STOCHASTICALLY PERTURBED BRED VECTORS

BRENT GIGGINS AND GEORG A. GOTTWALD

School of Mathematics and Statistics, University of Sydney, Australia

Abstract. The breeding method is a computationally cheap procedure to generate initial conditions for ensemble forecasting which project onto relevant synoptic growing modes. Ensembles of bred vectors, however, often lack diversity and align with the leading Lyapunov vector, which severely impacts their reliability. In previous work we developed stochastically perturbed bred vectors (SPBVs) and random draw bred vectors (RDBVs) in the context of multi-scale systems. Here we extend this method to systems without scale separation, and examine the performance of the stochastically modified bred vectors in the single scale Lorenz 96 model. In particular, we show that the performance of SPBVs crucially depends on the degree of localisation of the bred vectors. It is found that, contrary to the case of multi-scale systems, localisation is detrimental for applications of SPBVs in systems without scale-separation when generated from assimilated data. In the case of weakly localised bred vectors, however, ensembles of SPBVs constitute a reliable ensemble with improved ensemble forecasting skills compared to classical bred vectors, while still preserving the low computational cost of the breeding method. RDBVs are shown to have superior forecast skill and form a reliable ensemble in weakly localised situations, but for strongly localised bred vectors they do not constitute a reliable ensemble and are over-dispersive.

1. Introduction

The chaotic nature of the atmosphere and the climate system, and its sensitivity to small uncertainties in the initial conditions may render single forecasts meaningless. Probabilistic forecasts, which instead run an ensemble of forecasts, have become standard in numerical weather forecasting, providing a Monte-Carlo estimate of the probability density function \[13, 23, 26\]. Such ensemble forecasts issue the most probable forecast alongside measures of its uncertainty. A key question is how to initialise the ensemble members. There exist several methods to generate such ensembles, using singular vectors \[27, 37\], bred vectors \[47, 48\], analysis ensembles from ensemble Kalman filters \[49, 9\], and more recently analogs \[5\]. In this work we consider bred vectors and the so called ”breeding method” which constitutes a computationally very attractive method to produce an ensemble of initial conditions introduced by \[47, 48\]. In this method initial conditions are generated from finite perturbations, the bred vectors (BVs), which encapsulate information about fast growing modes. Such fast growing initial conditions are then likely to be pre-images of states of high probability. Ensemble-forecasting with bred vectors was successfully implemented for more than a decade since 1992 by the National Centre for Environmental Prediction (NCEP) for their operational 1-15 day ensemble forecasts, and has been widely used in atmosphere and climate probabilistic forecasts such as ENSO prediction \[10, 11\], seasonal-to-interannual forecasting

E-mail address: brent.giggins@sydney.edu.au and georg.gottwald@sydney.edu.au.
in coupled general circulation models (CGCMs) \cite{53} and forecasting Mars’ weather and climate \cite{31,18}.

In the breeding method a control trajectory alongside an ensemble of nearby trajectories is generated. The ensemble members are initialised from perturbed initial conditions with finite perturbation size $\delta$ from the initial condition of the control trajectory. Contrary to Lyapunov vectors, all ensemble members are propagated with the full nonlinear model. The perturbed trajectories are periodically rescaled to a specified finite-size distance $\delta$ away from the control trajectory, to avoid saturation of instabilities. Bred vectors are defined as the difference at the time of rescaling of the perturbed trajectories and the control trajectory. The perturbation size is often thought of as a filter of small scale instabilities, in the sense that BVs are insensitive to very fast growing instabilities which typically are associated with small scale processes and which nonlinearly saturate at an amplitude smaller than $\delta$. Choosing $\delta$ appropriately allows the forecast to be tuned to specific instabilities of interest. For example, BVs project onto baroclinic instabilities when the perturbation size is comparable to $1 - 10\%$ of the natural variability in the atmosphere \cite{48,12}.

Unfortunately, bred vector ensembles lack diversity and most of the ensemble forecast variability may be contained in a single BV \cite{49}. Indeed, for a range of small perturbation sizes $\delta$ bred vectors align with the leading Lyapunov vector and the ensemble collapses to a single member. This reduction in ensemble dimension severely impedes on their ability to reliably sample the forecast probability density function. To preserve the attractive features of BVs such as their low computational cost, several strategies have been devised to increase the ensemble spread in BVs. \cite{3,23} proposed to orthogonalize BVs. \cite{35} employed stochastic backscattering to increase the diversity. \cite{42,41,39} generated BVs employing the geometric rather than the Euclidean norm when rescaling. \cite{6} proposed a different rescaling procedure based on the largest BV. \cite{18} added small random perturbations to the BVs at each rescaling period.

In recent work, \cite{16} proposed a method of stochastically perturbing BVs in the context of ensemble forecasts of the slow dynamics in multi-scale systems to alleviate the problem of small ensemble diversity, introducing stochastically perturbed bred vectors (SPBV). SPBVs were constructed, it was argued, to sample the conditional probability function of the system conditioned on the slow variables by multiplicatively randomising the fast BV components. The localised character of the fast BV components lead, after a fast relaxation towards the attractor, to realisations of the fast variables of the resulting perturbation which sample the probability density function conditioned on the slow synoptic state. The ensemble dimension of SPBVs is significantly increased, in particular for small but finite values of the perturbation size $\delta$. It was shown that the subsynoptic variability of SPBVs allowed to generate the necessary synoptic uncertainty consistent with the uncertainty of the analysis, and the increased diversity leads to much better forecasting skill when compared to standard BVs. Important for probabilistic forecasts, SPBV ensembles were shown to be reliable in the sense that each ensemble member is equally likely to be closest to the truth. Furthermore, SPBVs were shown to be dynamically consistent and recover characteristic features of the temporal evolution of errors in chaotic dynamical systems. Additionally, random draw bred vectors (RDBVs), which are designed to sample from the marginal equilibrium density of the fast variables (and hence are not conditioned on the slow variables), were introduced.
While RDBVs are not dynamically consistent and are typically over-dispersive, they were found to still have improved forecast skill over standard BVs.

In this work we extend the ideas proposed in [16] to the situation of general dynamical systems without time-scale separation. In the realistic situation when the state of the atmosphere is given by the analysis output from a data assimilation procedure, a good forecast ensemble has to satisfy two constraints: It has to evolve into likely future states, and it has to account for the uncertainty of the analysis used to generate the ensemble. In the case of multi-scale dynamics the latter issue was resolved by generating the necessary small synoptic uncertainty required by the analysis covariance via the relaxation of the stochastic perturbations of the fast variables onto the attractor. The situation in single-scale dynamics is more complicated. Whereas localisation of the bred vectors was beneficial in the multi-scale case and allowed for the conditioning of the SPBVs on the slow synoptic dynamical state, localisation of BVs prohibits in the single-scale scenario perturbations outside the localised region. Hence, although the resulting perturbation will be close in phase space to the original BV and appropriately sample the probability density function around it, the resulting initial conditions may not contain sufficient variability in the regions of significant uncertainty of the analysis. To investigate the performance of stochastically modified bred vectors we consider the Lorenz 96 model [28] in two settings, which support strongly localised and weakly localised BVs. We investigate the performance of SPBVs and of RDBVs in an unrealistic scenario where the truth is assumed to be known as well as a realistic scenario where BVs are determined from an analysed state. Whereas in the unrealistic scenario SPBVs provide reliable ensembles with superior forecast skill compared to classical BVs for the strongly and for the weakly localised case, in the realistic case when the ensemble is seeded around the analysis, localisation prevents SPBVs to constitute a reliable ensemble as they are not consistent with the analysis error which may be non-negligible outside the region of significant activity of the SPBV. For the weakly localised case we will show that the strength of the multiplicative noise used for generating SPBV ensembles has to be judiciously chosen as a trade-off between providing the most ensemble diversity while preserving some of the spatial correlation of the fast growing instabilities. Our numerical simulations demonstrate that SPBVs have an increased ensemble dimension and reduced RMS error in comparison to classical BVs in both the strongly and the weakly localised case. Furthermore, by means of error-spread relationships and reliability diagrams we show that SPBV ensembles are reliable provided that the perturbations are not strongly localised. Their dynamical consistency is probed by projecting onto the subspace spanned by the dominant covariant Lyapunov vectors. RDBVs will be shown to be dynamically inconsistent but nevertheless feature improved forecast skill over SPBVs.

The paper is organised as follows. In Section 2 we introduce the Lorenz 96 model [28]. In Section 3 we briefly review the breeding method. Section 4 introduces our stochastically modified bred vector ensemble methods, namely SPBVs and RDBVs, and shows how they relate to covariant Lyapunov modes as a measure of their dynamic adaptivity. Section 5 introduces the diagnostics used to evaluate the performance and the efficiency of stochastically modified bred vectors. In Section 6 the forecast skill and the reliability of each ensemble type is analysed in an idealised experiment with perfect observations. The more realistic
experiment incorporating imperfect observations and data assimilation is performed in Section 7. We conclude in Section 8 with a discussion and an outlook.

2. The Lorenz 96 system

The Lorenz 96 (L96) system \^[28\]

\[
\frac{d}{dt}X_k = -X_k - X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F
\]

(1)

with cyclic boundary conditions \(X_{k+K} = X_k\) for \(k = 1, \ldots, K\), was introduced as a caricature for the midlatitude atmosphere and has been used as a test bed for numerous studies in atmospheric sciences. The dynamics of the Lorenz 96 system is characterized by energy conserving nonlinear transport, linear damping and forcing. The variables \(X_k\) can be interpreted as large scale atmospheric fields arranged in the midlatitudes on a latitudinal circle of 30,000 km, such as synoptic weather systems. The classical choice \(K = 40\) corresponds to a spacing between adjacent variables of roughly the Rossby radius of deformation of 750 km. We shall also consider \(K = 128\) which implies a spacing between adjacent sites of 234 km. In both cases we use as forcing amplitude \(F = 8\) which implies chaotic dynamics \^[29\]. The setting with \(K = 40\) reproduces dynamical patterns with a realistic number of Rossby-like waves and is frequently used in the context of data assimilation. On the other hand, the choice \(K = 128\) is used to study intrinsic properties of spatially extended dynamical systems \^[39\]. For \(F = 8\) the climatic variance is estimated as \(\sigma^2 = 13.25\) and the decorrelation (\(e\)-folding) time is \(\tau = 0.41\), for both \(K = 40\) and \(K = 128\). The maximal Lyapunov exponent is measured as \(\lambda_{\text{max}} = 1.69\) for \(K = 40\) and as \(\lambda_{\text{max}} = 1.775\) for \(K = 128\). The L96 system is extensive \^[22\], in the sense that many relevant quantities (such as surface width, attractor dimension, entropy) scale linearly with the size \(K\) and the Lyapunov exponents converge to a density in the limit \(K \to \infty\). This is illustrated in Figure 1, where we show the Lyapunov spectrum for \(K = 40\) and \(K = 128\). For \(K = 40\) there are 13 distinct positive Lyapunov exponents, while for \(K = 128\) there are \(42 \approx \frac{128}{40} \times 13\) distinct positive Lyapunov exponents.

To numerically simulate the L96 system we employ a fourth-order Runge-Kutta method with a fixed time step \(dt = 0.005\). In our simulations an initial transient time of 5000 time units is employed to assure that the dynamics has settled on the attractor.

3. Bred vectors and the breeding method

We briefly review the classical breeding method introduced by \^[47, 48\]. BVs are finite-size, periodically rescaled perturbations generated using the full non-linear dynamics of the system. Centred around a control trajectory \(z_c(t_i)\) at some time \(t_i\), perturbed initial conditions of size \(\delta\),

\[
z_p(t_i) = z_c(t_i) + \delta \frac{p}{\|p\|},
\]

are defined where \(p\) is an initial arbitrary random perturbation. The control and the perturbed initial condition are simultaneously evolved using the full non-linear dynamics for an integration time \(T\) until time \(t_i+1 = t_i + T\). At the end of the integration window the difference between the two trajectories

\[
\Delta z(t_{i+1}) = z_p(t_{i+1}) - z_c(t_{i+1})
\]
Figure 1. Lyapunov exponent spectrum for the L96 model with $F = 8$ for $K = 40$ and $K = 128$.

is determined, and the bred vector is defined as the difference rescaled to size $\delta$ with

$$b(t_{i+1}) = \delta \frac{\Delta z(t_{i+1})}{\|\Delta z(t_{i+1})\|}.$$ 

The perturbation $b(t_{i+1})$ then determines the initial condition of the perturbed trajectory $z_p(t_{i+1}) = z_c(t_{i+1}) + b(t_{i+1})$ at the start of the next breeding cycle. This process of breeding is repeated for several cycles until the perturbation maintains a sufficiently large growth rate and until the perturbations converge in the sense that at time $t_n$ an ensemble of BVs spans the same subspace as BVs obtained in a breeding cycle which had been initialised further in the past. The number of cycles needed for this convergence depends on the time-scales of the dominant instabilities [47]. For the L96 system a breeding cycle length of $T = 0.05$ time units is employed for all simulations, and we employ a spin-up time for the BVs of 500 time units (which amounts to 10000 breeding cycles). An ensemble of $N$ BVs is created by $N$ independent breeding cycles initialised from independent initial perturbations $p$. The resulting converged BV ensemble at time $t_i$ is then employed as initial conditions for ensemble forecasts. The breeding method is conceptually similar to the method for generating Lyapunov vectors. They differ though in that Lyapunov vectors are generated using the linearised dynamics and an infinitesimal perturbation $\delta$, whereas BVs are generated using the full nonlinear model and finite perturbation sizes. In contrast to covariant Lyapunov vectors which are mapped by the linear tangent dynamics onto each other, the dynamics of finite-size BVs is not given by a linear mapping and as such they technically do not form a vector space. Despite the similarities between BVs and Lyapunov vectors, we adopt here the point of view outlined in [16] that for probabilistic ensemble forecasts the object of interest are the perturbed states $z_p = z_c + \Delta z$, which constitute the sample points for the Monte-Carlo approximation of the probability density function, rather than the differences $\Delta z$. We nevertheless follow the general convention and label them as vectors. We shall investigate in how far BVs are close to Lyapunov vectors in Section 34.1.
The bred vectors of the L96 system for system sizes $K = 40$ and $K = 128$ have markedly different spatial structures. This is illustrated in Figure 2 where snapshots of a typical BV are shown for $K = 40$ and for $K = 128$ for $\delta = 0.1$. For the larger $K = 128$ system BVs are strongly localised with only a well-defined group of sites having significant entries, whereas for $K = 40$ the localisation is less well defined. To quantify the degree of localisation we introduce the localisation parameter

$$L = \frac{\sum_{k=1}^{K} b_k^4}{\left( \sum_{k=1}^{K} b_k^2 \right)^2},$$

where $L \in [1/K, 1]$. If the vector $b$ does not express any localisation with all components of $b$ being equal, we have $L = 1/K$. In the extreme case of localisation when $b$ has only one non-zero component, $b_{k^*} = 1$ for some $k = k^*$ and $b_k = 0$ for $k \neq k^*$, we have $L = 1$.

Since the localisation $L$ depends on $K$, one cannot directly compare the values of $L$ for different system sizes. For the BVs shown in Figure 2 we obtain $L = 0.1589$ for $K = 40$ and $L = 0.1375$ for $K = 128$. However, when comparing the average localisation of a BV to that of a spatially uncorrelated Gaussian random perturbation of the same size for reference, the two cases can be clearly separated. When averaging over 2500 realisations Gaussian random perturbations have localisation with $\bar{L}_{GRP} = 0.0715$ for $K = 40$ and $\bar{L}_{GRP} = 0.0230$ for $K = 128$, and the localisation $L$ is for $K = 128$ on average approximately 6.5 times larger than that of a random Gaussian perturbation, whereas for $K = 40$ it is only about 2.4 times larger.

For the non-localised case $K = 40$, BVs nevertheless exhibit a nontrivial spatial correlation structure. To measure the spatial organisation of BVs we consider the $K \times K$ covariance matrix

$$C = \frac{b(t) [b(t)]^T}{\|b(t)\|^2 \|b(t)\|^2},$$

and determine its average $\bar{C} = \langle C \rangle$, where the average is taken over realisations of independent BVs generated at different points in time. Since all components for the L96 system are statistically equivalent the $k$-th and $(k + l)$-th rows of $\bar{C}$ are identical up to a shift of $l$ components. Figure 3 shows the row-averaged $\bar{C}_{k, l}$ of the matrix $C$ for $K = 40$ for some arbitrary component $k$ for the BV depicted in Figure 2.

In probabilistic forecasting the aim is to approximate the density $\rho(X, \tau)$ at lead time $\tau$ given an initial density $\rho(X, t = 0)$, describing the current estimate of the system. BVs are designed to represent a good Monte Carlo estimate of $\rho(X, t = 0)$ utilising that fast growing initial conditions are likely to be observed at later times $\tau$. The capability of BVs to form an ensemble of independent initial conditions suited for a reliable probabilistic ensemble forecast, depends crucially on the perturbation size $\delta$. For too large perturbation sizes $\delta$, the initial conditions resemble random draws from the attractor (after a typically rapid transition towards it) and the forecast skill deteriorates. Contrary, for too small values of the perturbation size, BVs align with the leading Lyapunov vector (LLV) exhibiting ensemble collapse. An ensemble of $N = 20$ BVs with $\delta = 0.1$, which were initialised with different random perturbations, collapses and the ensemble members are indistinguishable by eye from the ones depicted in Figure 2 for both $K = 40$ and $K = 128$. This lack of
diversity of an ensemble of bred vectors presents a major drawback of bred vectors in ensemble forecasting. For example, [49] observed the collapse of BVs onto the LLV in the NCAR Community Climate Model. In the language of probabilistic forecasts the alignment of BVs with the LLV implies that only a single draw from $\rho(X, t = 0)$ is considered. In the next section we devise a method how to overcome this drawback while still preserving the desirable features of BVs such as their low computational cost and their dynamical consistency [40].

![Figure 2. Bred vectors of the L96 system with perturbation size $\delta = 0.1$ for different system size $K$. Top: $K = 40$. Bottom: $K = 128$.](image)

4. **Stochastically perturbed bred vectors**

We review here the method proposed in [16] to increase the diversity of BV ensembles for multi-scale systems and apply it to systems without scale separation. To generate a diverse ensemble of initial conditions conditioned on the current state of the system, BVs are generated from a parent BV by applying a multiplicative stochastic perturbation to it. The key idea to generating additional draws from the initial density function $\rho(X, t = 0)$ is to exploit the fact that in spatially extended dynamical systems, BVs are often localised
stochastically perturbed bred vectors

Figure 3. Row-averaged $\bar{C}_k$, of the covariance matrix (3) for the BV of the L96 system with perturbation size $\delta = 0.1$ displayed in Figure 2.

(as shown in Figure 2 for $K = 128$) or exhibit some non-trivial spatial structure (as shown in Figure 3 for $K = 40$), corresponding to some degree of spatial organisation of error growth. Stochastically perturbed bred vectors (SPBVs) are designed to preserve the spatial structure of BVs, which is paramount to conditioning the initial density $\rho(X, t = 0)$ on the current state $X$. SPBVs are defined as

$$b_{sp} = \delta \frac{(I + \Xi)b}{\| (I + \Xi)b \|},$$

(4)

where $I$ is the $K \times K$ identity matrix. The diagonal $K \times K$ matrix $\Xi$ with entries $\xi_{jj} \sim N(0, \sigma^2)$ for $j = 1, \ldots, K$ with variance parameter $\sigma^2$ represents the stochastic perturbation. The stochastic perturbation is performed only once as a post-processing step from a given parent BV when generating initial conditions for a forecast ensemble, and therefore does not significantly add to the computational cost. In Figure 4 we show a realisation of an SPBV, overlaid with their parent BV, for a perturbation size of $\delta = 0.1$ with noise strength $\sigma = 1.25$, for $K = 40$ and for $K = 128$. It is clearly seen that the spatial structure of the perturbation is preserved.

The stochastic perturbations generate initial conditions that are nearby the attractor and after a typically rapid relaxation towards the attractor along the stable manifold, approach the attractor close in phase space to the initial condition associated with the parent BV, which we know is capturing fast error growth. The stochasticity hence allows to sample the phase space on the attractor in the fastest growing region.

The noise strength $\sigma$ obviously plays a central role. When $\sigma \to 0$ SPBVs essentially reproduce the parent BVs they were generated from, and the spatial structure is exactly preserved but no diversity is gained. In the other extreme case $\sigma \to \infty$, the behaviour depends on the degree of localisation. For the strongly localised case $K = 128$ with many vanishing BV components (cf. Figure 4 (bottom)), the degree of localisation remains preserved since SPBVs are rescaled to size $\delta$, and the diversity is greatly enhanced. This is
the case discussed in the multi-scale setting in [16]. The weakly localised case when there are no significant regions with vanishing components of the BV (cf. Figure 4 (top)) is more complex. For sufficiently large magnitudes of the noise strength $\sigma$, SPBVs become spatially uncorrelated random perturbations of size $\delta$. This allows for (almost) maximal diversity of the ensemble which, however, comes at the cost of destroying the inherent spatial structure of the dynamically relevant fast growing perturbations. The destruction of the spatial structure implies that we typically do not sample the phase space region locally but instead generate initial conditions as random draws from the attractor, which are not conditioned on the current state. In Figure 6 we illustrate the loss of spatial structure by showing the average of rows $\overline{C_k}$ of the covariance matrix (3) for SPBVs for increasing values of the noise strength $\sigma$. It is seen that for $\sigma = 1.0$ and for $\sigma = 1.25$ the nontrivial correlations between adjacent sites are preserved albeit reduced in magnitude, whereas for $\sigma = 5$ the spatial structure is entirely lost and adjacent sites are uncorrelated. This suggests that $\sigma$ needs to be tuned in the weakly localised case to balance diversity and reliability of the ensemble with forecast skill and root-mean-square errors of the forecast. We shall provide numerical evidence for these statements in Section 6 after introducing the relevant diagnostics in Section 5.

We also consider so called random draw bred vectors (RDBVs) introduced in [16]. An ensemble of RDBVs is generated by randomly selecting classical BVs which were started from independent initial conditions randomly drawn from the attractor. To avoid storing a huge library of independent BVs, an ensemble of RDBVs is generated on the fly by evolving $N$ independent control trajectories started from independent initial conditions, each generating a single BV. Whereas SPBVs are designed to sample the phase space locally, RDBVs are dynamically inconsistent in the sense that they may, after a quick relaxation towards the attractor, evolve into states which are not close to the control trajectory. Example RDBVs for $K = 40$ and $K = 128$ are shown in Figure 5. We remark that, contrary to SPBVs, RDBVs form an (almost) orthogonal ensemble.

4.1. Dynamic properties of bred vectors: Backward and covariant Lyapunov vectors. We now probe how bred vectors and their stochastic modifications relate to dynamically relevant modes such as Lyapunov vectors which capture the asymptotic growth of infinitesimal perturbations, and thereby in how far they are dynamically adapted. The dynamic adaptivity of classical BVs was established in [40]. We now show that SPBVs inherit this property form their parent BVs. In particular, we consider the relationship between bred vectors and backward Lyapunov vectors and covariant Lyapunov vectors. Backward Lyapunov vectors are initialised in the asymptotically distant past and are generated by solving the linear tangent model of the dynamical system under a Gram-Schmidt orthogonalisation procedure. The orthogonal backward Lyapunov vectors are not covariant under the linear tangent dynamics and all of them typically evolve under the dynamics into the leading Lyapunov vector (LLV). Covariant Lyapunov vectors, on the contrary, form a typically non-orthogonal basis of the tangent space and are mapped onto each other by the linearised tangent dynamics. The associated asymptotic growth rates of backward and covariant Lyapunov vectors, the Lyapunov exponents, are shown in Figure 1 for the L96 system. As BVs, the first few leading covariant Lyapunov vectors exhibit a localised spatial
structure in the L96 system (not shown), with strong localisation for $K = 128$ and weak localisation for $K = 40$.

We quantify the relationship between the respective BV ensembles and Lyapunov vectors by measuring the average projection of BV ensembles onto backward and onto covariant Lyapunov vectors, and consider the following measure for the degree of projection

$$\pi_n^i(t) = \frac{\| b_n(t) \cdot l_i(t) \|}{\| b_n(t) \| \| l_i(t) \|},$$

where $b_n(t)$ denotes the $n$th bred vector ensemble member at time $t$ and $l_i(t)$ denotes the Lyapunov vector corresponding to the $i$th largest Lyapunov exponent at time $t$. We report here on the average degree of projection $\bar{\pi}_i$ where we average $\pi_n^i(t)$ over time and over the ensemble members. Note that $\bar{\pi}_i = 1$ corresponds to perfect alignment and $\bar{\pi}_i = 0$ corresponds to (on average) no alignment.

There exist several efficient numerical algorithms to calculate the covariant Lyapunov vectors $[52, 17]$. We use here the algorithm by $[17]$ as described in $[24]$ to numerically
calculate covariant Lyapunov vectors. We use a spin-up time of 2500 time units to converge to the set of backward Lyapunov vectors evolving forward in time, and a further 2500 time units to ensure convergence to the set of expansion coefficients of the covariant Lyapunov vectors that express the covariant Lyapunov vectors in the basis of forward and backward Lyapunov vectors respectively, evolving backward in time. Orthonormalisation of the backward Lyapunov vectors is performed at every time step.

Figures 7 and 8 display $\bar{\pi}_i$ for all $i = 1, \ldots, K$ backward Lyapunov vectors for $K = 40$ and $K = 128$ respectively, for classical BVs (top) and SPBV (bottom). It is clearly seen that both, classical BVs and SPBV, project almost completely onto the first backward Lyapunov vector (the LLV) for small $\delta < 0.1$ and are orthogonal to all other directions for both dimension sizes. When the perturbation size lies between $0.1 \lesssim \delta \lesssim 8$ for $K = 40$ and between $0.1 \lesssim \delta \lesssim 5$ for $K = 128$, BVs also project onto the next few backwards Lyapunov vectors. We shall see in the next section, that for these perturbation sizes, BV ensembles have collapsed to a single member and have an ensemble dimension (to be defined below in
Figure 6. Row-averaged $\bar{C}_k$, from the covariance matrix (3) for SPBVs of the L96 system with various noise strengths $\sigma = 1.0$, $\sigma = 1.25$ and $\sigma = 5.0$. The covariance of BVs is depicted as a reference.

(6)) strictly equal to 1 (cf. Figure 11). For $\delta > 1$ the non-vanishing projections of BVs onto the next Lyapunov vectors stem from increasing fluctuations of the BV ensemble around the LLV. We observe that $\pi_n^i(t)$ may strongly fluctuate in time and individual members of a BV/SPBV ensemble may exhibit, locally in time, strong projections on higher Lyapunov vectors. In such cases when BVs/SPBVs do not fully align with the LLV, they lie typically in the subspace spanned by the first few Lyapunov vectors (not shown). For even larger values of the perturbation size, BVs do not significantly project onto the linear Lyapunov vectors as they evolved into truly nonlinear objects.

Projections of BVs and SPBVs onto covariant Lyapunov vectors exhibit similar signatures as for backward Lyapunov vectors. Contrary to backward Lyapunov vectors, covariant Lyapunov vectors do not form an orthogonal basis. Furthermore, successive covariant Lyapunov vectors are likely to be localised in similar spatial regions to each other whereas this is not the case for backward Lyapunov vectors due to non-dynamical orthogonality constraint [20]. Hence fluctuations of BVs cause them to project onto several of the covariant Lyapunov vectors. In particular, we see strong projections of classical BVs onto covariant Lyapunov vectors with index $i \leq 7$ for $K = 40$, and onto those with index $i \leq 12$ for $K = 128$. As for backward Lyapunov exponents, the projection onto dynamically relevant low-index Lyapunov vectors drops off when BVs gain diversity at $\delta \approx 8$ for $K = 40$ and at $\delta \approx 5$ for $K = 128$. SPBVs feature weaker projections onto the higher-index covariant Lyapunov vectors with significant projections in the smaller range $i \leq 3$ and $i \leq 6$ for $K = 40$ and $K = 128$, respectively. This overall stronger projection of SPBVs to the low-index covariant Lyapunov vectors is caused by the ensemble averaging (SPBVs are generated from a single collapsed BV). As for backward Lyapunov vectors, the lower dimensional subspaces onto which BVs and SPBVs project onto can fluctuate over time, in particular for larger values of the perturbation size $\delta$ (not shown). It is pertinent to notice that the projection of SPBVs onto the Lyapunov vectors weakens as $\sigma$ is increased and saturates for sufficiently
large $\sigma$, in particular in the weakly localised case $K = 40$.

We conclude that BVs and SPBVs share similar localisation structure to that of the first few covariant Lyapunov vectors, and that for sufficiently small noise strength $\sigma$ SPBVs inherit from BVs the desirable property of dynamical adaptivity. RDBVs, on the other hand, do not exhibit any significant average projections onto any of the backward or covariant Lyapunov vectors as they are unrelated to the local dynamics (not shown). RDBVs are hence dynamically not adapted.

5. Diagnostics

To illustrate how SPBVs can be used as a reliable diverse ensemble with improved forecast skill we now introduce several diagnostics. In particular, we consider the ensemble dimension to measure the diversity of an ensemble, the root-mean-square error to quantify the forecast skill and several reliability measures to probe the probabilistic properties of an ensemble.
5.1. Ensemble dimension. We quantify the diversity of an ensemble using the "ensemble dimension" [8, 33], also known as the "bred vector dimension" [38]. The ensemble dimension is a measure for the dimension of the subspace spanned by a set of vectors. For an ensemble of $N$ BVs $\{b^{(n)}(t)\}_{n=1,...,N}$ at a given time $t$, the ensemble dimension is defined as

$$D_{ens}(t) = \left( \frac{\sum_{n=1}^{N} \sqrt{\mu_n}}{\sum_{n=1}^{N} \mu_n} \right)^2,$$

where the $\mu_n$ are the eigenvalues of the $N \times N$ covariance matrix $C$ (cf. [33]). The ensemble dimension takes values between $D_{ens} = 1$ and $D_{ens} = \min(N, D)$, where $D$ is the total dimension of the dynamical system, depending on whether the ensemble members are all aligned or are orthogonal to each other. We consider in our numerical experiments the temporal average $\overline{D}_{ens}$ to characterise the diversity of an ensemble.

Figure 8. Average projection $\bar{\pi}_i$ of backward Lyapunov vectors for $K = 128$. Results are shown on a logarithmic scale. Top: BVs. Bottom: SPBVs with $\sigma = 1.25$. 
5.2. Ensemble forecast skill. To measure the performance of an ensemble with $N$ members $X_k^{(n)}$, $n = 1, \ldots, N$, we consider the root-mean-square error (RMS error) of the ensemble average with respect to the truth. Denoting ensemble means by angular brackets, we introduce the ensemble mean

\begin{equation}
\langle X_k \rangle = \frac{1}{N} \sum_{n=1}^{N} X_k^{(n)}
\end{equation}

and the site-averaged root-mean-square error between the truth $X_k^{tr}$ and the ensemble average over $M$ realizations

\begin{equation}
\mathcal{E}(\tau) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \frac{1}{K} \sum_{k=1}^{K} \| X_k^{tr}(\tau) - \langle X_{k,m} \rangle(\tau) \|^2}
\end{equation}
as a function of the lead time $\tau$. The index $m = 1, \ldots, M$ denotes the realisation. Similarly, as a measure of the uncertainty of the ensemble average, we consider the site-averaged root-mean-square spread (RMS spread)

\begin{equation}
S(\tau) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \frac{1}{K} \sum_{k=1}^{K} \langle \| X_{k,m}^{(n)}(\tau) - \langle X_{k,m} \rangle(\tau) \|^2 \rangle}.
\end{equation}

### 5.3. Reliability.

The RMS error is not always the appropriate measure to quantify the performance of an ensemble in probabilistic forecasting. For example, in the case when the probability density function has disjoint support, the ensemble mean is a bad forecast and is not even physically meaningful. For probabilistic forecasts the reliability of an ensemble is more relevant. In a so called *perfect* ensemble each ensemble member and the truth are independent draws from the same probability density function $\rho(X)$. In perfect ensembles the ratio between the RMS error of the ensemble mean and the ensemble spread approaches 1 as the ensemble size increases \cite{51,26}. Ratios smaller or larger than 1 indicate that the
ensemble is either under or over-dispersive, respectively. Furthermore, in reliable ensembles the truth is statistically indistinguishable from any given ensemble member and each ensemble member has equal probability to be closest to the truth. This property can be probed in Talagrand or Rank histograms [1, 19, 45]. To generate a Talagrand histogram, the \( N \) ensemble members are sorted at each forecast time defining a set of \( N + 1 \) bins. A histogram of probabilities of the truth falling into a bin \( i \) is then produced by counting the frequency that the truth falls into the bin \( i \). A flat histogram indicates a reliable ensemble for which each ensemble member has equal probability of being nearest to the truth. A convex histogram indicates a lack of spread of the ensemble, and a concave diagram indicates an excess of spread of the ensemble [51].

6. Simulations with ensembles initialised from the truth

We now present numerical results demonstrating that SPBVs can be used as a reliable diverse ensemble with improved forecast skill. We shall present results for the strongly localised case with \( K = 128 \) and for the weakly localised case with \( K = 40 \) separately. We examine the ensemble diversity, forecast skill metrics such as the RMS error and RMS spread, as well as the reliability quantified by the error-spread relationship and the Talagrand diagram. We start with an idealised scenario in which BV ensembles are seeded from the actual truth which in realistic applications would not be accessible. We consider this idealised case to show that our premise that SPBVs allow for a reliable estimation of the conditional probability density function by generating initial conditions close in phase space to a reference point which then subsequently move into likely states of the probability density function at a later lead time \( \tau \).

The setup for the numerical simulations is as follows. The L96 model (1) is integrated from an arbitrary initial condition first for a transient period of 500 time units. After this transient period which ensures that the dynamics has settled on the attractor, BV ensembles are generated by perturbing from the true state as our control trajectory for a spin-up time of 500 time units and a breeding cycle length of \( T = 0.05 \) time units. We consider BV ensembles consisting of \( N \) independent BV perturbations of size \( \delta \). For \( K = 40 \) we use \( N = 10 \) and for \( K = 128 \) we use \( N = 20 \) independent ensemble members. Each ensemble member is then evolved freely under the L96 dynamics for some lead time \( \tau \). Forecasts are run for a total of 5 time units and we report the results for lead times \( \tau = 2.0 \) and \( \tau = 4.0 \) time units for the forecast metrics presented in Section 5. A new forecast is created each 1.0 time units. SPBVs have been generated using a noise strength of \( \sigma = 1.25 \). All metrics are averaged over \( M = 2500 \) forecasts.

6.1. Ensemble Dimension. Figure 11 shows the average ensemble dimension \( \bar{D}_{ens} \) (9) as a function of \( \delta \) for classical BVs, SPBVs with \( \sigma = 1.25 \), and RDBVs. For classical BVs the average ensemble dimension is \( \bar{D}_{ens} = 1 \) for \( \delta \lesssim 8 \) and for \( \delta \lesssim 5 \), for \( K = 40 \) and \( K = 128 \) respectively, indicating the collapse of BV ensembles. For these perturbation sizes, a BV ensemble typically collapses onto the LLV but can also for \( \delta > 1 \), when the dynamics of the perturbation begins to feel the nonlinearity of the dynamics, align in a different direction, spanned by the first few leading Lyapunov vectors (cf. Figures 78). For even larger perturbation sizes \( \delta > 5 \) for \( K = 40 \) and \( \delta > 8 \) for \( K = 128 \), the nonlinear dynamics becomes dominant and the ensemble dimension increases rapidly. Perturbation sizes corresponding to \( \bar{D}_{ens} > 1 \), however, are unrealistic in the sense that they are much larger than typical
analysis errors for the L96 model as reported in Bowler [7], Ng et al. [32] and Pazó et al. [39]. This implies that classical BVs in our setting lack sufficient diversity. We remark that the qualitative behaviour of $\bar{D}_{\text{ens}}$ does not change with the number of independent ensemble members $N$.

![Graph](image)

**Figure 11.** Average ensemble dimension $\bar{D}_{\text{ens}}$ as a function of $\delta$ for each ensemble generation method for the L96 system \((1)\). The SPBV ensemble was generated using $\sigma = 1.25$. Top: $K = 40$. Bottom: $K = 128$.

SPBVs and RDBVs exhibit a significant increase in the ensemble dimension. Both methods produce ensembles with a much larger ensemble dimension than the original BVs for all values of $\delta$. SPBVs maintain a consistent ensemble dimension of $\bar{D}_{\text{ens}} = 6.4$ for $K = 40$ and $\bar{D}_{\text{ens}} = 10.2$ for $K = 128$, before increasing in conjunction with the BVs when $\delta$ is large. RDBVs support the highest ensemble dimension as they are independent from each other. They do not attain the maximum ensemble dimension $\bar{D}_{\text{ens}} = N$ since they are not strictly orthogonal. The averaged ensemble dimension of SPBVs is closer to the maximum ensemble dimension of $\bar{D}_{\text{ens}} = 10$ for $K = 40$ than to $\bar{D}_{\text{ens}} = 20$ for $K = 128$, reflecting the
differing degree of localisation in the two cases; the multiplicative stochastic perturbation can generate a larger ensemble subspace the smaller the degree of localisation.

The ensemble dimension of SPBVs increases for increasing values of the noise strength $\sigma$ as shown in Figure 12 for SPBVs with $\delta = 0.1$. The ensemble dimension approaches a limiting value of $\bar{D}_{\text{ens}} = 7.4$ for $K = 40$ and of $\bar{D}_{\text{ens}} = 12.4$ for $K = 128$. The limiting ensemble dimension is smaller than $N$ for both $K = 40$ and $K = 128$. The difference is stronger for the strongly localised case $K = 128$ for the same reason as discussed above. The observed increase of the ensemble dimension with increasing noise strength may lead to the false conclusion that one should use sufficiently large noise strengths $\sigma$ and moreover that the performance is insensitive to changes in $\sigma$ past some threshold value. We will see that this is correct for the forecast skill but that increasing the noise strength leads to unreliable over-dispersive ensembles.

![Figure 12](image_url)

**Figure 12.** Average ensemble dimension $\bar{D}_{\text{ens}}$ of an SPBV ensemble with perturbation size $\delta = 0.1$ as a function of $\sigma$ for the L96 system (1) for $K = 40$ (with $N = 10$ ensemble members) and for $K = 128$ (with $N = 20$ ensemble members).

6.2. **Ensemble Forecast Skill.** Figure 13 shows the RMS error $\mathcal{E}$ for BVs, SPBVs and RDBVs as a function of the perturbation size $\delta$ for lead times $\tau = 2.0$ and $\tau = 4.0$, both for the strongly localised case $K = 128$ and the weakly localised case $K = 40$. It is seen that classical BVs exhibit the largest RMS error for both cases and both lead times and for all values of $\delta$. In the weakly localised case $K = 40$, BVs do not achieve a forecast error smaller than the climatic error $\mathcal{E}_{\text{clim}} = \sigma_{\text{clim}} \approx 3.64$ for $\delta > 0.1$ (for $\tau = 4.0$) and $\delta > 2.3$ (for $\tau = 2.0$). In the strongly localised case $K = 128$, BVs incur forecast errors larger than the climatic error for perturbation sizes $\delta > 0.5$ for $\tau = 4.0$. For very large values of $\delta \gtrsim 20$, when each ensemble member can be considered as a random draw from the attractor, all BV ensembles acquire the same RMS error which equals the climatic error $\mathcal{E}_{\text{clim}}$. The observed asymptotic values in Figure 13 for $\delta > 20$ are slightly larger than the climatic error, which can be attributed to the finite ensemble size, with a larger deviation for the $N = 10$ ensemble for $K = 40$ than for the $N = 20$ ensemble for $K = 128$. Classical BVs
Figure 13. RMS error $\mathcal{E}$ as a function of $\delta$ in the perfect observation setup, for each ensemble generation method for fixed lead times $\tau = 2.0$ and $\tau = 4.0$. SPBVs were generated with $\sigma = 1.25$. Top: $K = 40$. Bottom: $K = 128$.

approach the climatic error $\mathcal{E}_{\text{clim}}$ around $\delta \approx 5$ and $\delta \approx 8$. Note that this coincides exactly with the increase of the ensemble dimension from $\bar{D}_{\text{ens}} = 1$ (cf. Figure 11). It is further seen that BV ensembles with no diversity for which $\bar{D}_{\text{ens}} = 1$, the RMS error can become larger than $\mathcal{E}_{\text{clim}}$ and approaches $\sqrt{2}\mathcal{E}_{\text{clim}} \approx 5.2$. The observed general poor performance in forecast skill for small values of $\delta$ stems from the lack of diversity with $\bar{D}_{\text{ens}} = 1$ and the collapse (on average) onto the LLV (cf. Figure 11).

In contrast, SPBVs and RDBVs which exhibit increased ensemble diversity, perform significantly better with markedly smaller RMS errors than classical BVs in all cases. RDBVs consistently have the smallest RMS error for all values of $\delta$ and all lead times $\tau$. The difference in RMS error between RDBVs and SPBVs is more pronounced in the strongly localised case $K = 128$. In this case, by construction SPBVs are less diverse with smaller ensemble dimension (cf. Figure 11), as all SPBV members are localised around the very same sites; the increased ensemble dimension of RDBVs leads to less divergence of the ensemble mean
Figure 14. RMS spread $S$ as a function of $\delta$ in the perfect observation setup, for each ensemble generation method for fixed lead times $\tau = 2.0$ and $\tau = 4.0$. SPBVs were generated with $\sigma = 1.25$. Top: $K = 40$. Bottom: $K = 128$.

from the truth, improving the overall average forecast skill compared to SPBVs. Figure 15 shows a typical ensemble forecast for the L96 system for the arbitrary component $X_1$, with a perturbation size of $\delta = 0.1$. We show the truth together with the ensemble forecast mean, as well as the individual ensemble members. The figure illustrates how small sub-synoptic perturbations may cause the ensemble members to evolve into different parts of phase-space and acquire synoptic variance. The lack of diversity of BVs is clearly seen to be detrimental and the collapsed ensemble diverges from the truth for lead times larger than 2, causing the poor performance in the RMS error reflected in Figure 13. The stochastically modified bred vectors both feature more diversity, allowing some ensemble members to explore the different synoptic “futures” and different parts of the phase space. The spread is particularly large for the independently drawn RDBV ensemble, which can sample larger parts of phase space. This increased spread leads to less ensemble divergence from the truth.
and is reflected in the superior RMS error performance of RDBVs (cf. Figure 13).

The associated RMS spread $\mathcal{S}$ (9) for BVs, SPBVs and RDBVs as a function of the perturbation size $\delta$ is shown in Figure 14 and is consistent with the results on the ensemble dimension and on the RMS error $\mathcal{E}$ shown above. In particular, RDBVs exhibit the largest spread whereas BVs have vanishing spread, corresponding to $D_{\text{ens}} = 1$ for $\delta < 5$ and $\delta < 8$ when BVs have collapsed on average onto the LLV.

We now investigate how the RMS error for SPBVs changes as the noise strength $\sigma$ is varied. We recall that when $\sigma \to 0$ SPBVs essentially reproduce the original BVs they were generated from, while once $\sigma$ is sufficiently large, the ensemble dimension saturates at some fixed value due to the rescaling back to size $\delta$ (cf. Figure 12). Figure 16 shows the RMS error of SPBVs with perturbation size $\delta = 0.1$ as a function of the noise strength $\sigma$ for both system sizes and lead times. For comparison we also show the RMS error of BV and of RDBV ensembles (which both do not depend on $\sigma$). It is seen that the RMS error deviates rapidly from the value attained by BVs for increasing values of $\sigma$, and then asymptotes to a constant value for large $\sigma$. In the strongly localised case $K = 128$, SPBVs preserve the localisation structure for any value of $\sigma$. For the weakly localised case $K = 40$, however, sufficiently large values of $\sigma$ destroy any existing spatial correlations (cf. Figure 6). Again, this is reflected in the behaviour of the RMS error as a function of the noise strength $\sigma$. For the weakly localised case $K = 40$ the asymptotic value of the RMS error is close to the one of RDBVs - the spatial structure of both ensembles is not related to the current state and their associated initial conditions evolve into random draws from the attractor, so both ensembles have the same statistical properties. For the strongly localised case $K = 128$, on the other hand, the asymptotic RMS error of SPBVs is larger than the one of RDBVs. In the localised case, SPBV ensembles have markedly different statistical properties to RDBVs as they sample locally with all ensemble members exhibiting non-vanishing entries in the same spatial region. Note that the value of $\sigma = 1.25$ used in previous examples does not correspond to the minimal RMS error. In the next section we shall see that it is advantageous to use sub-optimal values of $\delta$ with regards to forecast skill, to achieve a reliable ensemble of SPBVs.

6.3. Reliability. We now use the error-spread ratio and the Talagrand histogram to evaluate if the additional spread acquired by the stochastic modifications of BVs is beneficial in the sense that it leads to a reliable ensemble or whether it causes the ensemble to be simply over-dispersive. The RMS error–spread ratio as a function of the lead time $\tau$ is shown in Figure 17 and the Talagrand diagram is shown in Figure 18 for $\delta = 0.1$ for BVs, SPBVs and RDBVs. We use again $\sigma = 1.25$ for SPBVs (this particular choice will finally be discussed further down). It is clearly seen that classical BV ensembles are highly under-dispersive for $\delta = 0.1$ as they have collapsed to a single member. Both reliability diagnostics clear indicate that SPBV ensembles are reliable whereas RDBV ensembles are significantly over-dispersive for both $K = 40$ and $K = 128$ for all lead times, despite their superior performance with respect to the forecast RMS error. We conjecture that the superior statistical properties of SPBV ensembles are linked to them preserving the spatial structure of the fastest growing modes and thereby sampling the initial probability density conditioned on the current state. The non-dynamically adapted RDBVs, on the other hand, typically do not evolve into states which reliably sample the probability density function.
Figure 15. A typical ensemble forecast for the L96 system \cite{11} with 10 ensemble members. Depicted are the truth, ensemble members and the ensemble average for the slow component $X_1$ for Top: BVs. Middle: SPBVs. Bottom: RDBVs. The perturbation size is $\delta = 0.1$. 
at later lead times. We remark that the error-spread ratio curves and the Talagrand diagram do not change qualitatively when changing the perturbation size $\delta$, provided $D_{\text{ens}} = 1$.

Now we look at how the reliability of an SPBV ensemble depends on the noise strength $\sigma$. We show in Figures 19 and 20 the error-spread ratio and the Talagrand histogram, averaged over all sites, for values of $\sigma = 1$, $\sigma = 1.25$ (the value we had so far employed) and for $\sigma = 5$. It is clearly seen that there is a sweet spot of the noise strength for which the system is reliable at $\sigma = 1.25$. For $\sigma = 1.25$ the error–spread ratio is exactly equal to 1 and the Talagrand histogram is flat, indicating a reliable ensemble, for both $K = 40$ and $K = 128$. For values of $\sigma < 1.25$ the SPBV ensemble is under-dispersive whereas it is over-dispersive for larger noise levels with $\sigma > 1.25$. Recall that for $\sigma = 1.25$ the RMS error was not minimal. Hence, the noise level $\sigma$ can be used to tune the SPBV forecast ensemble.

![Figure 16. RMS error of SPBVs (red) as a function of $\sigma$ for perturbation size $\delta = 0.1$. Shown for reference are the RMS errors for BVs (black) and RDBVs (blue). Top: $K = 40$. Bottom: $K = 128$.](image-url)
Figure 17. RMS error vs RMS spread for forecast ensembles in the perfect observation setup, parameterised by increasing lead times from $\tau = 0$ to $\tau = 5.0$ time units. The markers indicate the specific lead times $\tau = 2.0$, $\tau = 3.0$ and $\tau = 4.0$. The grey dot-dashed line indicates a one-to-one ratio of error and spread, corresponding to a reliable ensemble. SPBVs were generated with $\sigma = 1.25$. Top: $K = 40$. Bottom: $K = 128$.

to favour reliability over forecast RMS error or vice versa. We shall in the following continue to use $\sigma = 1.25$, maximizing reliability of SPBV ensembles.

7. Simulations with ensembles initialised from the analysis

Finally we turn our attention to the more realistic setup of incorporating analysis errors into the forecast and generating forecast ensembles as perturbations from the analysis. We employ an ensemble Kalman filter to perform the data assimilation and construct the analysis. Details are provided for completeness in the Appendix. The analysis is constructed from a perfect model forecast and noisy observations with variance 0.01 (corresponding
Figure 18. Talagrand diagrams for forecast ensembles in the perfect observation setup, for $\delta = 0.1$ at lead times $\tau = 2.0$ and $\tau = 4.0$. SPBV were generated with $\sigma = 1.25$. Top: $K = 40$. Bottom: $K = 128$.

to observational noise with 2.75% of the climatological standard deviation), following \[7, 39\]. To focus on the performance of the bred vector ensemble rather than on the data assimilation, we use a large ensemble for the ETKF with $K + 1$ members to produce the analysis avoiding the need for localisation and inflation. The ETKF ensemble is spun-up for 500 time units before commencing the breeding cycles. The average analysis error for the $K = 40$ and $K = 128$ systems is 0.10 and 0.18, respectively, over 2500 forecasts. Ideally, the value of $\delta$ that results in a local minima of the RMS forecast error $E$ matches the size of the analysis error. In practice, however, the perturbation size often needs to be larger to achieve acceptable forecast skill \[48, 30, 16\].

As is common practice in operational ensemble forecasting, pairs of positive/negative BVs are generated to ensure that the BV forecast ensemble represents the analysis mean at the initial forecast time. We employ 5/10 independent breeding cycles (generating a total of $N = 10/N = 20$ ensemble members) for $K = 40/K = 128$, respectively. Note that for perturbation sizes $\delta$ for which $D_{\text{ens}} = 1$ and BVs typically collapse onto the LLV, BVs generated from the analysis are the same as BVs generated from the truth. For larger values of $\delta$, when the alignment is only approximately valid on average, we find that BVs generated from the analysis and those generated from the truth are only statistically similar in the sense that they share the same ensemble dimension (not shown).
Seeding BV forecast ensembles from the analysis poses additional problems, as a good forecast ensemble now not only should evolve into likely future states but also has to account for the uncertainty of the analysis. This is particularly a problem in the case of strongly localised BVs (and SPBVs) for $K = 128$, as the localisation inhibits to sample the uncertainty of the analysis which typically extends outside of the region of localisation and is distributed across the whole domain. Figure 21 shows snapshots of BVs and the analysis error (scaled to have norm equal to 0.1 to facilitate comparison) for the weakly localised case $K = 40$ and the strongly localised case $K = 128$. We remark that by construction, the spatial structure of SPBVs is similar to that of BVs. In the strongly localised case, it is clearly seen that there are large regions of significant uncertainty of the analysis which are not perturbed by the BV. The average localisation of analysis errors is for $K = 128$
Figure 20. Talagrand histograms in the perfect observation setup for SP-BVs with different noise strengths $\sigma$ for $\delta = 0.1$. Top: $K = 40$. Bottom: $K = 128$.

$\bar{L}_a = 0.046$ (compared to $\bar{L}_{BV} = 0.15$ for BVs) quantifying this difference in localisation. For $K = 40$ the difference in localisation is not significant with $\bar{L}_a = 0.11$ for analysis errors and $\bar{L}_{BV} = 0.16$ for BVs. Moreover, the average spatial structure of the analysis error, measured via the correlation matrix (3), is very well reproduced by all three bred vector ensembles, including RDBVs, in the weakly localised case $K = 40$ (not shown), contrary to the localised case $K = 128$.

The lack of activity in sites remote of the region of their spatial localisation is likely to severely inhibit the BV/SPBV ensemble to evolve into states which contain the truth. We shall find below, that the property of localisation is detrimental for the dynamically adapted SPBVs in the L96 system (11) without scale separation, whereas it was essential in the multi-scale case in Giggins and Gottwald [16]. In particular we show that for $K = 128$ strong localisation implies poor reliability of SPBV ensembles. RDBVs, however, despite not being dynamically adapted, exhibit improved reliability and forecast skill compared to classical BVs. In the weakly localised case $K = 40$, SPBVs and RDBs both constitute a reliable forecast ensemble with superior forecast skill compared to classical BV ensembles.
7.1. **Ensemble Forecast Skill.** Figure 22 shows the RMS error $\mathcal{E}$ for BVs, SPBV and RDBV as a function of the perturbation size $\delta$ for lead times $\tau = 2.0$ and $\tau = 4.0$, both for the strongly localised case $K = 128$ and the weakly localised case $K = 40$.

BV exhibit again the largest RMS error in both the weakly localised $K = 40$ and the strongly localised $K = 128$ case. The RMS error exhibits a local minimum for a designated perturbation size $\delta_{\text{min}}$. For $K = 40$ we find $\delta_{\text{min}} = 0.05$ for BVs for both lead times $\tau = 2.0$ and $\tau = 4.0$. For $K = 128$ we find for BVs $\delta_{\text{min}} = 0.08$ and $\delta_{\text{min}} = 0.12$ for lead times $\tau = 2.0$ and $\tau = 4.0$, respectively. Hence BV ensembles exhibit their minimal RMS error at perturbation sizes which are not consistent with the average analysis error of 0.10 and 0.18 for $tK = 40$ and $K = 128$, respectively. For perturbation sizes around $\delta_{\text{min}}$, BVs are collapsed to a single member with ensemble dimension $\bar{D}_{\text{ens}} = 1$ at both lead times. As in the idealistic truth experiment presented in Section 6, the RMS error assumes unacceptable high values for perturbation sizes which allow for a non-collapsed BV ensemble with
\( \mathcal{D}_{\text{ens}} > 1 \) for \( \delta \gtrsim 8 \) (\( \delta \gtrsim 5 \) for \( K = 128 \)).

SPBV and RDBV ensembles exhibit a significant increase in forecast skill and consistently have smaller RMS error for all values of the perturbation size \( \delta \) and lead times. For the smaller lead time \( \tau = 2 \) we observe that all ensembles incur the same RMS error for sufficiently small values of the perturbation size \( \delta \). This is because for perturbation sizes \( \delta \) which are significantly smaller than the analysis error, all ensembles are highly under-dispersive and the forecast error is dominated by the analysis error. As in the idealised experiment presented in Section 6, SPBV and RDBV ensembles perform almost identically in the weakly localised case \( K = 40 \), due to their similar spatial structures, and both ensembles perturb significantly across the whole domain capturing the regions of non-trivial analysis error. Both the SPBV and RDBV ensembles feature an RMS error minimum at approximately \( \delta = 0.1 \) for all lead times, matching the size of the typical analysis error. This indicates that the ensembles are well-adapted to capturing the analysis error uncertainties in addition to capturing the dynamic error growth. In the strongly localised case \( K = 128 \), RDBVs consistently exhibit smaller forecast RMS errors compared to SPBVs, as already seen in the idealised simulations discussed in Section 6. In the strongly localised case, the optimal perturbation size associated with the smallest RMS error depends on the lead time \( \tau \) for SPBVs and RDBVs. For lead times \( \tau = 2.0 \) we find \( \delta_{\text{min}} = 0.18 \) for SPBVs and \( \delta_{\text{min}} = 0.22 \) for RDBVs, consistent with the average analysis error of 0.18. For \( \tau = 4.0 \) we find \( \delta_{\text{min}} = 0.39 \) for SPBVs and \( \delta_{\text{min}} = 0.46 \) for RDBVs, which are both inconsistent with the average analysis error. Hence, in the strongly localised case the optimal perturbation size \( \delta_{\text{min}} \) of SPBVs and RDBVs does not match the average analysis for all lead times. This suggests that we may not be efficiently capturing the uncertainties of the analysis.

The dependency of the RMS error on the noise strength \( \sigma \) used to generate SPBVs is similar to the one in the idealised situation depicted in Figure 16, albeit the transition from the RMS error associated with classical BVs for \( \sigma = 0 \) to the asymptotic RMS value for \( \sigma \to \infty \) is much more rapid and is reached already for \( \sigma = 1.25 \) (not shown).

The ensemble RMS spread \( S \) for BVs, SPBVs and RDBVs as a function of the perturbation size \( \delta \) is shown in Figure 23. It is again clearly seen that classical BVs are deficient in RMS spread. Contrary to the idealised simulations discussed in Section 4, classical BV ensembles exhibit a non-vanishing spread for small values of \( \delta \) when the ensemble dimension is \( \mathcal{D}_{\text{ens}} = 1 \). This is entirely due to the chosen set-up of using pairs of positive and negative BVs, and is not indicative of any non-trivial diversity of the ensemble. Once \( \mathcal{D}_{\text{ens}} > 1 \) (cf. Figure 11) the RMS spread of BVs increases significantly. SPBVs and RDBVs exhibit significantly larger RMS spread compared to BVs. As for the RMS error, the differences between the RMS spread of RDBVs and SPBVs is more pronounced in the strongly localised case \( K = 128 \), reflecting the reduced ensemble space of SPBVs which are generated from a collapsed strongly localised BV by multiplicative perturbations, preserving the localisation. In the weakly localised case \( K = 40 \) the smaller RMS spread of SPBVs compared to RDBVs implies that SPBVs achieve the same forecast skill with less ensemble spread. In the strongly localised case \( K = 128 \) the increased ensemble spread of RDBVs positively impacts on their forecast skill and their RMS error.
7.2. **Reliability.** The error-spread ratio, parameterised by lead time $\tau$, is shown in Figure 24. The markers indicate the lead times of $\tau = 2.0$, $\tau = 3.0$ and $\tau = 4.0$. We note that each curve was obtained using a different value of the perturbation size $\delta$, which corresponds to the optimal perturbation size producing the smallest RMS error at lead time $\tau = 4.0$ (cf. Figure 22). For the weakly localised case $K = 40$ both SPBVs and RDBVs are close to the ideal error-spread ratio of 1, suggesting a reliable ensemble. We point out that since the ensemble size $N = 10$ is relatively small, the error-spread curves lie just above the one-to-one ratio due to finite-size sampling error effects. On the other hand, for the strongly localized case $K = 128$ SPBV ensembles are over-dispersive for small lead times $\tau \leq 2$, becoming under-dispersive for lead times $\tau > 2$. RDBV ensembles are seen to be over-dispersive for all lead times $\tau \leq 4.0$. 

![Figure 22](image-url)
Figure 23. RMS spread $S$ for forecast ensembles generated from the analysis as a function of $\delta$ for each ensemble generation method for fixed lead times $\tau = 2.0$ and $\tau = 4.0$. Top: $K = 40$. Bottom: $K = 128$.

Talagrand histograms are shown for each of the three forecast ensembles, averaged over all sites, in Figure 25, for lead times $\tau = 2.0$ and $\tau = 4.0$. Each histogram was again obtained using the perturbation size $\delta$ corresponding to the respective minimal RMS error (cf. Figure 22). Consistent with the results on the error-spread ratio above, the Talagrand diagrams show that SPBV and RDBV ensembles are reliable with a flat histogram in the weakly localised case $K = 40$. The reliability of the stochastically modified BV ensembles is linked to the fact that they generate non-trivial variance in regions of non-vanishing analysis error.

On the other hand, in the strongly localised case $K = 128$, when there is a strong discrepancy between the spatial structure of the analysis error and all of the SPBV ensemble members, SPBV ensembles do not lead to a flat Talagrand histogram, indicating an unreliable under-dispersive ensemble. We observe a high probability for the true state to lie outside the ensemble for both lead times. Remarkably, the interior bins of the histogram are
Figure 24. RMS error vs RMS spread for forecast ensembles generated from the analysis, parameterised by increasing lead times from $\tau = 0$ to $\tau = 5.0$. Each ensemble was generated using a perturbation sizes $\delta_{\text{min}}$ corresponding to minimal RMS error for each ensemble type at lead time $\tau = 4.0$ (cf. Figure 22). The markers indicate the specific lead times $\tau = 2.0$, $\tau = 3.0$ and $\tau = 4.0$. The grey dot-dashed line indicates a one-to-one ratio of RMS error and RMS spread, corresponding to a reliable ensemble. Top: $K = 40$. Bottom: $K = 128$.

relatively evenly populated and we do not observe the “U” shape typically associated with under-dispersive ensembles. The unusual shape of the Talagrand histogram in the strongly localised case can be understood as follows. Consider an arbitrary component $k$ of an SPBV away from the localised region of the parent BV, which is not significantly perturbed. If $D_{\text{ens}} = 1$, then none of the members of the SPBV ensemble will be able to perturb this site. The initial conditions associated with these SPBVs at site $k$ are therefore approximately equal to the analysis mean at that site. However, typically the true state is much further away in phase space from the analysis mean. After evolving the SPBV ensemble forward
in time, for reasonable lead times the ensemble has likely not developed sufficient spread to enclose the truth within its support. Hence, in the corresponding Talagrand diagram the truth falls into one of the exterior bins. This explains the peaks at the edge of the Talagrand histogram of SPBVs observed in Figure 25. On the other hand, the non-trivial components of an SPBV corresponding to the localised region have comparable magnitude to that of the analysis errors. This ensures that there are several components of the L96 model for which the true state is contained within the ensemble, contributing to the evenly distributed tally marks in the middle of the Talagrand histogram.

RDBV ensembles display an unusual shape of the rank histogram for $\tau = 2.0$ with two distinct modes in the strongly localised case $K = 128$. This is again linked to the mismatch between the spatial structure of localised BVs and the analysis error. Individual RDBV ensemble member do not efficiently sample the analysis errors since each individual RDBV is localised. On the other hand it is likely that each site will be significantly perturbed by at least one of the RDBV members, implying that the true state will rarely be an outlier in the context of a Talagrand histogram. This combination of under- and over-dispersiveness leads to the bimodal structure observed in Figure 25 for RDBVs.

We remark that in the strongly localised case $K = 128$, increasing the perturbation size $\delta$ does not mitigate the issue of unreliability. We found that the values of $\delta$ needed to generate a flat Talagrand histogram feature significantly larger forecast errors (not shown). Likewise, improving the accuracy of the observations does not allow for SPBVs and RDBVs to be reliable, but the associated smaller analysis error only causes the poor reliability to occur for an associated smaller optimal perturbation size $\delta$.

In the idealised case, the noise strength $\sigma$ used to generate SPBV ensembles could be tuned to obtain reliable SPBV ensembles. Naively we may expect that increasing $\sigma$ will also lead to more reliable SPBV ensembles in the case when ensembles are generated from the analysis. However, in the strongly localised case $K = 128$ increasing $\sigma$ only increases the variance within the localised region and does not affect the components outside the region. As for the RMS forecast error, we find that reliability measures are insensitive to changes of $\sigma$ and reliability cannot be achieved by tuning $\sigma$. We find that for all values of $\sigma$ SPBV ensembles are under-dispersive. The error-spread ratio and Talagrand histogram are shown in Figures 26 and 27 respectively for $\sigma = 0.25$, $\sigma = 1.25$ and $\sigma = 5.0$. Again, the error-spread ratio and Talagrand histograms are obtained for the respective optimal values of $\delta$ for which the RMS error attains its minimum value for each ensemble type. The reliability of SPBV ensembles saturates for $\sigma \geq 1.25$ and is more under-dispersive for $\sigma < 1.25$ for both dimension sizes $K$.

We remark that the property of an ensemble to be dynamically adaptive, i.e. their relationship with covariant Lyapunov vectors (cf. Section 34.1) and that they are conditioned on the current state, does not seem to be necessarily promoting improved forecast skill and reliability. In fact, the dynamically non-adapted RDBVs perform better than the dynamically adapted SPBVs with regards to forecast skills for both $K = 40$ and $K = 128$, and in the weakly localised case $K = 40$ they are also slightly more reliable.
Figure 25. Talagrand diagrams for forecast ensembles generated from the analysis for lead times $\tau = 2.0$ and $\tau = 4.0$. Each ensemble was generated using a perturbation sizes $\delta_{\text{min}}$ corresponding to minimal RMS error for each ensemble type and lead time (cf. Figure 22). Top: $K = 40$. Bottom: $K = 128$.

8. Discussion and outlook

We have extended the framework of stochastically modified bred vectors from the original multi-scale setting developed in [16] to the case of systems without scale separation. We considered here two stochastic modifications, SPBVs which preserve any eventual localisation of the their parent BVs and their spatial correlation structure, and RDBVs which do not do so. SPBVs were constructed to sample the probability density function conditioned on the current state whereas RDBVs are not conditioned on the current state but may evolve into future states which do not reliably estimate the probability density function at a given lead time. The difference in construction renders SPBVs dynamically adapted in the sense that they project onto dynamically relevant covariant Lyapunov vectors whereas RDBVs are not dynamically adapted. Using numerical simulations of the single scale Lorenz 96 model we have shown that SPBVs and RDBVs successfully mitigate the collapse to a single ensemble member of classical BVs with significantly increased ensemble dimension for perturbation sizes $\delta$ in the range of typical analysis errors. Related to this, the forecast skill - as measured by the RMS error – and the ensemble reliability – probed by the error-spread ratio and the Talagrand diagram – are markedly improved by the stochastic modifications.
We identified the property of localisation of fast growing perturbations which is often observed in spatially extended systems to be a crucial aspect for the performance of stochastically modified BVs. Whereas localisation is advantageous to condition on the current state, it is detrimental in allowing the ensemble to perturb spatial regions of non-vanishing analysis error which are outside the localised region. This causes SPBVs to be under-dispersive (independent of the noise strength). RDBVs exhibit the better forecast skill, despite not being dynamically adapted. In the weakly localised case, RDVs and SPBVs perform equally well, and behave (per construction) statistically similarly, and both ensembles significantly improve the forecast skill and reliability of classical BV ensembles.

To counteract the detrimental effect of localisation in SPBVs one could apply additive noise at all sites outside the active localised region, similar to the method proposed in [18].
We tried this in the L96 model but did not find that it overcame the problem. The level of noise required to account for the analysis error was found to be such that the noise to BV-signal ratio was too large and the perturbation would be close to a Gaussian random perturbation. This may be though an artefact of the L96 model and the addition of spatially homogenous noise on SPBVs may still mitigate against the problem of localisation in more complex models.

We would like to stress that our work only considers bred vectors here as a method for probabilistic forecasting and we are concerned with improving the breeding method. We do not attempt to compare different methods such as ensemble Kalman filter ensembles, singular vectors and other methods, and to determine their individual merits. Which ensemble will perform optimally as a forecast ensemble is in fact situation dependent, as pointed out, for example, recently by [34]. The authors found that in a coupled atmosphere ocean model bred vectors, with perturbation sizes tuned to capture the tropical Pacific thermocline variability, are the most effective choices for ensemble initialization and ENSO forecasting, compared to ensembles based on ETKFs.

We caution the reader that one cannot simply extrapolate the bad performance of BVs observed here in the setting of the L96 toy model to an operational setting. In realistic operational forecasting situations, uncertainty in saturated sub-synoptic processes such as convective events often generate sufficient variability in the synoptic scales, and thereby...
prevent BVs from collapsing onto a single BV \cite{48}. However, our work shows how to improve BVs and mitigate against under-dispersive BV ensembles \cite{36} without significant additional computational cost, and our work shows that caution needs to be taken in case of strongly localised error growth in situations without strong scale separation.

**Acknowledgements**

BG thanks Diego Pazó and Juanma López for stimulating discussions and for their hospitality. BG acknowledges the support of an Australian Postgraduate Award. GAG acknowledges support from the Australian Research Council, grant DP180101385.

**Appendix**

9. **Ensemble Kalman filter**

We briefly describe the ensemble Kalman filter used to obtain the numerical results presented in Section 7. For more details on data assimilation and ensemble filters the reader is referred to textbooks such as \cite{21,13,44,43,4}. In ensemble Kalman filters (EnKF), proposed by \cite{14}, an ensemble with $N$ members $x_n \in \mathbb{R}^D \times \mathbb{N}$ is propagated by the full nonlinear dynamics $\dot{X} = f(X)$ with $f(X) = [f(x_1), f(x_2), \ldots, f(x_N)] \in \mathbb{R}^{D \times N}$. The ensemble is decomposed into its mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

and its ensemble deviation matrix

$$X' = X - \bar{x}e^T,$$

where $e = [1, \ldots, 1]^T \in \mathbb{R}^N$. The ensemble deviation matrix $X'$ is used to provide a Monte-Carlo estimate of the forecast covariance matrix

$$P_f(t) = \frac{1}{N-1} X'(t) X'(t)^T \in \mathbb{R}^{D \times D}.$$  

In addition to the forecast ensemble we are also given observations $y_o \in \mathbb{R}^d$ which we express as a perturbed truth with

$$y_o(t_i) = Hx(t_i) + r_o,$$

where the observation operator $H : \mathbb{R}^D \to \mathbb{R}^d$ maps from the whole space into observation space, and $r_o \in \mathbb{R}^d$ is i.i.d. observational Gaussian noise with associated error covariance matrix $R_o$ and zero mean.

Given a forecast $X_f = X(t_i - \epsilon)$ of a chaotic system and its associated forecast error covariance matrix (also known as the *prior*) $P_f(t_i - \epsilon)$ as well as noisy observations $y_o(t_i)$, data assimilation aims to find the best estimate of the system and updates a forecast into a so-called analysis (also known as the *posterior*). We adopt the convention that evaluation at times $t = t_i - \epsilon$ evaluates a quantity before taking observations $y_o$, taken at $t = t_i$, into account in the analysis step, and evaluation at times $t = t_i + \epsilon$ evaluates quantities after the analysis step when the observations have been taken into account.
In the first step of the analysis the forecast mean $\bar{x}_f$ is updated to the analysis mean
\begin{equation}
\bar{x}_a = \bar{x}_f - K [H \bar{x}_f - y_o],
\end{equation}
where the Kalman gain matrix is defined as
\begin{equation}
K = P_f H^T (H P_f H^T + R_o)^{-1}.
\end{equation}

The analysis covariance $P_a$ is given by
\begin{equation}
P_a = (I_d - K H) P_f.
\end{equation}

To calculate an ensemble $X_a$ which is consistent with the analysis error covariance $P_a$ in the sense that the ensemble satisfies
\begin{equation}
P_a = \frac{1}{N - 1} X_a X^T_a,
\end{equation}
we use the method of deterministic ensemble square root filters which expresses the analysis ensemble as a linear combination of the forecast ensemble. In particular we use the method proposed by [46, 50], the so called Ensemble Transform Kalman Filter (ETKF). A new forecast $X(t_{i+1} - \epsilon)$ is then obtained by propagating $X_a(t_i + \epsilon)$ with the full nonlinear dynamics to the next time of observation, where a new analysis cycle will be started.

A common problem encountered with ensemble Kalman filters is filter divergence which refers to the problem that in finite ensembles the estimated forecast error covariance $P_f$ may be too small, potentially prohibiting the analysis to be corrected towards incoming observations, which renders the analysis to be effectively a free running forecast. This underestimation of the error covariance is commonly mitigated by multiplying the error covariance with a so called inflation factor [2]. In our simulations we chose sufficiently large ensembles such that inflation was not needed.

To seed an ensemble of bred vectors to be used in a subsequent ensemble forecast from an analysis, the perturbation size $\delta$ for the bred vector ideally would be chosen in accordance with the uncertainty of the analysis with $\delta = \sqrt{\text{Tr}P^a}$.

**References**

[1] Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9 (7), 1518–1530.

[2] Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127 (12), 2741–2758.

[3] Annan, J. D., 2004: On the orthogonality of bred vectors. *Monthly Weather Review*, 132 (3), 843–849, doi:10.1175/1520-0493(2004)132<0843:OTOOBV>2.0.CO;2.

[4] Asch, M., M. Bocquet, and M. Nodet, 2016: *Data Assimilation*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[5] Atencia, A., and I. Zawadzki, 2017: Analogs on the Lorenz attractor and ensemble spread. *Monthly Weather Review*, 145 (4), 1381–1400, doi:10.1175/MWR-D-16-0123.1, URL https://doi.org/10.1175/MWR-D-16-0123.1

[6] Balci, N., A. L. Mazzucato, J. M. Restrepo, and G. R. Sell, 2012: Ensemble dynamics and bred vectors. *Monthly Weather Review*, 140 (7), 2308–2334, doi:10.1175/MWR-D-10-05054.1.
[7] Bowler, N. E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus A*, 58 (5), 538–548, doi:10.1111/j.1600-0870.2006.00197.x.

[8] Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Bladé, 1999: The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, 12 (7), 1990–2009.

[9] Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133 (5), 1076–1097, doi:10.1175/MWR2905.1. URL https://doi.org/10.1175/MWR2905.1 https://doi.org/10.1175/MWR2905.1.

[10] Cai, M., E. Kalnay, and Z. Toth, 2003: Bred vectors of the Zebiak-Cane model and their potential application to ENSO predictions. *Journal of Climate*, 16 (1), 40–56.

[11] Cheng, Y., Y. Tang, P. Jackson, D. Chen, and Z. Deng, 2010: Ensemble construction and verification of the probabilistic ENSO prediction in the LDEO5 model. *Journal of Climate*, 23 (20), 5476–5497, doi:10.1175/2010JCLI3453.1.

[12] Corazza, M., and Coauthors, 2003: Use of the breeding technique to estimate the structure of the analysis “errors of the day”. *Nonlinear Processes in Geophysics*, 10 (3), 233–243.

[13] Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, 21 (6), 739–759, doi:10.1111/j.2153-3490.1969.tb00483.x.

[14] Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99 (C5), 10143–10162.

[15] Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. Springer, New York.

[16] Giggins, B., and G. A. Gottwald, 2019: Stochastically perturbed bred vectors in multiscale systems. *Quarterly Journal of the Royal Meteorological Society*, 145, 642–658.

[17] Ginelli, F., P. Poggi, A. Turchi, H. Chaté, R. Livi, and A. Politi, 2007: Characterizing dynamics with covariant Lyapunov vectors. *Phys. Rev. Lett.*, 99, 130601, doi:10.1103/PhysRevLett.99.130601.

[18] Greybush, S. J., E. Kalnay, M. J. Hoffman, and R. J. Wilson, 2013: Identifying Martian atmospheric instabilities and their physical origins using bred vectors. *Quarterly Journal of the Royal Meteorological Society*, 139 (672), 639–653, doi:10.1002/qj.1990.

[19] Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta/RSM short-range ensemble forecasts. *Monthly Weather Review*, 125 (6), 1312–1327.

[20] Herrera, S., D. Pazó, J. Fernández, and M. A. Rodríguez, 2011: The role of large-scale spatial patterns in the chaotic amplification of perturbations in a Lorenz-96 model. *Tellus A*, 63 (5), 978–990, doi:10.1111/j.1600-0870.2011.00545.x.

[21] Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge.

[22] Karimi, A., and M. R. Paul, 2010: Extensive chaos in the lorenz-96 model. *Chaos: An interdisciplinary journal of nonlinear science*, 20 (4), 043105.

[23] Keller, J. D., A. Hense, L. Kornblueh, and A. Rhodin, 2010: On the orthogonalization of bred vectors. *Weather and Forecasting*, 25 (4), 1219–1234, doi:10.1175/2010WAF2223341.1.

[24] Kuptsov, P. V., and U. Parlitz, 2012: Theory and computation of covariant Lyapunov vectors. *Journal of Nonlinear Science*, 22 (5), 727–762, doi:10.1007/
s00332-012-9126-5.

[25] Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, 102 (6), 409–418, doi:10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

[26] Leutbecher, M., and T. Palmer, 2008: Ensemble forecasting. *Journal of Computational Physics*, 227 (7), 3515 – 3539, doi:https://doi.org/10.1016/j.jcp.2007.02.014.

[27] Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17 (3), 321–333, doi:10.3402/tellusa.v17i3.9076, URL https://doi.org/10.3402/tellusa.v17i3.9076.

[28] Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on predictability Vol. 1*, T. Palmer, Ed., ECMWF, Reading, UK, 1–18.

[29] Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55 (3), 399–414.

[30] Magnusson, L., M. Leutbecher, and E. Källén, 2008: Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF ensemble prediction system. *Monthly Weather Review*, 136 (11), 4092–4104.

[31] Newman, C. E., P. L. Read, and S. R. Lewis, 2004: Investigating atmospheric predictability on Mars using breeding vectors in a general-circulation model. *Quarterly Journal of the Royal Meteorological Society*, 130 (603), 2971–2989, doi:10.1256/qj.03.209.

[32] Ng, G.-H. C., D. McLaughlin, D. Entekhabi, and A. Ahanin, 2011: The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A*.

[33] Oczkowski, M., I. Szunyogh, and D. Patil, 2005: Mechanisms for the development of locally low-dimensional atmospheric dynamics. *Journal of the Atmospheric Sciences*, 62 (4), 1135–1156.

[34] O’Kane, T. J., and Coauthors, 2019: Coupled data assimilation and ensemble initialization with application to multiyear ENSO prediction. *Journal of Climate*, 32 (4), 997–1024, doi:10.1175/JCLI-D-18-0189.1, URL https://doi.org/10.1175/JCLI-D-18-0189.1.

[35] O’Kane, T. J., and J. S. Frederiksen, 2008: A comparison of statistical dynamical and ensemble prediction methods during blocking. *Journal of the Atmospheric Sciences*, 65 (2), 426–447, doi:10.1175/2007JAS2300.1.

[36] Palmer, T., 2019: The ECMWF ensemble prediction system: Looking back (more than) 25?years and projecting forward 25?years. *Quarterly Journal of the Royal Meteorological Society*, 145 (S1), 12–24, doi:10.1002/qj.3383, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3383.

[37] Palmar, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bulletin of the American Meteorological Society*, 74 (1), 49–66, doi:10.1175/1520-0477(1993)074<0049:ERAPAT>2.0.CO;2, URL https://doi.org/10.1175/1520-0477(1993)074<0049:ERAPAT>2.0.CO;2.

[38] Patil, D., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Physical Review Letters*, 86 (26), 5878.
[39] Pazó, D., J. López, and M. Rodríguez, 2013: The geometric norm improves ensemble forecasting with the breeding method. Quarterly Journal of the Royal Meteorological Society, 139 (677), 2021–2032.

[40] Pazó, D., M. Rodríguez, and J. López, 2010: Spatio-temporal evolution of perturbations in ensembles initialized by bred, Lyapunov and singular vectors. Tellus A, 62 (1), 10–23, doi:10.1111/j.1600-0870.2009.00419.x.

[41] Pazó, D., M. A. Rodríguez, and J. M. López, 2011: Maximizing the statistical diversity of an ensemble of bred vectors by using the geometric norm. Journal of the Atmospheric Sciences, 68 (7), 1507–1512.

[42] Primo, C., M. Rodríguez, and J. Gutiérrez, 2008: Logarithmic bred vectors. A new ensemble method with adjustable spread and calibration time. Journal of Geophysical Research: Atmospheres, 113 (D5).

[43] Reich, S., and C. Cotter, 2015: Probabilistic forecasting and Bayesian data assimilation. Cambridge University Press, New York, x+297 pp.

[44] Simon, D. J., 2006: Optimal State Estimation. John Wiley & Sons, Inc., New York.

[45] Talagrand, O., 1999: Evaluation of probabilistic prediction systems. Workshop proceedings "Workshop on predictability", 20-22 October 1997, ECMWF, Reading, UK.

[46] Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. Monthly Weather Review, 131 (7), 1485–1490.

[47] Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. Bulletin of the American Meteorological Society, 74 (12), 2317–2330.

[48] Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. Monthly Weather Review, 125 (12), 3297–3319.

[49] Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. Journal of the Atmospheric Sciences, 60 (9), 1140–1158.

[50] Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? Monthly Weather Review, 132 (7), 1590–1605, doi:10.1175/1520-0493(2004)132(1590:WIBAEO)2.0.CO;2.

[51] Wilks, D. S., 2006: Statistical Methods in the Atmospheric Sciences. Elsevier, Oxford.

[52] Wolfe, C. L., and R. M. Samelson, 2007: An efficient method for recovering Lyapunov vectors from singular vectors. Tellus A: Dynamic Meteorology and Oceanography, 59 (3), 355–366, doi:10.1111/j.1600-0870.2007.00234.x.

[53] Yang, S.-C., C. Keppenne, M. Rienecker, and E. Kalnay, 2009: Application of coupled bred vectors to seasonal-to-interannual forecasting and ocean data assimilation. Journal of Climate, 22 (11), 2850–2870.