpolated onto a regular $1.25° \times 1.25°$ latitude–longitude verification grid. Additionally, the number of ensemble members of the ECMWF EPS ensemble is reduced to 20 ensemble members from the full 50-member set. The reduction was tested using two methods: 1) randomly choosing 20 members and 2) taking the first 20 members of the full set. A comparison of the verification results of the two methods showed only small differences for nearly all verification scores. However, the second method exhibited a significantly better level of performance for the continuous ranked probability score. Therefore, we chose to present the results from the reduced set of EPS ensemble forecasts using the second method. The data for the ECMWF EPS and NCEP GEFS are retrieved from the The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble [TIGGE; information online at http://tigge.ecmwf.int TIGGE; see also WMO (2005), Bougeault (2008), and Raoult and Fuentes (2008)] database.

To simplify the notation, the GME–EFS ensemble forecasts based on the simple BV initialization will be referred to as SBVs and those based on the orthogonalized BV perturbations as OBVs hereafter.

b. Verification methods

To study the impacts of the BV orthogonalization on the quality of the resulting ensemble forecasts, different verification methods are applied to the forecast data. In section 4, several verification techniques will be used that are presented in the following. For a general overview on forecast verification methods, the reader is referred to several examples in the literature (e.g., Stanski et al. 1989; Jolliffe and Stephenson 2003).

Two different parameters are used in our verification analysis to represent the tropospheric upper-air structure, namely the 500-hPa geopotential for the northern extratropics (40°–80°N) and the 200-hPa zonal wind speed for the tropics (30°S–30°N). The parameters will be abbreviated as Φ_{500} and U_{200} hereafter. The verification is conducted by using the corresponding operational ECMWF analyses at the forecast lead times as representations of the true state of the atmosphere. Tests showed that differences in the verification results are negligible when using GME or NCEP analyses.

The first characteristic we want to examine is the reliability of the ensemble forecasts, that is, the ensemble's ability to properly estimate the uncertainty of the forecast itself. Therefore, the ensemble spread score is calculated for Φ_{500} and U_{200}, which represents the ratio of the ensemble spread E_{spd} and the forecast error E_{err}. The score equals 1 for a perfect ensemble spread and is lower (higher) than 1 for an underestimation (overestimation) of the forecast uncertainty. Both contributions are calculated in terms of a mean square as proposed in Palmer et al. (2006);

$$E_{\mathrm{spd}} = \overline{\|\mathbf{X} - \langle \mathbf{X} \rangle\|_{\mathrm{MS}}} \tag{24}$$

is the spread of an ensemble, where $\mathbf{X}$ are the forecasted values of the single ensemble members, $\langle \mathbf{X} \rangle$ is the ensemble mean, and

$$E_{\mathrm{err}} = \overline{|\mathbf{x}_T - \langle \mathbf{X} \rangle|_{\mathrm{MS}}} \tag{25}$$

is the error of the ensemble forecast with $\mathbf{x}_T$ the ''true'' state of the atmosphere (represented by the verifying analysis). The overbar denotes the mean over all grid points under consideration.

In addition, the reliability is examined using analysis rank histograms or Talagrand diagrams (e.g., Talagrand et al. 1998; Toth et al. 2003). Such a diagram is constructed by accumulating the number of events when the

observed value lies within a specific histogram bin. The $k + 1$ bins are defined by the k values of the single ensemble member forecasts sorted in increasing order. Accordingly, the first (last) bin is the number of verifying events when the observation is lower than the smallest (higher than the largest) ensemble member forecast. The histogram bins are filled with values by accumulating over space and/or time.

The next property of the ensembles to be verified is the forecast error variance, which can be explained by the ensemble forecast. The *perturbation versus error correlation analysis* (PECA) is used to examine how well an ensemble can account for the corresponding error of the deterministic forecasts. PECA evaluates the quality of a forecast by calculating the pattern anomaly correlation between a perturbed forecast ($\mathbf{P}$) and the error of the deterministic forecast ($\mathbf{E}$) of the same meteorological center (i.e., GME, ECMWF, and NCEP forecasts). Therefore, the perturbed forecast is constructed as an optimal combination, $\mathbf{P}_{\text{optimal}} = \sum \alpha_i P_i$, of the original perturbed ensemble forecasts $\mathbf{P}_i$. This optimal combination is estimated by finding the set of α_i, which provides the minimum for $|\mathbf{E} - \sum \alpha_i \mathbf{P}_i|$. More details about PECA can be found in Wei and Toth (2003).

Both of the aforementioned verification methods benefit primarily from a larger spread in the ensemble forecasts. This might tempt the forecaster to exaggerate the ensemble spread in order to obtain higher scores. Hence, we conduct an analysis to assess the ensemble resolution, that is, the gain in forecast quality with respect to a reference forecast (climatology). The performance of the OBVs of upper-air fields is therefore tested by using the continuous ranked probability score (CRPS). The CRPS is a continuous extension of the ranked probability score (RPS) defined as the integral of the Brier score over all possible threshold values. Here, the CRPS is calculated following the approach proposed by Hersbach (2000) together with a decomposition of the CRPS into reliability and resolution components. In addition, the CRPS is presented as a skill score that expresses the improvement of the forecast under consideration, in relation to a climatological forecast.

The forecast quality is also assessed by using relative operating characteristics (ROCs). Therefore, the hit rate (HR) and the false-alarm rate (FAR) of a forecast are computed and sorted by ascending FAR. In that way, the HR can be plotted as a function of the FAR that constitutes the ROC curve. A diagonal line in such a diagram would represent the skill of a climatological forecast, whereas curves to the upper left (lower right) would represent a higher (lower) level of forecast skill. Hence, the area under the ROC curve (ROCA) can be used as a measure for the forecast quality. Here, an HR–FAR pair is computed for each ensemble member using several threshold values at each grid point. Then, the ROCA is calculated as the area under the ROC curve minus 0.5 (the area under the diagonal) times 2, which leads to a ROCA value of 1 for a perfect forecast and 0 (and below) for a forecast without skill. For further information on ROC, the reader is referred to Harvey et al. (1992) and Fawcett (2006).

To estimate the uncertainty of the aforementioned verification scores, confidence intervals are calculated using the bootstrap method (Efron and Tibshirani 1991). The idea of the bootstrap technique is to create a new dataset $\mathbf{y}^*$ of the given data $\mathbf{y} = [y_1, y_2, y_3 \ldots, y_n]$ (both of size n), which has the same empirical probability distribution as the original. This leads to a permutated sample drawn from y with replacements, i.e., $\mathbf{y}^* = [y_4, y_9, y_6, y_2, y_9, \ldots]$, to which the statistic (i.e., the verification score) can be applied. This procedure is repeated N_{boot} times for independent bootstrap samples $\mathbf{y}^*$, leading to N_{boot} different verification scores. Then, the confidence intervals can be estimated from this set of scores.

c. Case study: 2007 Midwest flooding

In addition to the verifications based on the overall scores, we present a case study to illustrate the actual improvement of the OBV in relation to the SBV in a specific forecast situation. The case study focuses on the so-called 2007 Midwest flooding (MF07), which was a flood event in the midwestern United States that took place during August 2007. This flood event impacted six states (Illinois, Indiana, Iowa, Minnesota, Ohio, and Wisconsin), claiming 18 lives and causing over $500 million in damage.

The flooding was caused by heavy rainfall from numerous thunderstorms constantly redeveloping along the west–east axis of a warm front that became stationary on 18 August 2007. For the following days, the front extended into the states of Indiana and Ohio while large amounts of moisture originating from Tropical Storm Erin were transported northward over the warm front, thus intensifying the thunderstorm activity along the front.

Rainfall amounts reached record levels in several counties with an average precipitation sum of more than 100 mm for all areas along the front. The highest amount of rainfall was observed in La Crescent, Minnesota, with a total of 462 mm for the whole event. Figure 1 shows a map of the affected region. The contours denote the rainfall sums estimated in the National Weather Service (NWS) Advanced Hydrologic Prediction Service (AHPS) precipitation dataset (information online at http://water.weather.gov/) by combining data from radar and rain gauge observations. The data shown were interpolated to the verification grid and therefore the maximum values

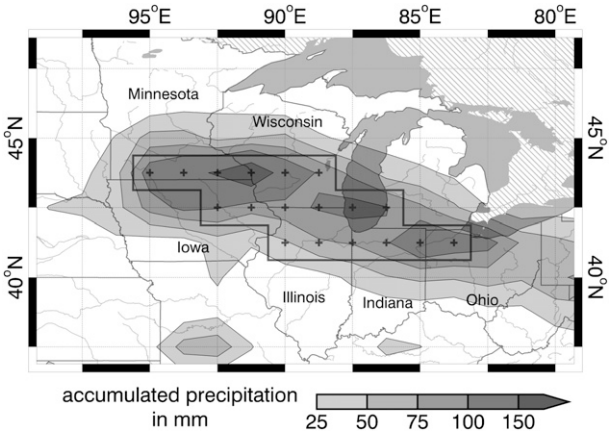

FIG. 1. Map of the region impacted by the MF07 event. Points on the verification grid inside the area struck by heavy rainfall and subsequent flooding are marked by dark crosses and outlined with a dark thick line. In addition, a contour map overlay shows the accumulated precipitation from 1200 UTC 18 Aug to 1200 UTC 22 Aug 2007 estimated from radar and rain gauge data and interpolated to the verification grid. Canadian territories are shaded, as precipitation data is only provided within the boundaries of the United States.

are lower (maximum value of 191 mm compared to 382 mm in the original AHPS dataset).

Weather forecasts of a horizontal resolution of approximately 120 km, at which we are looking, have limited ability to represent the convective processes that led to the large amounts of rainfall in MF07. In addition, the precipitation forecast values of the forecast model have to be interpreted as area means. Hence, the observed precipitation data are interpolated to the verification grid, thus representing comparable area-mean values. One has to keep in mind that MF07 was an extreme event, one that normally falls into the outermost upper tail of a probabilistic forecast. Therefore, we define a new measure for forecast errors using ensemble quantiles for further evaluation. Let

$$\delta^{(p)}(i) = |P_f^{(p)}(i) - P_o(i)| \tag{26}$$

be the absolute deviation of the precipitation forecast $P_f^{(p)}$ of ensemble quantile p from the observation P_o of the AHPS dataset. Both values are interpolated to the verification grid at a given grid point i. Then,

$$\Delta^{(p)} = \sqrt{\frac{1}{N_g}\sum_i [\delta^{(p)}(i)]^2} \tag{27}$$

is the RMS of the forecast error given the ensemble forecast quantile p for all N_g grid points under consideration.

To test the performance of the GME–EFS simulations in the medium range and to account for the time scale given by the MF07 event, ensemble forecasts leading to SBV as well as OBV are constructed as follows:

- Accumulated precipitation is calculated as the difference of the 108- and 84-h precipitation forecasts, denoting the 24-h rainfall sum at 1200 UTC on the fourth forecast day.
- These 24-h sums are accumulated for forecasts initiated from 0000 UTC 15 August to 0000 UTC 18 August 2007, thus covering a period from 1200 UTC 18 August to 1200 UTC 22 August 2007.

The forecasted precipitation sums are used to calculate the ensemble spread, representing forecast uncertainty, and $\delta^{(p)}(i)$ and $\Delta^{(p)}(i)$, to account for the forecast error.

4. Results

In the following, results from the OBV are presented and compared to the SBV experiment described in K08 as well as the aforementioned operational ensembles. The scores and methods are described in section 3b.

A first look at the differences between the initializations of OBV and SBV is given in the Figs. 2 and 3. Figures 2 and 3 show maps of the northern Atlantic and European region, with gray contours representing the positive or negative 850-hPa temperature anomalies of the perturbed analysis from the control analysis for 1 July 2007. Anomalies in Fig. 2 are calculated with four of the simple BVs, while the anomalies in Fig. 3 originate from the orthogonalized set of BVs.

When comparing the initial perturbations for SBV, it is obvious that most patterns are similar in all of the analyses; for example, between Iceland and Scandinavia all of the analyses show a large-scale positive anomaly (light gray shading). The perturbation patterns for OBV show much more variability in the horizontal structure; for example, through the orthogonalization the aforementioned large-scale positive anomaly is altered in shape or replaced by negative anomalies (dark gray shading) or small-scale variations. However, not all features are changed similarly: the negative anomaly over southern Scandinavia, for example, remains in all of the orthogonalized BVs. Hence, the similar structures in the BVs leading to small ensemble spread growth are altered. Some features in the perturbation patterns may however remain unchanged.

In the following, we investigate if the induced changes in the BV structure also enhance the performance of the ensemble forecasts.

a. Ensemble reliability

To examine the performance of the ensembles regarding ensemble reliability, the ensemble spread score

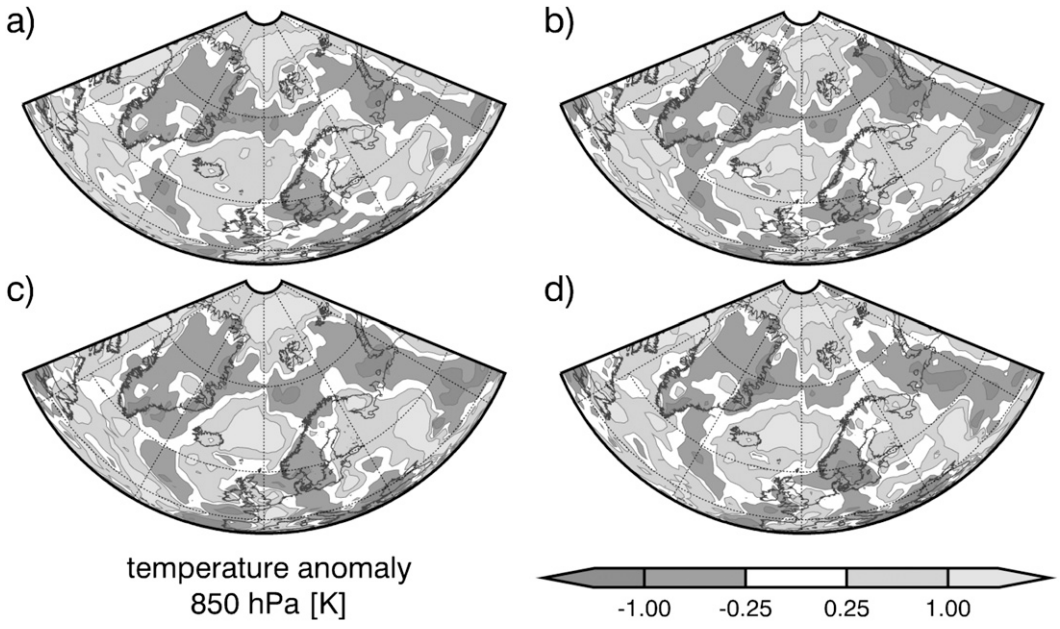

FIG. 2. Structure of the 850-hPa temperature anomaly determined by the difference of (a)–(d) four perturbed analyses from the control analysis at 0000 UTC 1 Jul 2007. The perturbations are taken from the breeding cycle. Patterns are shown for the northern Atlantic and Europe.

is calculated for Φ_{500}. The results for the score as a function of the forecast lead time are shown in Fig. 4. Comparing OBV and SBV, a significantly higher ensemble spread score is clearly visible as a consequence of the orthogonalization procedure. Although the uncertainty is heavily overestimated in the beginning (maybe as a result of exaggerated ensemble spread), the OBV simulations perform very well in the medium range (3–6 days) with an ensemble spread score near one.

In comparison to the operational ensembles, the ensemble spread score reveals a better level of performance for the OBV against the NCEP GEFS, which

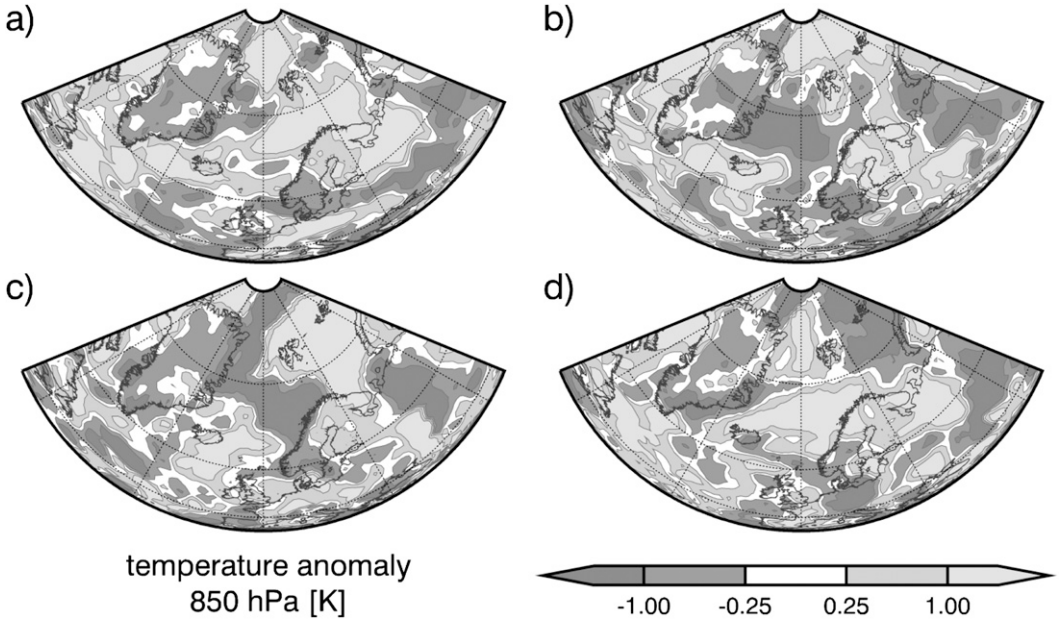

FIG. 3. As in Fig. 2, but from the control analysis at 0000 UTC 1 Jul 2007. The perturbations are taken from the orthogonalized BV set.

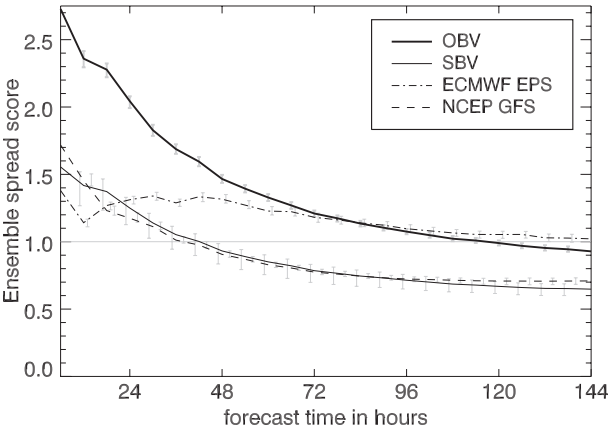

FIG. 4. Ensemble spread score calculated with 500-hPa geopotential in the Northern Hemisphere extratropics (40°–80°N) over JJA07 vs forecast lead time. The vertical lines denote the 95% confidence intervals.

drops below 1 after about 30 h of simulation. The ECMWF EPS performs best as the ensemble spread score stays on a level above 1 for the whole forecast lead time.

The ensemble reliability regarding Φ_{500} is further assessed by using Talagrand diagrams shown in Fig. 5. Here, the diagrams are constructed for forecast lead times from 6 to 48, 54 to 96, and 102 to 144 h. The results for the OBV (Fig. 5a) indicate an overestimation of forecast uncertainty for the first 48 h. The gain in uncertainty estimation quality achieved by the orthogonalized perturbations becomes apparent with progressing forecast time. The ensemble forecast distribution significantly enhances toward a flat histogram representing reasonable spread growth characteristics for orthogonalized BVs. However, a small bias to higher forecast values can still be determined.

The results of the SBV shown in Fig. 5b exhibit a strong U shape, which increases with forecast lead time. Despite the inflation of the initial perturbation, the underestimation of the uncertainty is still quite large even for the first 48 h. When comparing SBV to OBV, the strong beneficial effects of the SVD orthogonalization on the BV structure (i.e. the growing mode characteristics) become obvious.

To compare the GME–EFS results to the operational ensembles, Talagrand diagrams for the ECMWF EPS and the NCEP GEFS are provided in Figs. 5c and 5d, respectively. The results from the ECMWF EPS forecasts exhibit a similarity to the OBV results for lead times greater than 48 h, both in terms of shape and bias to higher forecast values. On the contrary, the NCEP GEFS produces a U shape for longer lead times comparable to that of the SBV, although the uncertainty underestimation is not as strong as that of the SBV.

For the tropical region, the results of the ensemble spread score analysis of U_{200} are given in Fig. 6. Obviously, all ensembles have problems generating sufficient ensemble spread in this region. However, the OBV simulations provide large spread in the beginning, but exhibit a strong decrease in the first 48 h. This may be an indication of the perturbations not being in balance with the atmospheric background, which could be caused by the missing moisture perturbations. The ensemble spread score stabilizes a value of around 0.7 in the medium range and lies within the performance range of the ECMWF ensembles and above the NCEP GEFS simulations. In comparison to the SBV, the OBV simulations provide

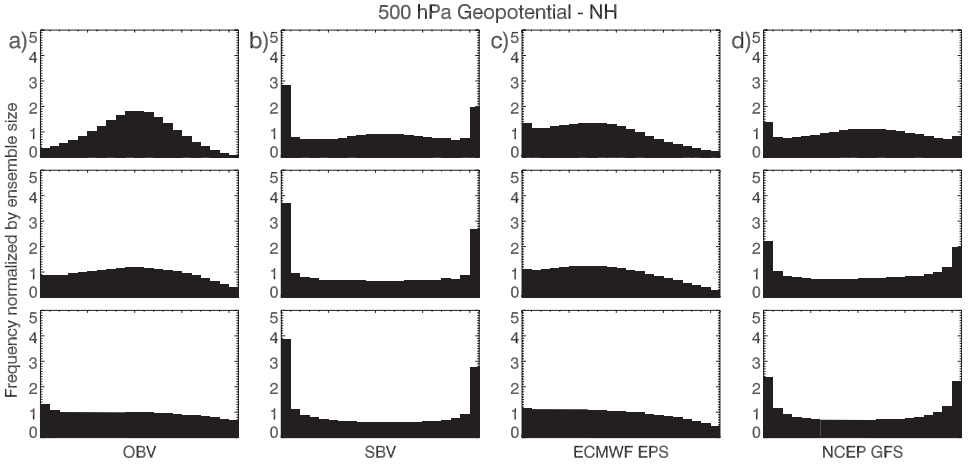

FIG. 5. Talagrand diagrams calculated with 500-hPa geopotential in the Northern Hemisphere extratropics (40°–80°N) over JJA07 for (a) GME with OBV, (b) GME with SBV, (c) ECMWF EPS, and (d) NCEP GEFS. The results for three different forecast time intervals: (top) 6–48, (middle) 54–96, and (bottom) 102–144 h.

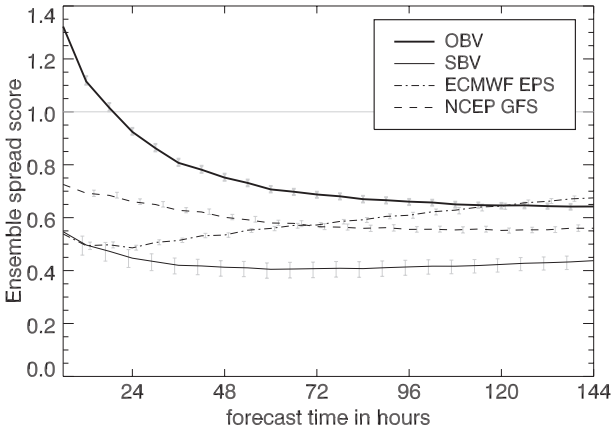

FIG. 6. As in Fig. 4, but for the 200-hPa zonal wind component in the tropical region (30°N–30°S).

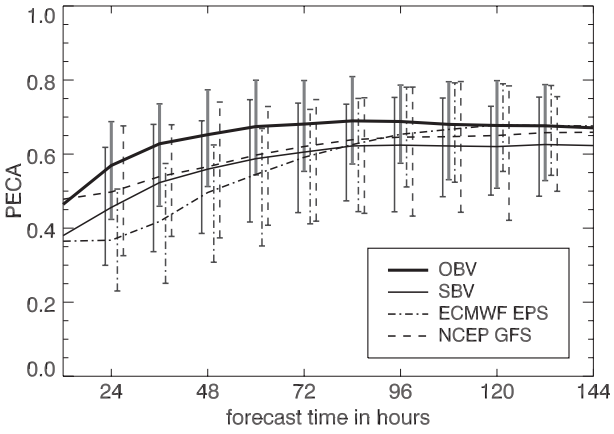

FIG. 7. PECA score of the 500-hPa geopotential calculated for the Northern Hemisphere extratropics (40°–80°N) over JJA07. The vertical lines denote the 95% confidence intervals.

a significantly higher ensemble spread score for all forecast lead times.

b. Explained variance

To evaluate the ability of the ensemble to represent the deterministic forecast error, PECA scores are computed for the upper-air fields. The results for Φ_{500} can be found in Fig. 7. The OBV and SBV PECA values overlap for forecast lead times shorter than 72 h. Thereafter, the OBV exhibits higher PECA values than SBV with the difference increasing with forecast lead time. This indicates an enhanced ensemble quality due to the orthogonalization of the BVs, which becomes apparent in the medium range.

For short forecast lead times, the optimal perturbation vectors of the OBV and SBV ensembles constructed as linear combinations of the corresponding perturbed forecasts are very similar (not shown). This is a result of the implemented orthogonalization procedure, which works as a linear operator on the BVs, combining them according to their covariance structure to overcome the similarity among them. Therefore, there is nearly no improvement in the forecast error variance explained by the ensembles. However, under the nonlinearities of the model, the OBV forecasts evolve into a larger variety of directions in the subspace of growing error modes spanned by the initial BVs. Hence, in the medium range the OBV ensembles allow for a better explanation of the forecast error variance.

The same illustration is given in Fig. 8 for U_{200}. Here, the OBV clearly outperforms the SBV for all forecast lead times. The different results of the Φ_{500} and U_{200} PECA scores can be explained with different structures of perturbations in the tropics and the midlatitudes. The single BVs are even more alike in the tropics and their amplitudes are quite small. Hence, the orthogonalization of the BVs leads to a significantly larger ensemble spread and therefore it is more likely that the OBV ensemble may cover the true analysis state and future state of the atmosphere than the SBV ensemble.

c. Ensemble resolution

To also account for the forecast quality, the skill score of the CRPS is computed. The results of the Φ_{500} CRPS for all four ensembles are presented in Fig. 9. The SBV simulations exhibit the worst CRPS of the ensembles under consideration. The improvement achieved through the orthogonalization is considerable as the OBV's CRPS is significantly higher than the SBV's CRPS and is positive over the whole forecast lead time. In addition, comparison to the operational ensembles shows that the OBV produces even higher CRPS values than the operational ensembles for the medium range. The good performance

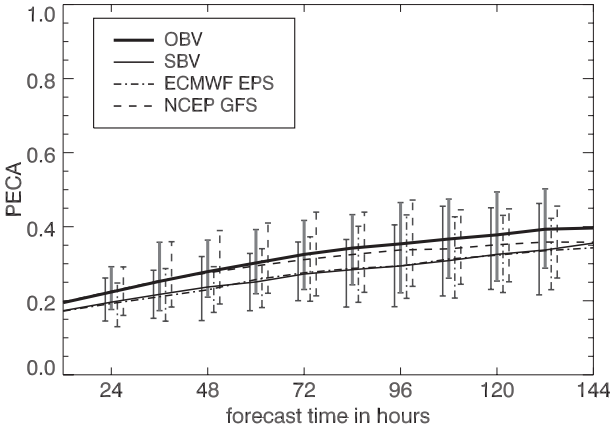

FIG. 8. As in Fig. 7, but for the 200-hPa zonal wind component calculated for the tropical region (30°N–30°S).

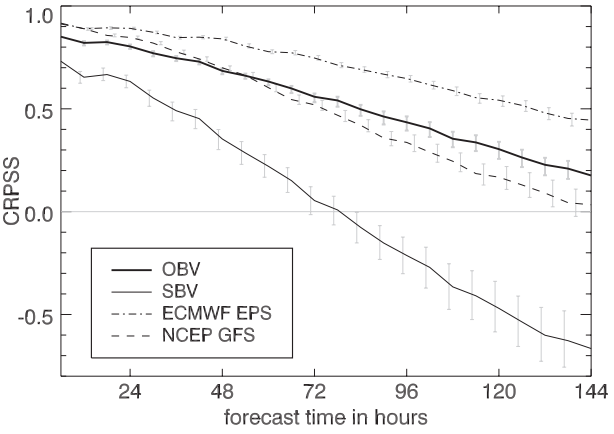

FIG. 9. CRPS score of the 500-hPa geopotential calculated for the Northern Hemisphere extratropics (40°–80°N) over JJA07. The vertical lines denote the 95% confidence intervals.

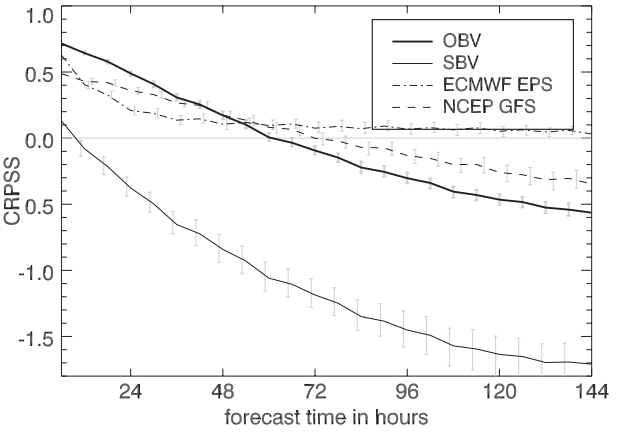

FIG. 10. As in Fig. 9, but for the 200-hPa zonal wind component calculated for the tropical region (30°N–30°S).

of the OBV can also be seen when looking at U_{200}, as presented in Fig. 10. Again, the OBV exhibits a much higher CRPS than the SBV. In comparison to the operational ensembles, the forecast quality of the OBV is slightly worse in the medium range. However, it still provides reasonable forecast quality in terms of CRPS.

With the decomposition of the CRPS, a reliability and a resolution component can be retrieved. The resolution is the difference between the potential CRPS (which in turn is the difference between the CRPS and the reliability component) and the climatological uncertainty (cf. Hersbach 2000). Therefore, the resolution of an ensemble system is positive (negative) if it performs better (worse) than the climatologic probabilistic forecast; that is, larger values of resolution indicate better ensemble performance.

A comparison of OBV and SBV for the CRPS resolution component of Φ_{500} (Fig. 11) shows a significant gain in resolution for the OBV, with the difference between the two ensembles increasing with forecast lead time. Again, the OBV performs better than the NCEP GEFS in the medium range, while the ECMWF EPS is superior to all of the ensembles presented.

For U_{200}, (Fig. 12) the SBV resolution becomes negative for larger forecast lead times, indicating poor performance that is even worse than a climatological forecast. As with the CRPS itself, the NCEP GEFS performs best for the medium range. The OBV ensemble exhibits the highest resolution for the first 2 days, but a stronger decrease with forecast lead time nearing zero for the end of the simulation period. However, the increase in ensemble resolution induced by the orthogonalization is still considerable.

Another analysis regarding forecast resolution is conducted using ROCs. Results are given in Figs. 13 and 14, which show the ROCA as a function of forecast lead time. For Φ_{500} (Fig. 13), the ROCA of both GME ensembles exhibits lower values than the operational ensembles. However, the OBV simulations provide a higher ROCA than the SBV ensembles with the difference between the two slightly increasing after the fourth forecast day.

For the U_{200} ROCA, the values are significantly lower for all ensembles, which is an indication of the poorer performance of the ensemble forecast in the tropics compared to the midlatitudes. While the ECMWF and NCEP ensembles exhibit a nearly constant ROCA for all forecast lead times, the ROCAs of the GME simulations decline with progressing forecast time. Again, the OBV produces higher values than the SBV with the difference being larger than in the midlatitudes and the difference between the two decreasing over forecast lead

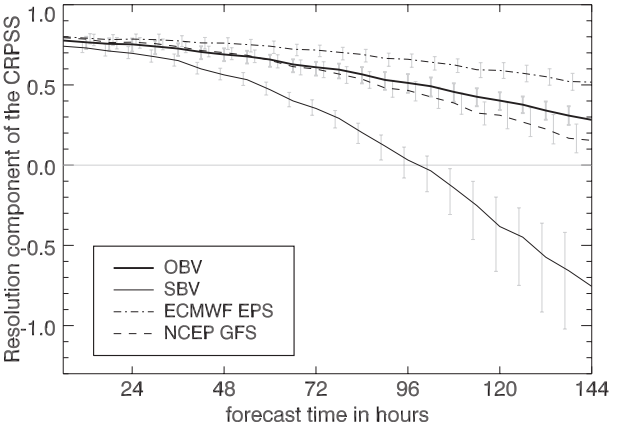

FIG. 11. Resolution component of the CRPS of the 500-hPa geopotential calculated for the Northern Hemisphere extratropics (40°–80°N) over JJA07. The vertical lines denote the 95% confidence intervals.

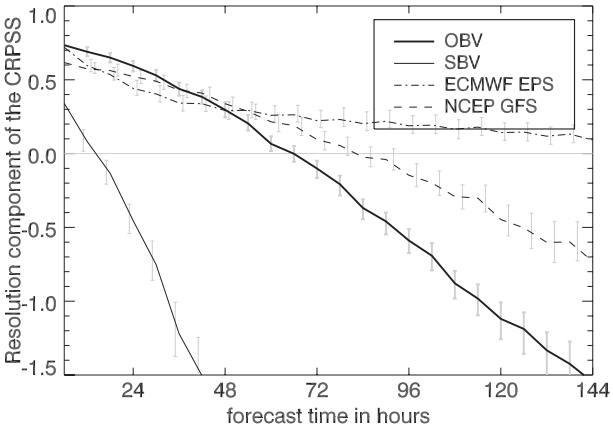

FIG. 12. As in Fig. 11, but for the 200-hPa zonal wind component calculated for the tropical region (30°N–30°S).

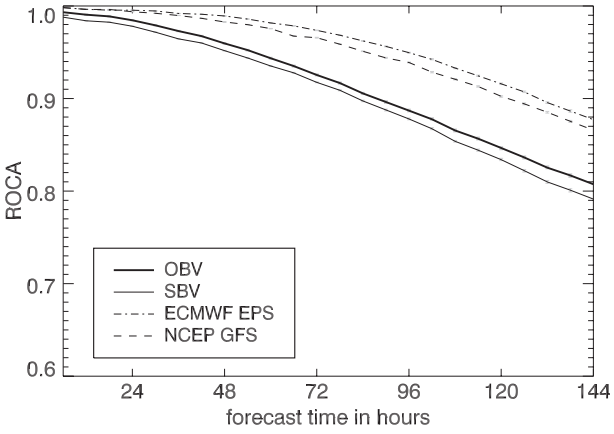

FIG. 13. ROCA vs forecast lead time of the 500-hPa geopotential calculated for the Northern Hemisphere extratropics (40°–80°N) over JJA07. The vertical lines denote the 95% confidence intervals.

time. This points to difficulties of the GME model forecasts for the tropical region that cannot be compensated by the ensemble spread growth induced by the orthogonalized BV initialization.

d. Case study

Figure 15 shows the results of δ^{95} (absolute forecast error of the 95th quantile of the ensemble at each grid point) for SBV (Fig. 15a) and OBV (Fig. 15b) in the region affected by MF07. Both GME–EFS forecasts have a northeastern shift of the stationary warm front leading to the large forecast errors over the Great Lakes and the Canadian border region. However, significant differences between the two ensembles can be found. When concentrating on the region of major flooding (depicted by crosses and outlined with a dark thick line), the errors of the OBV are generally lower than those of the SBV. Especially in the northwestern part of the area (southern Minnesota) where the most significant amount of damage was reported, the OBV forecasts perform better as a result of the larger ensemble spread.

Similar results can be found with the uncertainty forecast expressed through the ensemble spread (Fig. 16). Here, the spread is calculated as the difference between the 0.975th and the 0.025th quantiles, which are estimated by kernel dressing using standard Gaussian distribution functions as the kernel [for further information on kernel dressing, see Broecker and Smith (2008)]. Again, the OBV (Fig. 16b) performs better as the uncertainty estimation for the forecast is higher in the MF07 region than in the SBV simulations (Fig. 16a). The spread also indicates higher uncertainties for southern Minnesota, where the rainfalls reached up to 190 mm (area mean).

The impacts of orthogonalization on the ensemble forecast spread can also be seen in Fig. 17. Here, the ratio of the RMS forecast errors for the SBV and OBV simulations ($\Delta^{(p)}_{\mathrm{SBV}}/\Delta^{(p)}_{\mathrm{OBV}}$), calculated for the region of major flooding, is shown. For quantiles smaller than 0.45, the ratio is positive, indicating larger forecast errors for OBV than for SBV. With higher forecast quantiles, the ratio increases, climbing above 1 beyond the quantile value of 0.45. The poorer performance for the lower quantiles and the significantly better performance for the higher quantiles of the OBV are consequences of the larger ensemble spread. Due to the fact that the probabilistic forecast becomes broader in terms of the probability density function, a larger number of possible developments is covered by the ensemble. Especially in cases of extreme events like MF07, this can be beneficial, for example, in estimating the probabilities of different scenarios (worst-case scenario).

5. Conclusions

The experiment conducted in K08 revealed several shortcomings of the GME–EFS ensemble initialization

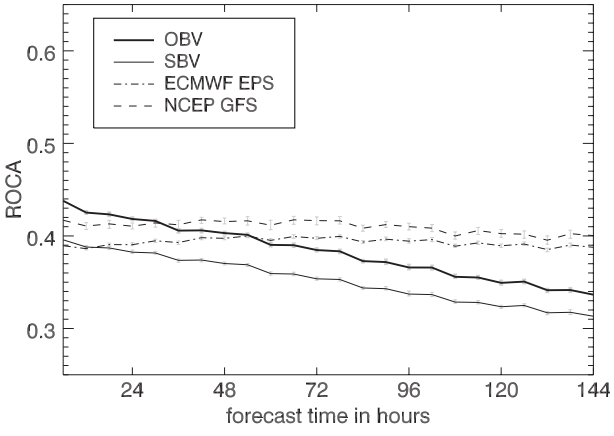

FIG. 14. As in Fig. 13, but for the 200-hPa zonal wind component calculated for the tropical region (30°N–30°S).

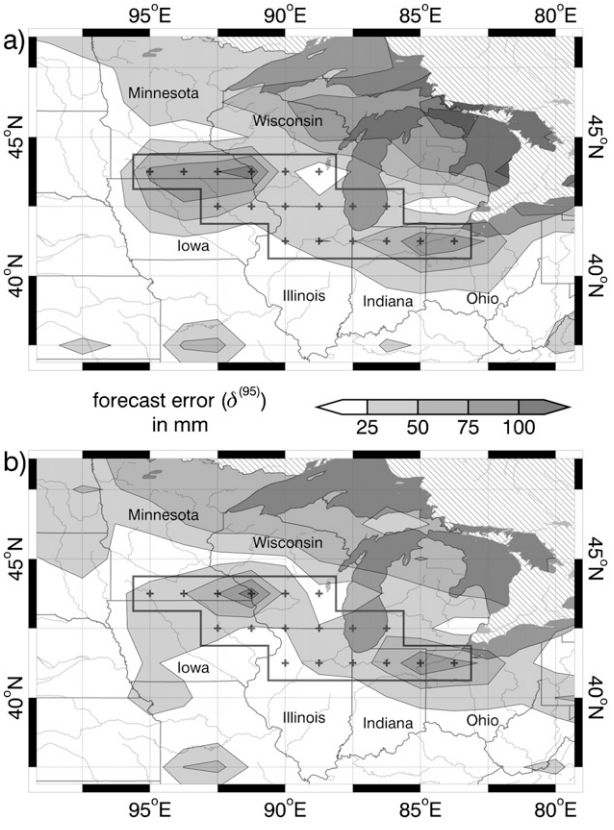

FIG. 15. Map of the absolute forecast error for the 0.95% quantile of the ensemble for (a) SBV and (b) OBV simulations [$\delta^{95}(i)$]. Designation of the main area of flooding is as in Fig. 1.

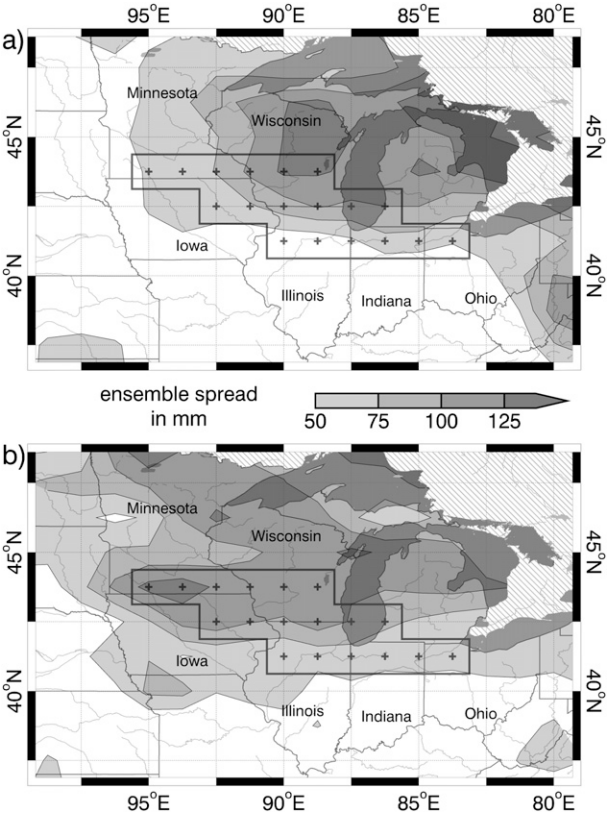

FIG. 16. Map of the forecast uncertainty expressed through ensemble spread for (a) SBV and (b) OBV simulations. The spread is calculated as the difference between the 0.025% and 0.975% quantiles of the ensemble forecast estimated by kernel dressing. Designation of the main area of flooding is as in Fig. 1.

procedure using simple BVs. The major issue is the similarity among the BVs (cf. Bowler 2006), which is a result of the BVs' tendency to point in the direction of the leading Lyapunov vector. As a consequence, an orthogonalization of the BVs leads to an enhanced utilization of the subspace of growing error modes spanned by the BVs. This is because the orthogonalized perturbation patterns are similar to Lyapunov vectors different from the leading one. While Kalnay et al. (2002) state that such an orthogonalization is not needed, as the BVs remain distinct from another due to the different initialization and the nonlinearities of the system, this study shows otherwise. The analysis exhibits significant improvement in forecast skill as an effect of the orthogonalization.

As the spread of the SBV simulations is quite small, its ensemble reliability is poor. Hence, considerable improvement in reliability can be achieved by applying an SVD orthogonalization to the BVs. In the medium range, the OBV even outperforms the NCEP GEFS operational ensemble in terms of ensemble spread score and CRPS for Φ_{500}. Forecast enhancements due to the orthogonalization can also be found for specific weather situations. The MF07 case study provides an example of the impacts of the larger ensemble spread on precipitation forecasts. These forecasts can be an enhancement to the probabilistic assessment of near-future hazardous weather events.

However, the results of the ROCA analysis point to a shortcoming of the orthogonalized BV implementation. The orthogonalized ensemble perturbations are not centered around the unperturbed analysis and, therefore, lead to a shift in the ensemble mean. Although the ensemble itself provides a far better estimation of the forecast uncertainty, the quality of single forecast quantities such as the ensemble mean or median is poorer than in the SBV approach. Implementing a simplex transform (ST) to recenter the orthogonalized perturbations as described in Wei et al. (2008) in the forthcoming version of the GME–EFS is planned.

For the tropical region, the systematic difficulties of the GME model forecasts have become apparent. The higher ensemble spread of the OBV simulations leads to a gain in the explained error variance and better forecast

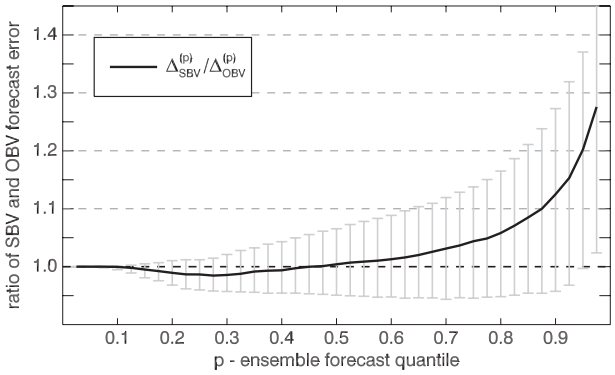

FIG. 17. Ratio of SBV and OBV forecast error ($\Delta^{(p)}_{SBV}/\Delta^{(p)}_{OBV}$) independent of the ensemble forecast quantile p. A value lower (higher) than 1 denotes smaller (larger) forecast error in the SBV in relation to the OBV forecasts. The vertical lines denote the 95% confidence intervals.

quality in terms of ROCA. This is because the GME–EFS (as well as many operational ensembles) has difficulties generating reasonable spread in the tropical region. The increased ensemble spread of the orthogonalized initialization therefore instantaneously increases the quality of the ensemble forecasts, but cannot account for deficiencies in the tropical region forecast of the model itself. Another reason for the poor performance of the GME–EFS compared to the operational ensembles in the tropics may be the lack of moisture perturbations in the ensemble initialization as well as the purely dynamical error representation given by BVs, as the operational ensembles also make use of stochastic parameterizations.

Acknowledgments. The authors want to thank the German Science Foundation (DFG) for funding this study within the framework of the Priority Program 1167, Quantitative Precipitation Forecast, as well as the German Meteorological Service (DWD), the German High Performance Computing Centre for Climate- and Earth System Research (DKRZ), and the Centre for Marine and Atmospheric Sciences (ZMAW) for providing the model, data, and computational resources. The authors also want to thank three anonymous reviewers for their constructive comments as well as Zoltan Toth, Mozheng Wei, and Yuejian Zhu from the National Centers for Environmental Prediction (NCEP) for their useful suggestions regarding the verification methods.

REFERENCES

Bergemann, K., G. Gottwald, and S. Reich, 2009: Ensemble propagation and continuous matrix factorization algorithms. *Quart. J. Roy. Meteor. Soc.,* **135,** 1560–1572.

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

Bougeault, P., 2008: The THORPEX Interactive Grand Global Ensemble (TIGGE): Concept and objectives. *ECMWF Newsletter,* No. 116, 9–10. [Available online at http://www.ecmwf.int.]

Bowler, N. E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus,* **58,** 538–548.

Broecker, J., and L. A. Smith, 2008: From ensemble forecasts to predictive distribution functions. *Tellus,* **60A,** 663–678, doi:10.1111/j.1600-0870.2008.00333.x.

Buizza, R., and T. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.,* **52,** 1434–1456.

Eckmann, J.-P., and D. Ruelle, 1985: Ergodic theory of chaos. *Rev. Mod. Phys.,* **57,** 617–656.

Efron, B., and R. Tibshirani, 1991: Statistical data analysis in the computer age. *Science,* **253,** 390–395, doi:10.1126/science.253.5018.390.

Fawcett, T., 2006: An introduction to ROC analysis. *Pattern Recognit. Lett.,* **27,** 861–874.

Harvey, L., K. Hammond, C. Lusk, and E. Mross, 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev.,* **120,** 863–883.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting,* **15,** 559–570.

Jolliffe, I., and D. B. Stephenson, Eds., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 240 pp.

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

——, M. Corazza, and M. Cai, 2002: Are bred vectors the same as Lyapunov vectors? Preprints, *Symp. on Observations, Data Assimilation, and Probabilistic Prediction,* Orlando, FL, Amer. Meteor. Soc., 6.8. [Available online at http://ams.confex.com/ams/pdfpapers/28833.pdf.]

Keller, J. D., L. Kornblueh, A. Hense, and A. Rhodin, 2008: Towards a GME ensemble forecasting system: Ensemble initialization using the breeding technique. *Meteor. Z.,* **17,** 707–718.

McLay, J., and C. A. Reynolds, 2009: Two alternative implementations of the ensemble-transform (ET) analysis-perturbation scheme: The ET with extended cycling intervals, and the ET without cycling. *Quart. J. Roy. Meteor. Soc.,* **135,** 1200–1213, doi:10.1002/qj.437.

Palmer, T., R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher, and L. Smith, 2006: Ensemble prediction: A pedagogical perspective. *ECMWF Newsletter,* No. 104, 10–17. [Available online at http://www.ecmwf.int.]

Raoult, B., and M. Fuentes, 2008: Implementation of TIGGE Phase I. *ECMWF Newsletter,* No. 116, 10–16. [Available online at http://www.ecmwf.int.]

Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. Forecast Research Division Research Rep. 89-5, Atmospheric Environment Service, Downsview, ON, Canada, 77 pp.

Talagrand, O., R. Vautraud, and B. Strauss, 1998: Evaluation of probabilistic prediction systems. *Proc. Workshop on Predictability,* Reading, United Kingdom, ECMWF, 1–25.

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

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.,* **125,** 3297–3319.

——, O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science,* I. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 137–163.

Trevisan, A., and R. Legnani, 1995: Transient error growth and local predictability: A study in the Lorenz system. *Tellus,* **47A,** 103–117.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.

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

Wei, M., and Z. Toth, 2003: A new measure of ensemble performance: Perturbation versus error correlation analysis (PECA). *Mon. Wea. Rev.,* **131,** 1549–1565.

——, ——, R. Wobus, Y. Zhu, C. Bishop, and X. Wang, 2006: Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus,* **58A,** 1–28.

——, ——, ——, and ——, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus,* **60A,** 62–79.

Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano, 1985: Determining Lyapunov exponents from a time series. *Physica,* **16D,** 285–317.

WMO, 2005: First workshop on the THORPEX Interactive Grand Global Ensemble (TIGGE): Final report. WMO/TD-No. 1273, WWRP/THORPEX No. 5, World Meteorological Organization, 31 pp. [Available online at www.wmo.int/thorpex.]