Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants

Reinhard Furrer[a], Thomas Bengtsson[b],[*]

[a]*Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO, USA*
[b]*Bell Labs, Statistics and Data Mining Department, Murray Hill, NJ, USA*

Abstract

This work studies the effects of sampling variability in Monte Carlo-based methods to estimate very high-dimensional systems. Recent focus in the geosciences has been on representing the atmospheric state using a probability density function, and, for extremely high-dimensional systems, various sample-based Kalman filter techniques have been developed to address the problem of real-time assimilation of system information and observations. As the employed sample sizes are typically several orders of magnitude smaller than the system dimension, such sampling techniques inevitably induce considerable variability into the state estimate, primarily through prior and posterior sample covariance matrices. In this article, we quantify this variability with mean squared error measures for two Monte Carlo-based Kalman filter variants: the ensemble Kalman filter and the ensemble square-root Kalman filter. Expressions of the error measures are derived under weak assumptions and show that sample sizes need to grow proportionally to the square of the system dimension for bounded error growth. To reduce necessary ensemble size requirements and to address rank-deficient sample covariances, covariance-shrinking (tapering) based on the Schur product of the prior sample covariance and a positive definite function is demonstrated to be a simple, computationally feasible, and very effective technique. Rules for obtaining optimal taper functions for both stationary as well as non-stationary covariances are given, and optimal taper lengths are given in terms of the ensemble size and practical range of the forecast covariance. Results are also presented for optimal covariance inflation. The theory is verified and illustrated with extensive simulations.

[*] Corresponding author.
E-mail addresses: furrer@ucar.edu (R. Furrer), tocke@stat.berkeley.edu (T. Bengtsson).

1. Introduction

Many modern geophysical problems are characterized by extremely large-scale systems and pose difficult challenges for real-time assimilation of observations and system information. One important application area of high-dimensional data analysis is numerical weather prediction (NWP), where solutions to the general (inverse) problem of combining data and model quantities are commonly required. For instance, to produce real-time weather forecasts (including hurricane and severe weather warnings), satellite and radar observations must be combined with time-integrated solutions of atmospheric and oceanic models. To this end, recent focus in NWP has been on representing the knowledge of the atmospheric state using a probability density function (pdf) [18,31,11]. This framework treats a weather forecast as a draw from a prior distribution, and seeks to update the knowledge of the atmospheric state once new weather observations are available. Various sample-based techniques have been developed to address the problem of real-time updating and forecasting (of this pdf) in extremely high-dimensional systems [32,2,27,55]. Motivated by the aforementioned research, this work studies errors due to sampling variability in Monte Carlo-based Kalman filter variants used for NWP. Our results are derived in the context of atmospheric data assimilation, but extend to large-scale multivariate systems with spatial structure and to the study of sample matrix variability in general.

The best known filtering (data assimilation) algorithm is in the context of Gaussian distributions and linear system dynamics, where the prior and posterior pdfs are described by the Kalman filter (Kf) recursion [35]. To address the heavy computational expense of the Kf recursions in very large-scale problems, Evensen [18] proposes the *ensemble* Kalman filter (enKf). Conceptually, the enKf implements Bayes theorem by perturbing a (Gaussian) forecast sample to produce a posterior sample with the correct first two moment structures. The enKf is optimal only in Gaussian settings, but because it samples using the empirical forecast distribution the method is know to have excellent non-Gaussian properties in various settings [8]. For more general non-Gaussian forecast distributions there exists a rich recent statistical literature based on sequential importance sampling methods (e.g. [17,40]). However, since the computational requirements of importance sampling methods increase rapidly with dimension, calculation of the posterior pdf can only be envisioned for systems with a relatively small number of degrees of freedom [39]. Thus, we note that for large-scale, non-linear and non-Gaussian systems, there exists only sub-optimal approaches to real-time sequential state estimation. We therefore focus on Monte Carlo-based filtering methods derived from the Kf update step, which asymptotically provide best linear prediction of the involved states as the sample size grows large.

The influence of sample variability may seriously deteriorate the quality of Monte Carlo-based update techniques, especially when the sample size is small (or moderate) compared to system dimension. In the case of ensemble-based Kf methods for NWP, the need to accurately estimate large-scale covariance structures from finite samples is a challenging task, and has led to the development of ad hoc tuning methods, e.g. covariance inflation and tapering [4,27,55]. As will be shown here, spurious sample correlations and the non-linear form of the Kalman gain matrix result in posterior samples with too little spread. To understand such effects, our work delineates the consequences of matrix sample variability in two enKf variants: the perturbed observations scheme [19,26] and the square-root enKf [55,49].

This article is structured as follows. Section 2 reviews the Kf recursions and introduces the (perturbed observation scheme) enKf as a Monte Carlo-based approximation to the Kf update step. A simple univariate example serves to illustrate bias effects in the gain due to sample variability, and analytical results show the bias to hold also in the multivariate case. In terms

of mean squared error (MSE) properties, Section 3 describes the effects of sampling variability on enKf-based forecast and update covariance matrices. Analytical results are delineated in terms of the eigen-structure of the forecast covariance matrix, system dimension, and sample size. Sections 4 and 5 give results to reduce the effects of sampling variability on enKf methods. Among several ideas, Section 4 introduces a new, non-distance-based taper function, and Section 5 provides results for optimal covariance inflation. Section 6 presents simulations. Our work is concluded with a discussion in Section 7. Proofs and other technical details are given in the Appendix.

2. Background and problem formulation

Weather forecasts are generated by large-scale numerical models that require initialization at the beginning of each forecast cycle and data assimilation is the problem of estimating initial conditions given the observations and the model. For instance, to obtain initial conditions for the next forecast cycle, satellite observations of up-welling radiation (related to vertical integrals of atmospheric temperature and water vapor) and of radar backscatter from the sea surface (related to surface winds), must be combined with previous numerical forecasts from atmospheric models.

Naturally, there exists a large body of research in NWP on how to select such initial conditions. One approach is given by variational techniques which (essentially) focus on producing a single estimate of the atmospheric state through non-linear optimization (e.g. [42,12]). These methods frame the assimilation problem in probabilistic terms (similarly to the methods discussed in this paper), and a statistician may think of the operational method of 4D-Var (e.g. [36]) as a fixed lag-smoother that, subject to dynamical constraints, produces the maximum likelihood estimator (or posterior mode) of the state at the end of the assimilation window. To be implemented, variational techniques require specification of a stationary, high-dimensional *background covariance* structure from a "finite" information set (e.g. [48]). Another approach to data assimilation in NWP is given by ensemble techniques which use Monte Carlo methods to obtain updated estimates of atmospheric quantities ([50,51,44,31]). Through a Bayesian formulation, these methods create initial conditions by sampling from a posterior density describing the atmospheric state. The resulting sequential (in time) propagation of a sample produces (in theory) flow dependent covariance structures.

Sequential modeling of dynamical systems consists of an *update-* and a *forecast* step. The goal of the update step is to modify an available forecast (prior) pdf for the system once new data is available. Then, in the forecast step, the updated pdf is propagated forward in time using known or approximate dynamical laws that are typically specified by partial differential equations. In the context of NWP, both updating the pdf given new observations and forecasting the pdf forward in time present formidable obstacles: the dimension of the state-vector in most oceanic and atmospheric models is extremely high, often including 10^6–10^8 components. Constructing a sample from the updated pdf in such high-dimensional systems is a challenging computational task and this update step is the primary focus of this paper.

The Kalman filter recursions are crucial to our work and are described next.

2.1. Kalman filter recursions

Let $\mathbf{x}_t \in \mathbb{R}^q$ represent the unobserved atmospheric state of the system at time t, and let the $\mathbf{y}_t \in \mathbb{R}^r$ denote a new set of weather observations. The data and the state are related by the

observation equation,

$$\boldsymbol{y}_t = \mathbf{H}_t \boldsymbol{x}_t + \boldsymbol{e}_t. \tag{1}$$

In (1), $\mathbf{H}_t$ is a linear observation operator, $\boldsymbol{e}_t \in \mathbb{R}^r$ is a Gaussian random vector (taken independent of $\boldsymbol{x}_t$) with mean $\mathbf{0}$ and variance $\mathbf{R}_t$, here denoted $\boldsymbol{e}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R}_t)$.

We wish to update our knowledge of the unobserved state $\boldsymbol{x}_t$ in light of the new data $\boldsymbol{y}_t$. Initial knowledge of the system is given by the conditional forecast distribution $p(\boldsymbol{x}_t | \mathcal{Y}_{t-1})$, where $\mathcal{Y}_{t-1}$ denotes all past data up to and including time $t - 1$. The standard Kf assumes $p(\boldsymbol{x}_t | \mathcal{Y}_{t-1}) \sim \mathcal{N}_q(\boldsymbol{x}_t^f, \mathbf{P}_t^f)$, where $\boldsymbol{x}_t^f$ and $\mathbf{P}_t^f$ represent the mean forecast and forecast error covariance matrix, respectively. The *update* step combines the forecast distribution and the new vector of observations $\boldsymbol{y}_t$, yielding the posterior distribution $p(\boldsymbol{x}_t | \mathcal{Y}_t)$. A straightforward application of Bayes theorem yields

$$p(\boldsymbol{x}_t | \mathcal{Y}_t) \sim \mathcal{N}_q(\boldsymbol{x}_t^a, \mathbf{P}_t^a), \tag{2}$$

where

$$\boldsymbol{x}_t^a = \boldsymbol{x}_t^f + \mathbf{K}_t(\boldsymbol{y}_t - \mathbf{H}_t \boldsymbol{x}_t^f),$$

and

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^f. \tag{3}$$

Here, $\mathbf{I}$ represents the identity matrix and $\mathbf{K}_t$ is the Kalman gain matrix given by

$$\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}_t' (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t' + \mathbf{R}_t)^{-1}.$$

(The superscript $'$ denotes matrix transpose.)

To obtain the forecast distribution for the next time step, $p(\boldsymbol{x}_{t+1} | \mathcal{Y}_t)$, the update distribution (2) is propagated using the system dynamics represented in the state equation

$$\boldsymbol{x}_{t+1} = \mathbf{G}(\boldsymbol{x}_t) + \boldsymbol{v}_t,$$

where $\boldsymbol{v}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{Q}_t)$. Note that if the system dynamics are linear, so that $\mathbf{G}(\boldsymbol{x}_t) = \mathbf{G}_t \boldsymbol{x}_t$ for some matrix $\mathbf{G}_t$, the forecast distribution $p(\boldsymbol{x}_{t+1} | \mathcal{Y}_t)$ will again be multivariate normal with closed forms for the mean and covariance

$$\boldsymbol{x}_{t+1}^a = \mathbf{G}_t \boldsymbol{x}_t^f,$$
$$\mathbf{P}_{t+1}^f = \mathbf{G}_t \mathbf{P}_t^a \mathbf{G}_t' + \mathbf{Q}_t. \tag{4}$$

However, since the subsequent methods approximate the forecast distribution through the propagation of a sample (ensemble), the closed form in (4) will not be used in our developments.

Next we describe the creation of the posterior sample.

2.2. The ensemble Kalman filter

In operational NWP the state-dimension q prevents direct implementation of the covariance recursions specified in (3) and (4). To deal with the computational expense, variants of the Kf that approximate the forecast and update steps by Monte Carlo methods have recently been developed. The basic sampling scheme, originally proposed by Evensen [18] and Evensen and van Leeuwen [19] in the context of oceanographic data assimilation, is given by the enKf.

The enKf assumes a sample of size n from the forecast distribution $p(\mathbf{x}_t | \mathcal{Y}_{t-1}) \sim \mathcal{N}_q(\mathbf{x}_t^f, \mathbf{P}_t^f)$. We denote the sample $\{\mathbf{x}_{t,i}^f\}$. Then, given a forecast ensemble member $\mathbf{x}_{t,i}^f$, the enKf algorithm generates synthetic data (by adding a perturbation to $\mathbf{y}_t$), and updates $\mathbf{x}_{t,i}^f$ according to

$$\mathbf{x}_{t,i}^a = \mathbf{x}_{t,i}^f + \mathbf{K}_t \left(\mathbf{y}_t + \boldsymbol{\varepsilon}_{t,i} - \mathbf{H}_t \mathbf{x}_{t,i}^f \right). \tag{5}$$

For (5), with $\boldsymbol{\varepsilon}_{t,i} \stackrel{\text{iid}}{\sim} \mathcal{N}_r(\mathbf{0}, \mathbf{R}_t)$ and $\boldsymbol{\varepsilon}_{t,i}$ independent of $\mathbf{x}_{t,j}^f$, $\forall i, j$, it can be shown that $\{\mathbf{x}_{t,i}^a\}$ is a sample from the update distribution given in (2) (e.g. [32,11]).

In practice, with $\hat{\mathbf{P}}_t^f$ a sample estimate of $\mathbf{P}_t^f$ based on $\{\mathbf{x}_{t,i}^f\}$, the update step (5) is implemented by replacing $\mathbf{K}_t$ by $\hat{\mathbf{K}}_t = \hat{\mathbf{P}}_t^f \mathbf{H}_t' (\mathbf{H}_t \hat{\mathbf{P}}_t^f \mathbf{H}_t' + \mathbf{R}_t)^{-1}$. We refer to the scheme where $\hat{\mathbf{K}}_t$ replaces $\mathbf{K}_t$ in (5) as the *sample-based* enKf. The resulting sampling scheme, which is further delineated in (8), is the primary focus of our paper, and henceforth, unless explicitly stated, we always refer to the sample-based enKf. If $\hat{\mathbf{P}}_t^f$ is a consistent estimator of $\mathbf{P}_t^f$, Slutzky's theorem (e.g. [7]) can be used to show that the sample-based enKf converges to that of the Kf as the ensemble size n converges to infinity. However, for small n and large q, $\hat{\mathbf{P}}_t^f$ will likely be a poor estimate of $\mathbf{P}_t^f$, and the influence of sample variability in $\hat{\mathbf{P}}_t^f$ will adversely affect the sample-based enKf update.

Many enKf techniques have been developed to improve the quality of the enKf update step, in particular to address the situation $n \ll q$, for which spurious sample covariances in $\hat{\mathbf{P}}_t^f$ need to be attenuated. Examples of update techniques addressing the adverse effects of sampling variability on filter performance are given by square-root enKf approaches [2,10,49], double enKf [32,38], and covariance localization approaches [27,3]. Moreover, as will be shown in Section 2.4 (see also Section 5), the non-linear form of the Kalman gain matrix typically results in $\mathbf{P}_t^a$ being underestimated. Ad hoc tuning methods to increase posterior spread are proposed by Anderson and Anderson [4] and Whitaker and Hamill [55]. We do not delineate the computational details of any of these filter variants here; rather, our focus is on the errors due to matrix sample variability in the enKf algorithm. To this end, we are interested in the effects of variability in $\{\mathbf{x}_{t,i}^f\}$ and $\{\boldsymbol{\varepsilon}_{t,i}\}$ on the enKf, and for the remainder of our work all expectations are taken with respect to (wrt) the forecast density, $\mathcal{N}_q(\mathbf{x}_t^f, \mathbf{P}_t^f)$, and the density of the data perturbations, $\mathcal{N}_r(\mathbf{0}, \mathbf{R}_t)$.

For a simple univariate system we next show the effects of sample variability induced by the enKf.

2.3. Motivating example

To illustrate the finite sample size effects of the enKf algorithm we consider a univariate system. In this setting, we can accurately approximate the bias of the (sample) Kalman gain and the posterior covariance, and delineate how to correct the bias by artificial inflation of the forecast spread.

We assume a univariate system with $\mathbf{H}_t = 1$ and $\mathbf{R}_t = 1$. For this setting, each scalar ensemble member $\mathbf{x}_{t,i}^f$ is updated according to (5) with $\hat{\mathbf{K}}_t = \hat{\lambda}_t / (\hat{\lambda}_t + 1)$, where $\hat{\lambda}_t$ represents the forecast ensemble sample variance $\sum_{i=1}^n (\mathbf{x}_{t,i}^f - \bar{\mathbf{x}}_t^f)^2 / (n - 1)$. Note that $\hat{\mathbf{K}}_t - \mathbf{K}_t = (\lambda_t + 1)^{-1} - (\hat{\lambda}_t + 1)^{-1}$ and, since $g(u) = (u + 1)^{-1}$ is strictly convex, $E(\hat{\mathbf{K}}_t) < \mathbf{K}_t$ by Jensen's inequality. With expectation evaluated wrt the distribution of $\mathbf{x}_{t,i}^f$, the bias can be approximated by a Taylor expansion of $g(\hat{\lambda}_t)$, yielding a third order bias of

$$E \left(\frac{\hat{\lambda}_t}{\hat{\lambda}_t + 1} - \frac{\lambda_t}{\lambda_t + 1} \right) = -\frac{2}{n - 1} \frac{\lambda_t^2}{(\lambda_t + 1)^3} + \mathcal{O}(n^{-2}). \tag{6}$$

Simulations show that the approximation is informative for sample sizes as small as $n = 4$ and a wide range of λ_t. Further, for this system $\mathbf{P}_t^a = \mathbf{K}_t$, and $\mathrm{E}(\hat{\mathbf{P}}_t^a) = \mathrm{E}(\hat{\mathbf{K}}_t)$ (see Proposition 1). Thus, the sample variance $\hat{\mathbf{P}}_t^a$ is a negatively biased estimate of the (true) posterior variance $\mathbf{P}_t^a$.

Following Anderson and Anderson [4], Whitaker and Hamill [55], the bias of $\hat{\mathbf{K}}_t$ and $\hat{\mathbf{P}}_t^a$ can be reduced by artificially inflating the sample forecast spread (variance), e.g. by replacing $\hat{\lambda}_t$ with $\rho \hat{\lambda}_t$ where $\rho > 1$. Then, the optimal ρ eliminating the bias term which involves λ_t in (6) is obtained by solving

$$\rho^3 - \rho^2 \left(1 - \frac{2}{\lambda_t} + \frac{2}{(n-1)\lambda_t} + \frac{2}{(n-1)\lambda_t^2} \right) - \rho \left(\frac{2}{\lambda_t} - \frac{1}{\lambda_t^2} \right) - \frac{1}{\lambda_t^2} = 0. \tag{7}$$

In the example given here the bias is maximized for $\lambda_t = 2$ and yields a simulated value of approximately 0.091 (representing 13.6% of the true value) for $n = 5$. For this choice of λ_t and n the bias is essentially removed by setting $\rho = 1.576$. Surprisingly, in this case, the sample variance $\hat{\lambda}_t$ must be increased by 57.6% to provide an approximately unbiased sample gain and a posterior sample $\{x_{t,i}^a\}$ with correct spread. It should be noted that (7) yields complicated solutions in terms of the parameters λ_t and n, but it is easily verified that $\rho = 1$ is the unique positive solution as n tends to infinity.

Next we generalize the results illustrated in this section to multivariate problems.

2.4. Multivariate case

For the multivariate case, we are interested in quantifying the effects of the enKf scheme on the sample covariance matrix of the posterior ensemble members. To describe the effects, we give a key decomposition of $\hat{\mathbf{P}}_t^a$ and use this result to motivate the need for bias-corrective measures in enKf, as well as to provide impetus for the study of high-dimensional sample covariance matrices in Sections 3–5.

Analogously to the univariate case, we now show that $\hat{\mathbf{P}}_t^a$ is negatively biased. To obtain this result we first decompose $\hat{\mathbf{P}}_t^a$ and argue that propagation of $\hat{\mathbf{P}}_t^f$ using the covariance recursion (3) produces a biased estimate of $\mathbf{P}_t^a$. Jensen's inequality then establishes the bias by showing $\mathrm{tr}\,\mathrm{E}(\hat{\mathbf{P}}_t^a) < \mathrm{tr}\,\mathbf{P}_t^a$.

Proposition 1. *Assume a state-space model as specified in Section* 2.1. *Let the sample Kalman gain matrix* $\hat{\mathbf{K}}_t$ *be defined using the covariance of the forecast ensemble, i.e.* $\hat{\mathbf{K}}_t = \hat{\mathbf{P}}_t^f \mathbf{H}_t' (\mathbf{H}_t \hat{\mathbf{P}}_t^f \mathbf{H}_t' + \mathbf{R}_t)^{-1}$, *where* $\hat{\mathbf{P}}_t^f$ *is the sample covariance of* $\{x_{t,i}^f\}$. *Also, let* $\hat{\mathbf{R}}_t$ *be the sample covariance of the data perturbations* $\{\varepsilon_{t,i}\}$, *and* $\hat{\mathbf{C}}_t$ *the sample covariance between the data perturbations and the forecast ensemble. With* $\hat{\mathbf{P}}_t^a = (n-1)^{-1} \sum_{i=1}^n (x_{t,i}^a - \bar{x}_t^a)(x_{t,i}^a - \bar{x}_t^a)'$ *the sample covariance matrix of the posterior ensemble* $\{x_{t,i}^a\}$, *generated according to*

$$x_{t,i}^a = x_{t,i}^f + \hat{\mathbf{K}}_t \left(y_t + \varepsilon_{t,i} - \mathbf{H}_t x_{t,i}^f \right), \quad i = 1, \ldots, n, \tag{8}$$

we have (a)

$$\hat{\mathbf{P}}_t^a = \hat{\mathbf{P}}_t^f + \hat{\mathbf{K}}_t (\mathbf{H}_t \hat{\mathbf{P}}_t^f \mathbf{H}_t' + \hat{\mathbf{R}}_t) \hat{\mathbf{K}}_t' - 2\hat{\mathbf{K}}_t \mathbf{H}_t \hat{\mathbf{P}}_t^f + (\mathbf{I} - \hat{\mathbf{K}}_t \mathbf{H}_t) \hat{\mathbf{C}}_t' \hat{\mathbf{K}}_t' + \hat{\mathbf{K}}_t \hat{\mathbf{C}}_t (\mathbf{I} - \mathbf{H}_t' \hat{\mathbf{K}}_t'), \tag{9}$$

and (b)

$$E(\hat{\mathbf{P}}_t^a) = E\left((\mathbf{I} - \hat{\mathbf{K}}_t\mathbf{H}_t)\hat{\mathbf{P}}_t^f\right). \tag{10}$$

The decomposition in (a) is obtained by expanding the outer product defining $\hat{\mathbf{P}}_t^a$, and the result in (b) is given by the independence of $\{\mathbf{x}_{t,i}^f\}$ and $\{\boldsymbol{\varepsilon}_{t,i}\}$. (Details are given in the Appendix).

It is important to distinguish the error covariance $\hat{\mathbf{P}}_t^a$ based on the sample scheme (8) from that produced by substituting $\hat{\mathbf{P}}_t^f$ and $\hat{\mathbf{K}}_t$ in (3). The posterior covariance matrix obtained through the latter method, i.e. $(\mathbf{I} - \hat{\mathbf{K}}_t\mathbf{H}_t)\hat{\mathbf{P}}_t^f$, will henceforth be referred to as the *propagated analysis covariance matrix*. (In Section 3.4 we show this method of producing a posterior covariance matrix to be equivalent to that obtained through use of the square-root enKf.)

The expectation given in (10) is not analytically tractable for arbitrary $\mathbf{H}_t$ and $\mathbf{R}_t$. However, using Proposition 1(b) we can show the following result.

Corollary 2. *Let* $\mathbf{H}_t'\mathbf{H}_t = \mathbf{I}$ *and* $\mathbf{R}_t = \sigma^2\mathbf{I}$. *Then,*

$$\operatorname{tr} E(\hat{\mathbf{P}}_t^a) < \operatorname{tr} \mathbf{P}_t^a.$$

Matrix algebra along with Jensen's inequality establishes the result. (See Appendix for technical details).

It may seem that the assumptions of Corollary 2 are restrictive. However, it is always possible define a projected state-space system $\{\tilde{\mathbf{x}}_t, \tilde{\mathbf{y}}_t\}$ by letting $\tilde{\mathbf{y}}_t = \mathbf{R}_t^{-1/2}\mathbf{y}_t$ and $\tilde{\mathbf{x}}_t = \mathbf{R}_t^{-1/2}\mathbf{H}_t\mathbf{x}_t$, and for this system the assumptions of the corollary are satisfied.

Corollary 2 shows that the enKf yields posterior ensemble members with too little spread when $\{\mathbf{x}_{t,i}^a\}$ is produced using (8). We emphasize that the lack of spread described here is due entirely to the inverse in the sample Kalman gain (as established by Jensen's inequality), and that this holds even as $E(\hat{\mathbf{P}}_t^f) = \mathbf{P}_t^f$. In the case of $\mathbf{H}_t = \mathbf{I}$ we have a similar corollary for the Kalman gain.

Corollary 3. *Let* $\mathbf{H}_t = \mathbf{I}$ *and* $\mathbf{R}_t = \sigma^2\mathbf{I}$. *Then,*

$$\operatorname{tr} E(\hat{\mathbf{K}}_t) < \operatorname{tr}(\mathbf{K}_t).$$

In Section 5, based on an expansion of $\hat{\mathbf{P}}_t^a$, we show that the bias induced by Corollary 3 can be effectively removed.

To further understand the effects of sample matrix variability in high-dimensional Kf settings, including the effects of spurious sample correlations, we next study the MSE properties of $\hat{\mathbf{P}}_t^f$ and $\hat{\mathbf{P}}_t^a$ as a function of the eigen-structure of $\mathbf{P}_t^f$, system dimension, and sample size.

3. Sample variability in Kalman filter covariance matrices

In this section we develop expressions for the MSE of $\hat{\mathbf{P}}_t^f$ and $\hat{\mathbf{P}}_t^a$. Exact results for the MSE of $\hat{\mathbf{P}}_t^f$ are given in Section 3.2 in terms of the eigen-structure of $\mathbf{P}_t^f$, system dimension, and sample size. Section 3.3 is devoted to the posterior covariance matrix. An orthogonal decomposition of the MSE of $\hat{\mathbf{P}}_t^a$ along with a matrix series expansion are used to provide a lower bound of the error, again as a function of eigen-structure, q and n. Based on these results, Section 3.4 discusses the effects of sample variability on the so-called square-root enKf. Asymptotic results for $q \to \infty$ but with fixed n are discussed, and conditions under which errors are bounded are given throughout.

It should be noted that the results presented in this section pertain to general, un-tapered sample covariance matrices. (Tapering results are considered in Section 4).

We start by defining an appropriate distance measure and by introducing notation for the eigen-structure of $\mathbf{P}_t^f$.

3.1. Distance measure

To determine the similarity of $\hat{\mathbf{P}}_t^f$ and $\mathbf{P}_t^f$, we define a distance measure based on the MSE of $\hat{\mathbf{P}}_t^f$, i.e. let

$$\|\mathbf{P}_t^f - \hat{\mathbf{P}}_t^f\|_{\mathrm{E}}^2 = \mathrm{E}\left(\mathrm{tr}\left((\mathbf{P}_t^f - \hat{\mathbf{P}}_t^f)^2\right)\right). \tag{11}$$

As before, expectation is taken with respect to the forecast distribution. To simplify the notation we henceforth drop the time index t.

Let $\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'$ be the spectral decomposition of $\mathbf{P}^f$, where $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_q > 0$ of $\mathbf{P}^f$ as entries and where $\mathbf{\Gamma} = (\gamma_{ij})$ contains the associated eigenvectors γ_j in the columns. We take the forecast members $\{\mathbf{x}_i^f\}$ to be zero-mean Gaussian with covariance matrix $\mathbf{P}^f$. Let $\hat{\mathbf{P}}^f = \sum_{i=1}^n \mathbf{x}_i^f \mathbf{x}_i^{f'}/n$ denote the unbiased estimator of $\mathbf{P}^f$ of rank $\min(n, q)$. To evaluate the distance measure defined by (11) we set $\hat{\mathbf{\Lambda}} = \mathbf{\Gamma}'\hat{\mathbf{P}}^f\mathbf{\Gamma}$, a rotation necessary to obtain unbiased estimates of the eigenvalues $\{\lambda_i\}$. Note that we are merely using the rotation in our developments, and do not assume $\mathbf{\Gamma}$ to be known. Thus, $\hat{\lambda}_{ij}$ is an unbiased estimator of $\lambda_{ij} = \lambda_i \delta_{ij}$, where δ_{ij} is the Kronecker-delta. However, $\hat{\mathbf{\Lambda}} = (\hat{\lambda}_{ij})$ is not diagonal. Further, $\hat{\lambda}_{ii}$ does not equal the ith sample eigenvalue obtained by the eigenvalue/eigenvector decomposition of $\hat{\mathbf{P}}^f$. Finally, to be employed in Section 3.3, we use the spectral radius $\varrho(\mathbf{A})$, which is the maximum eigenvalue for positive definite matrices $\mathbf{A}$.

The next two sections develop expressions for $\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2$ and $\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|_{\mathrm{E}}^2$ in terms of the eigen-structure of $\mathbf{P}^f$, q, and n.

3.2. MSE of forecast covariances

Although the asymptotic properties of sample eigenvalues are well-established [5,6], the distributional properties are complicated for finite samples [52]. In particular, when $n/q \to 1$ (but $n > q$) the sample eigen-structure is severely biased [34]. To complicate matters $n \ll q$ in our setting, and for this highly rank-deficient case we expect the sample singular values to be heavily biased. To see this, note simply that $\mathrm{E}\,\mathrm{tr}(\hat{\mathbf{P}}^f) = \sum_{i=1}^q \lambda_i$, from which it follows that the true eigenvalues must 'pile up' in the n positive singular values of $\hat{\mathbf{P}}^f$, consequently causing bias. In terms of the eigenvalues of $\mathbf{P}^f$ and n, the next result quantifies the MSE of $\hat{\mathbf{P}}^f$.

Lemma 7 of the Appendix can be used to derive the following exact form of the MSE of $\hat{\mathbf{P}}^f$:

$$\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2 = \frac{1}{n}\sum_{i=1}^q \lambda_i^2 + \frac{1}{n}\sum_{i,j=1}^q \lambda_i \lambda_j. \tag{12}$$

Since we need only a distribution of $\{\mathbf{x}_i^f\}$ satisfying the fourth moment equality given in (29) (see Appendix), it is possible to relax the hypothesis of normality. However, for other multivariate distributions it is not always possible to reduce the fourth moment to expressions in $\{\lambda_i\}$. Note that the above expression is valid for arbitrary n. We now consider the behavior of (12) as $q \to \infty$.
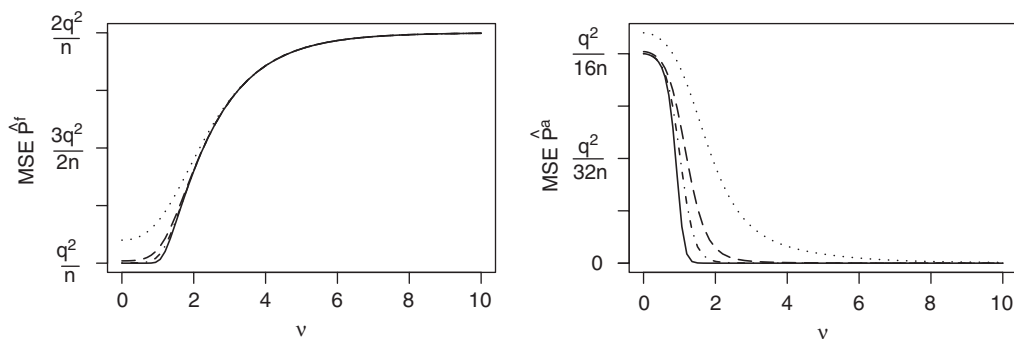
Fig. 1. Left panel: $\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2$ evaluated through (12) as a function of the decay rate of the eigen-structure of $\mathbf{P}^f$. The ith eigenvalue of $\mathbf{P}^f$ is given by $\lambda_i = ci^{-v}$, with c chosen such that $\mathrm{tr}(\mathbf{P}^f) = \mathrm{tr}(\mathbf{R}) = q$. The four curves in each plot correspond to $q = 10$, (dotted), $q = 10^3$ (dashed), $q = 10^6$ (dash-dotted), and $q \to \infty$ (solid). Right panel: Lower bound of $\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|_{\mathrm{E}}^2$ evaluated through the second order approximation (14).

To specifically evaluate (12) for high-dimensional, rank deficient sample forecast covariance matrices we assume that the eigenvalues of $\mathbf{P}^f$ can be described by reciprocal powers, i.e. with $v > 0$ let $\lambda_i = ci^{-v}$, $1 \leqslant i \leqslant q$, $v > 0$. For large q and $v > .5$ the sum in (12) is given by the well-known Riemann zeta function,

$$\lim_{q \to \infty} \|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2 = \frac{c^2}{n} \left(\zeta(2v) + \zeta^2(v) \right),$$

where $\zeta(v) = \int_0^\infty t^{v-1}/(e^t - 1)\, \mathrm{d}t$ (e.g. [1]). However, since it only has closed form for some special values (e.g. for $v = 2$, $\lim_{q \to \infty} \|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2 = 7c^2\pi^4/180n$), the zeta function is not convenient to work with, but if the sequence of eigenvalues $\{\lambda_k\}$ can be described as $\{g(k)\}$ for a continuous function $g(\cdot)$, the MSE of $\hat{\mathbf{P}}^f$ may be approximated using an integral approximation.

Based on evaluation of (12), the left panel of Fig. 1 illustrates how $\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2$ depends on the decay rate v and system dimension q. For all v, the forecast error matrix is scaled such that $\mathrm{tr}(\mathbf{P}^f) = q$, so that the total forecast error grows proportionally with system dimension. We see that a rapidly decaying spectrum represents the worst case scenario, with limit $\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2 = 2q^2/n$. As can be seen, a lower bound for $\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2$ is given by a flat spectrum. However, even for a flat spectrum, n has to be $\mathcal{O}(q^2)$ for bounded error growth.

The MSE for $\hat{\mathbf{P}}^a$ is described next.

3.3. MSE of analysis covariances

The previous evaluation of the MSE of $\hat{\mathbf{P}}^f$ is in terms of a standard bias and variance decomposition. A similar approach for $\hat{\mathbf{P}}^a$ would require evaluation of

$$\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|_{\mathrm{E}}^2 = \|\mathbf{P}^a - \mathrm{E}(\hat{\mathbf{P}}^a)\|^2 + \|\hat{\mathbf{P}}^a - \mathrm{E}(\hat{\mathbf{P}}^a)\|_{\mathrm{E}}^2. \tag{13}$$

However, as is the case with $\mathrm{E}(\hat{\mathbf{P}}^a)$, the terms in (13) are difficult to tract analytically for arbitrary systems (see Eq. (9), Proposition 1). Instead, we provide an alternative orthogonal decomposition from which a lower bound of $\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|_{\mathrm{E}}^2$ is derived.

Let $\{x_i^a\}$ be generated by the enKf algorithm described in (8) and let $\mathbf{Z} = \hat{\mathbf{K}}(\hat{\mathbf{R}} - \mathbf{R})\hat{\mathbf{K}}' + (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{C}}'\hat{\mathbf{K}}' + \hat{\mathbf{K}}\hat{\mathbf{C}}(\mathbf{I} - \mathbf{H}'\hat{\mathbf{K}}')$. Then, by Proposition 1, we have $\hat{\mathbf{P}}^a = (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}^f + \mathbf{Z}$. Thus, $\hat{\mathbf{P}}^a$ is expressed as a sum of the propagated analysis covariance and a residual term $\mathbf{Z}$. We can now state the following result.

Proposition 4. *Assume a state-space model as specified in Section* 2.1, *along with the enKf sampling scheme delineated in* (8). *Then, the propagated analysis covariance and the residual term* $\mathbf{Z} = \hat{\mathbf{K}}(\hat{\mathbf{R}} - \mathbf{R})\hat{\mathbf{K}}' + (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{C}}'\hat{\mathbf{K}}' + \hat{\mathbf{K}}\hat{\mathbf{C}}(\mathbf{I} - \mathbf{H}'\hat{\mathbf{K}}')$ *are independent, i.e.*

$$\mathrm{E}\left((\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}^f \mathbf{Z}\right) = 0.$$

The proof of the above proposition is based on the fact that $\{x_i^f\}$ and $\{\varepsilon_i\}$ are independent (see Appendix). The following orthogonal decomposition holds as a direct consequence of Proposition 4.

Corollary 5. $\left\| \mathbf{P}^a - \hat{\mathbf{P}}^a \right\|_{\mathrm{E}}^2 = \left\| \mathbf{P}^a - (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}^f \right\|_{\mathrm{E}}^2 + \left\| \mathbf{Z} \right\|_{\mathrm{E}}^2.$

Corollary 5 gives $\left\| \mathbf{P}^a - (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}^f \right\|_{\mathrm{E}}^2$ as a lower bound for the MSE $\left\| \mathbf{P}^a - \hat{\mathbf{P}}^a \right\|_{\mathrm{E}}^2$.

Due to the inverse in the Kalman gain, it is impossible to evaluate the lower bound analytically. A matrix series expansion is developed in Lemma 9 (see Appendix) to approximate $\left\| \mathbf{P}^a - (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}^f \right\|_{\mathrm{E}}^2$, and the proposition below utilizes the expansion to express the lower bound as a function of the eigenvalues of $\mathbf{P}^f$ and n.

Proposition 6. *Assume a state-space model as specified in Section* 2.1 *and set* $\mathbf{R} = \mathbf{H} = \mathbf{I}$. *Let* $\hat{\mathbf{P}}^f$ *be the sample covariance of* $\{x_i^f\}$. *Suppose* $\varrho\big((\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}})\mathbf{D}\big) < 1$, *where* $\mathbf{D} = (\boldsymbol{\Lambda} + \mathbf{I})^{-1}$. *Then, a third order approximation of* $\left\| \mathbf{P}^a - (\mathbf{I} - \hat{\mathbf{K}})\hat{\mathbf{P}}^f \right\|_{\mathrm{E}}^2$ *is given by*

$$\frac{1}{n}\left(\sum_i \frac{\lambda_i^2}{(\lambda_i + 1)^4} + \sum_{i,j} \frac{\lambda_i \lambda_j}{(\lambda_i + 1)^3 (\lambda_j + 1)} \right) \tag{14}$$

$$- \frac{2}{n^2}\left(\sum_i \frac{\lambda_i^3}{(\lambda_i + 1)^5} + \sum_{i,j} \frac{\lambda_i \lambda_j^2}{(\lambda_i + 1)^3 (\lambda_j + 1)^2} + \sum_{i,j} \frac{2\lambda_i^2 \lambda_j}{(\lambda_i + 1)^4 (\lambda_j + 1)} \right.$$

$$\left. + \sum_{i,j,\ell} \frac{\lambda_i \lambda_j \lambda_\ell}{(\lambda_i + 1)^3 (\lambda_j + 1)(\lambda_\ell + 1)} \right). \tag{15}$$

The above result represents the first three terms of the asymptotic expansion obtained through Lemma 9 (see Appendix). Note that (14) is a second order approximation, with error of order $\mathcal{O}(1/n^2)$. We believe that the error associated with the third order approximation is $\mathcal{O}(1/n^3)$. (This could be shown by evaluating higher-order terms, but these result in very lengthy expressions and will not be pursued here).

We acknowledge that not all samples satisfy $\varrho\big((\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}})\mathbf{D}\big) < 1$ and that there are instances for which the expansion does not hold. However, note that the employed rotation of $\hat{\mathbf{P}}^f$ using $\boldsymbol{\Gamma}$ ensures $\mathrm{E}(\boldsymbol{\Lambda} - \hat{\boldsymbol{\Lambda}}) = 0$, and that this difference is further suppressed by multiplication by $\mathbf{D}$. Moreover, as n increases, the number of samples for which the expansion holds increases. In particular,

Table 1

$\left\| \mathbf{P}^f - \hat{\mathbf{P}}^f \right\|_{\mathrm{E}}^2$ and the lower bound of $\left\| \mathbf{P}^a - \hat{\mathbf{P}}^a \right\|_{\mathrm{E}}^2$ for the two limit cases of the decay of the spectrum

Error matrix	Spectrum decay of $\mathbf{P}^f$	
	"Slow", $\lambda_i \equiv 1$	"Rapid", $\lambda_1 = q, \lambda_i \equiv 0, i > 1$
$\hat{\mathbf{P}}^f$	$\dfrac{q^2 + q}{n}$	$\dfrac{2q^2}{n}$
$\hat{\mathbf{P}}^a$	$\dfrac{q^2 + q}{16n} + \mathcal{O}(q^3 n^{-2})$	$\dfrac{2q^2}{(q+1)^4 n} + \mathcal{O}(q^{-2} n^{-2})$

since the forth moment exists and is finite, and as each individual term $|\hat{\lambda}_{ij} - \lambda_{ij}| = \mathcal{O}_p(n^{-1/2})$, we have

$$\left\| (\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D} \right\|_{\mathrm{E}}^2 = \sum_{i,j=1}^{q} d_i (\lambda_{ij} - \hat{\lambda}_{ij})^2 = \mathcal{O}_p(1/n).$$

Therefore, $\varrho\big((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D}\big)$ tends to zero as n increases, and convergence of the matrix expansion may be expected to be rapid, a claim supported by the simulations of Section 6.

A detailed picture of the relationship between the decay rate and the lower bound of $\left\| \mathbf{P}^a - \hat{\mathbf{P}}^a \right\|_{\mathrm{E}}^2$ is provided by the right panel of Fig. 1. Note that, when compared with $\hat{\mathbf{P}}^f$, the relationship between the MSE and the decay rate v is reversed for $\hat{\mathbf{P}}^a$. Thus, an enKf forecast with relatively few modes is more likely to produce an accurate analysis (posterior) ensemble. Table 1 further summarizes the MSEs of $\hat{\mathbf{P}}^f$ and the lower bound of $\hat{\mathbf{P}}^a$ for the case of a flat spectrum and the limiting case where the spectrum has only a single mode. As can be seen, a slow eigen-decay requires a sample size of order q^2 for bounded error growth.

The results presented in Table 1 are consistent with those of Künsch [37], who argues that in the univariate particle filter setting, as t grows, n has to be of order t^2 to ensure bounded error growth. However, in our work, because of the Gaussian assumptions, we are able to more exactly describe the error in terms of the system properties λ_i, q, and n.

The next section provides a justification why enKf algorithms which do not explicitly rely on the Monte Carlo technique of (8) are more efficient.

3.4. MSE of square-root enKf

An alternative approach to addressing the update step in high-dimensional systems is given by the so-called square-root Kalman filter (e.g. [55,49]). In our context, the general idea of the square-root filter algorithm is to manipulate a given forecast ensemble in order to produce a posterior sample with variance equal to that given by the Kalman filter.

With $\mathbf{F} = n^{-1/2}((\mathbf{x}_1^f - \bar{\mathbf{x}}^f), (\mathbf{x}_2^f - \bar{\mathbf{x}}^f), \ldots, (\mathbf{x}_n^f - \bar{\mathbf{x}}^f))$ define $\mathbf{A} = (\mathbf{I} - \mathbf{F}'\mathbf{H}'(\mathbf{H}\mathbf{F}\mathbf{F}'\mathbf{H}' + \mathbf{R})^{-1}\mathbf{H}\mathbf{F})^{1/2}$. Now consider post-multiplying the forecast deviations by $\mathbf{A}$ and let $\mathbf{U} = \mathbf{F}\mathbf{A}$, we have

$$\begin{aligned}
\mathbf{U}\mathbf{U}' &= \mathbf{F}\mathbf{A}\mathbf{A}'\mathbf{F} \\
&= \mathbf{F}(\mathbf{I} - \mathbf{F}'\mathbf{H}'(\mathbf{H}\mathbf{F}\mathbf{F}'\mathbf{H}' + \mathbf{R})^{-1}\mathbf{H}\mathbf{F})\mathbf{F}' \\
&= (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}^f.
\end{aligned}$$

We see that post-multiplication of the forecast deviations by $\mathbf{A}$ produces deviations $\mathbf{U}$ with an outer product equal to the propagated analysis covariance. Hence, this method of producing a posterior sample with the desired covariance does not require any data-perturbations.

By Corollary 5, sampling errors in $\hat{\mathbf{P}}^a$ due to the data-perturbations $\{\boldsymbol{\varepsilon}_i\}$ are captured by the term $\mathbf{Z}$. That is, by construction, $\mathbf{Z} \equiv \mathbf{0}$ and the square-root filter is devoid of the deteriorating sample effects of $\{\boldsymbol{\varepsilon}_i\}$. Thus, although the enKf asymptotically reproduces the Kalman recursions, the square-root filter update is always more efficient for finite n. The ensemble adjustment Kf of Anderson [2] provides similar advantages.

4. Minimizing sample variability using tapered covariances

The previous section shows that we require $n \propto q^2$ for bounded error growth in large-scale systems. We now address the situation $n \ll q$, common in operational NWP settings where the sample size is typically several orders of magnitude smaller than the system dimension: e.g. $n = 17$ at the Canadian Meteorological Centre (CMC) [45], and $n = 51$ at European Centre for Medium-Range Weather Forecasts (ECMWF) [46]. An effective approach to mitigate the spurious sample variability resulting from $n \ll q$, as well as to introduce sparseness, consists of multiplying $\hat{\mathbf{P}}^f$ element-wise by a compactly supported, positive definite matrix $\mathbf{C} = (c_{ij})$. In the context of NWP, this method is proposed by Houtekamer and Mitchell [33] and Hamill et al. [27] and produces the so-called 'tapered' forecast matrix $\hat{\mathbf{P}}_{\mathbf{C}}^f = \hat{\mathbf{P}}^f \circ \mathbf{C}$. (Here, $\circ$ represents the Schur product).

In our work tapering is used as a tool for variance stabilization and reduction and produces a shrinkage type estimator of the forecast covariance matrix. The technique is also considered in the context of spatial prediction to introduce sparseness in the covariance structure [21]. Their approach is based on asymptotic mean squared prediction criteria; here, in light of the computational expense of NWP, we derive optimal taper functions (and matrices) in finite sample size settings.

As motivation, we illustrate the effects of using an optimal taper in the simple case of a diagonal forecast matrix. (This is of course not a realistic assumption, but provides the only instance in which we can analytically relate the error of the tapered covariance matrix to the eigenvalues of $\mathbf{P}^f$). By Lemma 8 (see Appendix), the MSE $\|\mathbf{P}^f - \hat{\mathbf{P}}^f \circ \mathbf{C}\|_{\mathrm{E}}^2$ is a complicated expression, depending in part on the elements of the rotation matrix $\boldsymbol{\Gamma}$. However, for a diagonal forecast matrix $\mathbf{P}^f$, i.e. with $\boldsymbol{\Gamma} = \mathbf{I}$, by application of the same techniques as in Section 3.2, we have

$$\|\mathbf{P}^f - \hat{\mathbf{P}}^f \circ \mathbf{C}\|_{\mathrm{E}}^2 = \sum_i (c_{ii} - 1)^2 \lambda_i^2 + \frac{1}{n} \sum_i c_{ii}^2 \lambda_i^2 + \frac{1}{n} \sum_{i,j} c_{ij}^2 \lambda_i \lambda_j. \tag{16}$$

For the optimal choice $\mathbf{C} = \mathbf{I}$ the above expression simplifies to $2 \sum_i \lambda_i^2 / n$, and for a flat spectrum the MSE is reduced by a factor q (compare to Table 1). However, even for the optimal taper matrix, (16) tends to zero only if $q/n \to 0$. Further, due to bias introduced by the Schur product, the MSE does not tend to zero as $n \to \infty$ unless $\mathbf{C}$ is a correlation matrix.

Based on the tapered forecast covariance $\hat{\mathbf{P}}^f \circ \mathbf{C}$, similar developments can be carried out for the MSE of the propagated analysis covariance $\hat{\mathbf{P}}_{\mathbf{C}}^a$, and a second order approximation for the lower bound is given by

$$\|\mathbf{P}^a - \hat{\mathbf{P}}_{\mathbf{C}}^a\|_{\mathrm{E}}^2 \approx \sum_i \frac{(c_{ii} - 1)^2 \lambda_i^2}{(\lambda_i + 1)^4} + \frac{1}{n} \sum_i \frac{c_{ii}^2 \lambda_i^2}{(\lambda_i + 1)^4} + \frac{1}{n} \sum_{i,j} \frac{c_{ij}^2 \lambda_i \lambda_j}{(\lambda_i + 1)^3 (\lambda_j + 1)}. \tag{17}$$
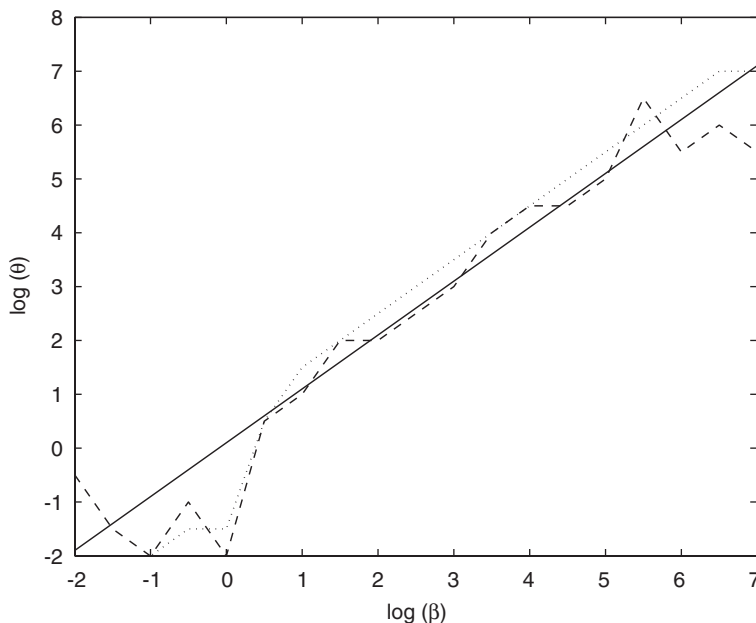
Fig. 2. Simulated optimal taper values θ as a function of β in (natural) log-scale. For $q = 40$ and $n = 10$, $p(h) = \exp(-3h/\beta)$ and $c(h) = \exp(-3h/\theta)$. The approximation given by (22) (solid) follows closely the optimal values of θ minimizing $\|\mathbf{P}^f - \hat{\mathbf{P}}^f_\mathbf{C}\|^2_E$ (dashed) and $\|\mathbf{P}^a - \hat{\mathbf{P}}^a_\mathbf{C}\|^2_E$ (dotted). (See Section 6 for the simulation setup).

As can be seen, in the special case of $\mathbf{\Gamma} = \mathbf{I}$, (16) and (17) are minimized for approximately the same taper, and for forecast covariance matrices that are dominated by small scale variability (i.e. with $\mathbf{P}^f$ "close" to a diagonal matrix), optimal tapering of $\hat{\mathbf{P}}^f$ should produce near optimal results for $\hat{\mathbf{P}}^a$. (This claim is supported by the simulation results shown in Fig. 2, provided in Section 4.2.)

Next we develop optimal taper matrices for general covariance structures.

4.1. Non-isotropic forecast covariance

Here we consider arbitrary covariance structures that cannot be described in terms of distance or other spatial features. We present a new taper result which is not dependent on distance.

Similarly to the idea of improved estimation of the mean it is possible to find an estimator $\hat{\mathbf{P}}^f \circ \mathbf{C}$ such that

$$\left\| \mathbf{P}^f - \hat{\mathbf{P}}^f \circ \mathbf{C} \right\|^2_E \leqslant \left\| \mathbf{P}^f - \hat{\mathbf{P}}^f \right\|^2_E.$$

We seek $\mathbf{C} = (c_{ij})$ to minimize

$$\left\| \mathbf{P}^f - \hat{\mathbf{P}}^f \circ \mathbf{C} \right\|^2_E = \operatorname{tr}((\mathbf{P}^f)^2) - 2\mathrm{E}\operatorname{tr}\left(\mathbf{P}^f(\hat{\mathbf{P}}^f \circ \mathbf{C}) \right) + \mathrm{E}\operatorname{tr}\left((\hat{\mathbf{P}}^f \circ \mathbf{C})^2 \right). \tag{18}$$

Using Lemma 7 one can show that minimizing (18) is equivalent to minimizing

$$\sum_{i,j} \left(p_{ij}^2 - 2c_{ij}\, p_{ij}^2 + c_{ij}^2 \left(p_{ij}^2 + \frac{1}{n}(p_{ij}^2 + p_{ii}\, p_{jj}) \right) \right). \tag{19}$$

To ensure that the estimator $\hat{\mathbf{P}}^f \circ \mathbf{C}$ is positive definite, the above expression must be minimized over the set of positive definite matrices $\mathbf{C}$ [30, Theorem 5.2.1]. Ignoring this constraint, which poses a non-trivial problem, we minimize (19) term by term to get

$$c_{ij} = \frac{p_{ij}^2}{p_{ij}^2 + (p_{ij}^2 + p_{ii}\, p_{jj})/n}. \tag{20}$$

For $p_{ij} = 0$ ($i \neq j$) we get $c_{ij} = 0$. Also, for $p_{ij} \neq 0$ ($i \neq j$), c_{ij} tends to one at rate $1/n$. It is important to note, however, that the optimal matrix $\mathbf{C}$ is not a correlation matrix (diagonal entries are $n/(n+2)$) and not necessarily positive definite. One can address these issues by several heuristic approaches. For instance, setting the diagonal entries of $\mathbf{C}$ equal to one and minimizing (19) over all $i \neq j$ results in a taper matrix with fewer negative eigenvalues. This approach can be generalized by boosting the optimal taper matrix to $\alpha\mathbf{I} + \beta\mathbf{C}$, where $\alpha, \beta > 0$ may be chosen to ensure positive definiteness. However, since this approach shifts the entire eigen-spectrum, α and β must be chosen with care. Another modification is to take the best full rank approximation of $\mathbf{C}$, i.e. take the positive eigenvalues of $\mathbf{C}$ and set the remaining ones equal to any small number $\varepsilon > 0$. Due to the computational cost of calculating eigenvalues, this latter approach is not interesting in high dimensions.

Although (20) may not provide a positive definite taper matrix it has appealing features. Primarily, it provides a non-distance-based taper scheme for general covariance structures, is computationally inexpensive, and, given sample data, a plug-in estimator based on $\hat{p}_{ij}$ is straightforwardly obtained. Also, sparseness may be introduced by setting $c_{ij} = 0$ whenever $\hat{p}_{ij} \approx 0$.

By imposing spatial structure on the forecast covariance, we next describe how to obtain optimal, positive definite taper functions.

4.2. Isotropic forecast covariance

We now consider covariance structures which depend only on distance. Such covariances can often be specified by a few parameters, e.g. the range describing the correlation length scale and the sill describing the magnitude of variation.

Suppose we observe a process on a linear transect with unit spacing and suppose further that the covariance between two locations i spacings apart is given by $p(i)$. It is straightforward to show that the expected difference (18) can be written as

$$\left\| \mathbf{P}^f - \hat{\mathbf{P}}^f \circ \mathbf{C} \right\|_{\mathrm{E}}^2 = \mathrm{E}\left(\sum_{i=1}^{q-1} 2(q-i)\left(p(i) - \hat{p}(i)c(i) \right)^2 + q\left(p(0) - \hat{p}(0)c(0) \right)^2 \right).$$

(Expressions for regular grids in higher dimensions are given in Furrer [20].) The above expression can be minimized over a class $\mathcal{C}$ of valid covariance functions $c(\cdot)$. For isotropic correlation functions, this minimization will not depend on $c(0)$, and motivates optimal taper functions by

$$\min_{c(\cdot) \in \mathcal{C}} \mathrm{E} \int_{\ell_q}^{u_q} \omega(h)\left(p(h) - \hat{p}(h)c(h) \right)^2 \, \mathrm{d}h.$$

Here, $\omega(\cdot)$ is a weight function and expectation is taken over the distribution of $\hat{p}(i)$. (The bounds ℓ_q and u_q of integration may depend on both q and $\omega(\cdot)$ to ensure a finite integral.) Under weak regularity conditions, using Lemma 7, the minimization reduces to

$$\min_{c(\cdot)\in\mathcal{C}} \int_{\ell_q}^{u_q} \omega(h) \left(p(h)^2 \left(1 - c(h)\right)^2 + \frac{1}{n} \left(p(h)^2 + p(0)^2 \right) c(h)^2 \right) \mathrm{d}h. \tag{21}$$

Eq. (21) has analytic solutions only in some special cases. For example, assume the exponential covariance function $p(h) = \alpha \exp(-3h/\beta)$, $\alpha, \beta > 0$, with parametrization such that $p(\beta) \approx 0.05\alpha$, i.e. β is the practical range [14]. Restrict the class of taper functions to $\mathcal{C} = \{c_\theta(h) = \exp(-3h/\theta), \theta > 0\}$. Then, for a constant weight function $\omega(h) \equiv \omega$, $\ell_q = 0$ and $u_q \to \infty$, the optimal θ in (21) is the root of a fourth-order polynomial with one real positive root, and for the weight function $\omega(h) \propto 1/h$, θ_{opt} can be approximated by

$$\theta_{\mathrm{opt}}(n) \approx \frac{\beta}{4} \left(\sqrt{9 + 8n} - 5 \right). \tag{22}$$

The optimal parameter is proportional to $\beta\sqrt{n}$ implying that, for fixed h, $c_{\theta_{\mathrm{opt}}}(h) \to 1$ as $n \to \infty$. Hence, the taper effect is $\mathcal{O}(n^{-1/2})$ and vanishes as n increases. Note that the isotropy assumption reduces the effective dimension of $\mathbf{P}^f$ (and the sample variability of $\hat{\mathbf{P}}^f$), resulting in a less pronounced taper effect.

For finite q and n, the validity of the approximation given by θ_{opt} is illustrated by Fig. 2. For $q = 40$ and $n = 10$ the figure depicts simulated optimal taper values based on the exponential covariance- and taper functions. The solid line is based on (22) and for these choices of q and n we get $\log(\theta_{\mathrm{opt}}) \approx \log(\beta)$. With $\hat{\mathbf{P}}_\mathbf{C}^a$ denoting the error covariance obtained by propagating $\hat{\mathbf{P}}_\mathbf{C}^f = \hat{\mathbf{P}}^f \circ \mathbf{C}$, Fig. 2 also shows that (22) accurately minimizes $\left\| \mathbf{P}^a - \hat{\mathbf{P}}_\mathbf{C}^a \right\|_\mathrm{E}^2$. Overall, for the exponential case, we see that the optimal taper length is closely related to the practical range.

In practice, Eq. (22) may serve as a rule of thumb to determine optimal taper lengths. That is, with an available estimator $\hat{\beta}$ (e.g. [14]), use $\hat{\beta}$ as a plug-in estimator to construct $\hat{\theta}_{\mathrm{opt}}$.

We next define pseudo-optimal tapers using compactly supported covariance functions.

4.3. Pseudo-optimal taper

Because the exponential taper considered in the previous section does not introduce sparseness it is not computationally attractive. To address the lack of sparsity, we derive sub-optimal taper functions with compact support.

In the isotropic case with covariance function $p(h)$, the optimal taper (20) is

$$c(h) = \frac{1}{1 + \left(1 + p(0)^2/p(h)^2\right)/n}. \tag{23}$$

The above formulation does not yield a positive definite taper function for all $p(\cdot)$. As a practical tool, we therefore suggest the use of a pseudo-optimal, compactly supported covariance function. We define such a taper function by finding a positive definite taper $c_{\mathrm{p\text{-}o}}(\cdot)$ that minimizes

$$\int_h \tilde{\omega}(h) \left(c_{\mathrm{p\text{-}o}}(h) - c(h) \right)^2 \mathrm{d}h, \tag{24}$$

where $\tilde{\omega}(\cdot)$ is a weight function to be specified and $c(\cdot)$ is given by (23).
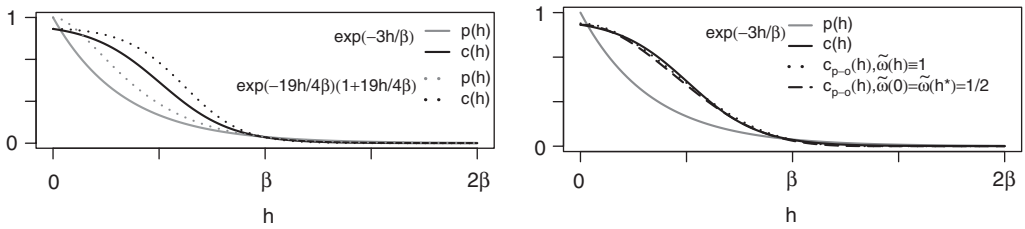
Fig. 3. Left panel: Covariance functions and optimal tapers based on (20). The exponential covariances $p(h) = \exp(-3h/\beta)$ (solid) and $p(h) = \exp(-19/4 \cdot h/\beta)(1 + 19/4 \cdot h/\beta)$ (dotted) are depicted in gray (lower curves). For $n = 20$, the optimal tapers $c(h)$ are depicted in black (top curves). Right panel: For $n = 20$, the optimal taper $c(h)$ (solid black) for the exponential covariance $p(h) = \exp(-3h/\beta)$ (solid gray), along with compactly supported covariance functions given by Gaspari and Cohn [22]. The pseudo-optimal taper functions are fitted using (24) by matching the sill and effective range of $c(h)$ (dashed) and by using a constant weight function $\tilde{\omega}(h)$ (dotted).

In NWP, a commonly used compactly supported taper is the fifth-order piecewise rational function obtained by self-convolving a triangular function over $\mathbb{R}^3$ (see [22, Eq. (4.10)]). Other compactly supported covariance functions can be found in Gneiting [23–25], Wendland [53,54], and Wu [56]. These functions are parameterized by their support length and sill (i.e. the value at zero). Therefore, a simple approach to obtaining a pseudo-optimal taper is to determine the two parameters of $c_{p\text{-}o}(\cdot)$ based on the sill and the practical range of $c(\cdot)$. (The practical range is defined as the value h^* such that $c(h^*) = 0.05c(0)$.)

The left panel of Fig. 3 shows two widely used exponential covariance functions (see caption) with their respective optimal taper functions defined by (23) and $n = 20$. This panel suggests that covariance functions with linear behavior at the origin are more likely to produce tapers $c(\cdot)$ (from (23)) which are positive definite for small n. For the exponential covariance $p(h) = \exp(-3h/\beta)$, the right panel of Fig. 3 shows pseudo-optimal tapers based on the compactly supported, positive definite function of Gaspari and Cohn [22]. To obtain $c_{p\text{-}o}(h)$ we use a discretized approximation of (24) where $\tilde{\omega}(0) = \tilde{\omega}(h^*) = 1/2$ and $\tilde{\omega}(h) = 0$ otherwise, as well as a numerical optimization of (24) with a constant weight function. As can be seen, since there is little difference between the pseudo-optimal tapers and $c(\cdot)$, this approach to defining compactly supported tapers is very effective.

5. Covariance boosting in large geophysical applications

As indicated in Corollary 3, the Kalman gain is negatively biased. In the univariate example of Section 2.3 the bias is effectively eliminated by artificial inflation (boosting) of the forecast covariance, a technique well-known in NWP (e.g. [4,55]). Typically, the inflation factor ρ is selected based on numerical experiments, but can also be based on the support length of the taper function [2]. Here, we extend the ideas of Section 2.3 and develop a theoretical framework for covariance inflation in the multivariate setting.

We assume that $\mathbf{H} = \mathbf{R} = \mathbf{I}$ and seek an inflation factor $\rho > 1$ such that

$$\mathrm{E}\,\mathrm{tr}(\hat{\mathbf{K}}_\rho - \mathbf{K}) = 0,$$

where $\hat{\mathbf{K}}_\rho$ is based on $\rho\hat{\mathbf{P}}^f$. Assuming that $\varrho\big(\rho(\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D}_\rho\big) < 1$ for $\mathbf{D}_\rho = (\rho\mathbf{\Lambda} + \mathbf{I})^{-1}$, developments similar to those of Section 3.3 and Lemma 9 can be used to express the bias in gain matrix by the

following expansion

$$\mathrm{E\,tr}\left((\mathbf{P}^f + \mathbf{I})^{-1} - (\rho\hat{\mathbf{P}}^f + \mathbf{I})^{-1}\right) = \sum_i \frac{1}{\lambda_i + 1} - \sum_{k=0}^{\infty} \mathrm{E\,tr}\left(\mathbf{D}_\rho\left((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D}_\rho\right)^k\right). \quad (25)$$

From (25) we obtain a second order approximation of the bias given by

$$B(\rho) = \sum_i \frac{\lambda_i(\rho - 1)}{(\rho\lambda_i + 1)(\lambda_i + 1)} - \frac{\rho^2}{n}\sum_i \frac{\lambda_i^2}{(\rho\lambda_i + 1)^3}$$
$$- \frac{\rho^2}{n}\sum_{i,j} \frac{\lambda_i\lambda_j}{(\rho\lambda_i + 1)^2(\rho\lambda_j + 1)}. \quad (26)$$

We can numerically determine a boosting factor ρ satisfying $B(\rho) = 0$ for arbitrary spectra $\{\lambda_i\}$. For all flat spectra $\lambda_i \equiv \lambda$, the bias reduces to

$$B(\rho) = \frac{q\lambda(\rho - 1)}{(\rho\lambda + 1)(\lambda + 1)} - \frac{\rho^2\lambda^2(q + q^2)}{n(\rho\lambda + 1)^3},$$

and for $B(\rho) = 0$ the boosting factor ρ is the solution of a third order polynomial similar to that obtained in the univariate case. The bias $B(1)$ is $\mathcal{O}(q^2/n)$, implying that bounded error is only guaranteed when $n \propto q^2$, a result consistent with those presented in Table 1.

Given the computational cost of determining representative estimates of the eigenvalues in large state spaces, Eq. (26) is not practical. Instead, similarly to Section 3, by imposing assumptions on the decay rate v of λ_i, the expression can be used to provide insight about the relationship between v and ρ. Moreover, as illustrated by the simulations of Section 6.3, applying an appropriate taper to the sample covariance matrix greatly reduces the need for boosting.

Using Monte Carlo simulations, the next section illustrates the theoretical results of Sections 3–5.

6. Simulations

By varying the decay of the eigen-spectrum of $\mathbf{P}^f$ and the form of the taper function we simulate the distributions of the squared errors $\|\mathbf{P}^f - \hat{\mathbf{P}}_\mathbf{C}^f\|^2$ and $\|\mathbf{P}^f - \hat{\mathbf{P}}_\mathbf{C}^a\|^2$. From these numerical evaluations it will (again) be clear that tapering yields an extremely efficient tool to reducing MSEs. Moreover, we show that the exact form of the taper function plays only a secondary role in improving accuracy. As a supplement to tapering we illustrate the effects of covariance boosting.

6.1. Setup

We assume an isotropic processes defined on a transect at locations $1, \ldots, q$, where, unless explicitly stated, $q = 1000$. The ensemble members are taken as Gaussian $x_i^f \sim \mathcal{N}_q(\mathbf{0}, \mathbf{P}^f)$, where $(\mathbf{P}^f)_{\ell k} = \exp\left(-3|\ell - k|/\beta\right)$. Thus, with $h = |\ell - k|$, we have $p(h) = \exp(-3h/\beta)$. As in the theoretical developments, we set $\mathbf{H} = \mathbf{R} = \mathbf{I}$ and $\mathrm{tr}(\mathbf{P}^f) = q$. The simulations are performed by varying the ensemble size n, the range parameter β, and the choice of taper function. The chosen range parameters $\beta = 10, 100$, and 333, correspond to a spectrum explaining 80% of the total variation (in $\mathbf{P}^f$) with 300, 100, and 10 modes, respectively. Fig. 4 details the decay of the
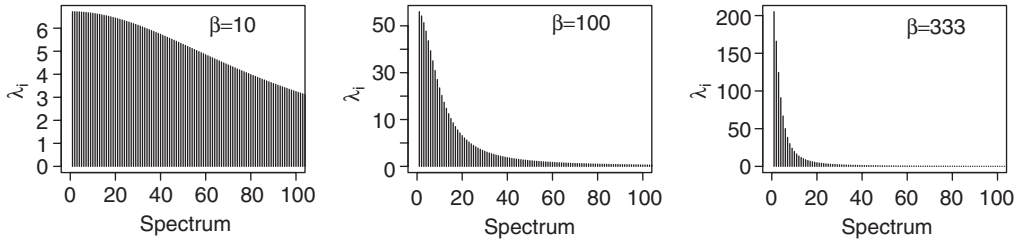
Fig. 4. The eigen-spectrum of the forecast covariance matrix used in the simulations. Here, $\mathbf{P}^f$ is based on the exponential covariance function $p(h) = \exp(-3h/\beta)$. The three plots show the first 100 modes of $\mathbf{P}^f$ for $\beta = 10$ (left), $\beta = 100$ (middle), and $\beta = 333$ (right).

eigen-spectrum as a function of β. For tapers we use the best full rank approximation of (20), the exponential taper with θ_{opt} given by (22), and the fifth-order piecewise rational function given by Gaspari and Cohn [22] with sill and effective range matched according to (24). The tapers are labeled 'FR', 'Exp' and 'GC', respectively.

6.2. MSE of sample forecast and analysis matrices

6.2.1. The forecast matrix

Fig. 5 depicts the sample distribution of $\|\mathbf{P}^f - \hat{\mathbf{P}}_{\mathbf{C}}^f\|^2$ for nine different combinations of β and n: the rows correspond to $\beta = 10, 100, 333$, and the columns correspond to $n = 10, 40, 160$. Each of the nine graphs show side-by-side box-plots comparing the effects of the three different tapers FR, Exp and GC.

We note that as β increases, implying fewer dominating modes, the error in $\hat{\mathbf{P}}_{\mathbf{C}}^f$ increases. The depicted results should be compared to the MSEs of the un-tapered forecast covariance matrices, which are, for the three choices of β used here, equal to $(9983 + q^2)/n$, $(95003 + q^2)/n$, and $(277916 + q^2)/n$ (see also Table 1). Thus, in this setting, the un-tapered error is at least $10^6/n$ and for $\beta = 10$ the error reduction is almost two orders of magnitude, emphasizing the extraordinary effect of the tapers. For rapidly decaying spectra ($\beta = 333$) the effect of tapering is approximately a 50% error reduction.

A crucial implication of Fig. 5 is that the choice of taper function is relatively unimportant for minimizing errors in $\hat{\mathbf{P}}^f$. Moreover, as emphasized in Fig. 6, which shows results evaluating the effects of varying the taper length of a GC-taper on $\|\mathbf{P}^f - \hat{\mathbf{P}}_{\mathbf{C}}^f\|^2$ with $n = 40$, the reduction in MSE is robust to the choice of taper length.

6.2.2. The analysis matrix

Simulated values of $\|\mathbf{P}^a - \hat{\mathbf{P}}_{\mathbf{C}}^a\|^2$ are given by the box-plots in Fig. 7. The results are based on the propagated analysis error, $\hat{\mathbf{P}}_{\mathbf{C}}^a = (\mathbf{I} - \hat{\mathbf{K}}_{\mathbf{C}})\hat{\mathbf{P}}_{\mathbf{C}}^f$, and pertain to the square-root enKf described in Section 3.4. The left panel shows simulated squared error values for the un-tapered case, while the three plots in the right panel depict tapered errors as a function of β and taper function (note the difference in scale from the left and right panels). The results are based on an ensemble size of $n = 10$.

Compared to $\hat{\mathbf{P}}_{\mathbf{C}}^f$, the effect of the decay rate of the spectrum on the MSE of $\hat{\mathbf{P}}_{\mathbf{C}}^a$ is reversed; i.e. rapidly decaying spectra produce smaller errors. This effect is consistent with the results of Section 3.3 (see Fig. 1 and Table 1). Interestingly, for $\hat{\mathbf{P}}_{\mathbf{C}}^a$, the GC-taper function performs worst,
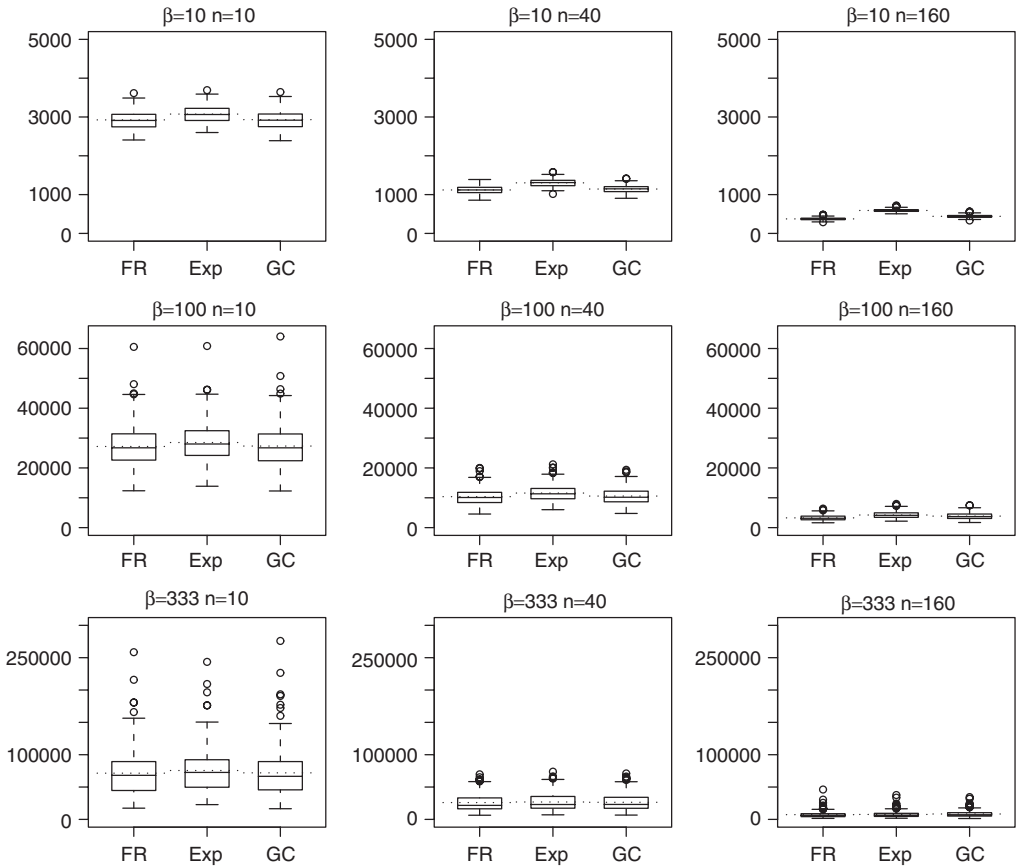
Fig. 5. Box-plots of simulated values of $\|\mathbf{P}^f - \hat{\mathbf{P}}_{\mathbf{C}}^f\|^2$. The rows correspond to range parameters $\beta = 10, 100,$ and 333, and the columns to ensemble sizes $n = 10, 40$ and 160. For each of the nine plots, the relative effectiveness of the taper variants (FR, Exp, GC) may be assessed through comparison of side-by-side box-plots. The means are dotted.
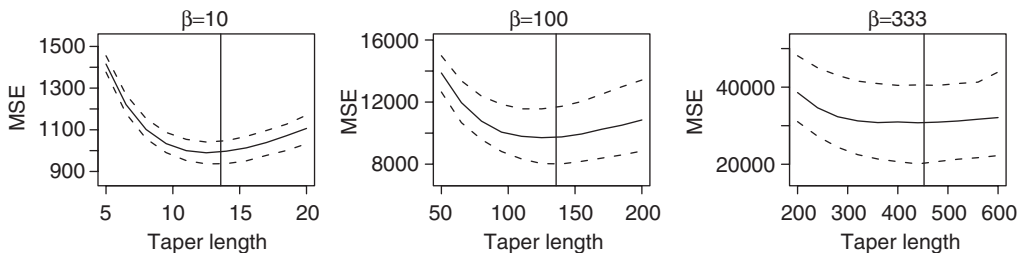

Fig. 6. $\|\mathbf{P}^f - \hat{\mathbf{P}}_{\mathbf{C}}^f\|^2$ as a function of the taper length of a GC-taper for $n = 40$. The vertical lines give the optimal taper lengths according (24) with $\tilde{\omega}(0) = \tilde{\omega}(h^*) = 1/2$ and $\tilde{\omega}(h) = 0$ otherwise. The dotted lines are point-wise 10% and 90% empirical confidence intervals.

but the choice of taper function is again a small factor in the MSE and this was found to be especially true for small n. Comparing the left and right panels of Fig. 7 we see a reduction in MSE of one order of magnitude for $\beta = 10$ and about a 50% reduction for $\beta = 333$. Thus, we
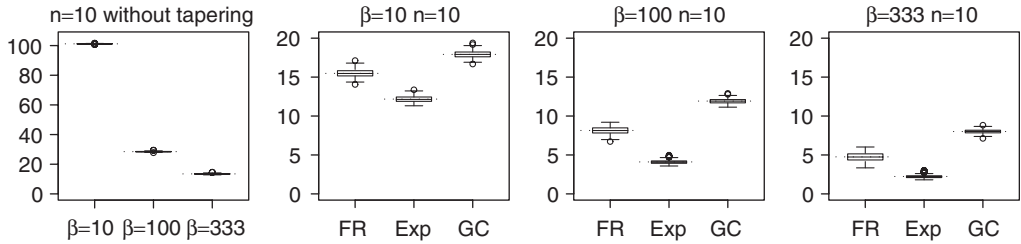
Fig. 7. Box-plots of simulated values of $\|\mathbf{P}^a - \hat{\mathbf{P}}^a_{\mathbf{C}}\|^2$. Left panel: un-tapered squared errors $\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|^2$. Right panel: tapered squared errors $\|\mathbf{P}^a - \hat{\mathbf{P}}^a_{\mathbf{C}}\|^2$ as a function of β and the choice of taper function. Each plot is based on 100 simulations and $n = 10$. The means are dotted. Note the difference in scale in the left and right panels.
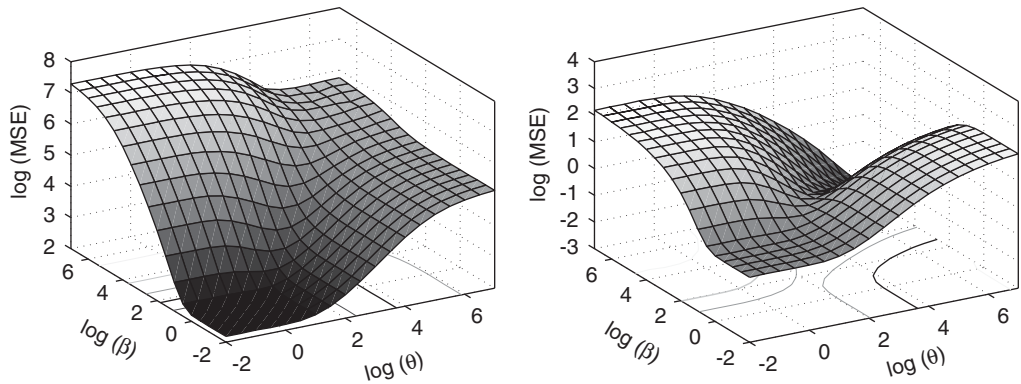

Fig. 8. $\log(\|\mathbf{P}^f - \hat{\mathbf{P}}^f_{\mathbf{C}}\|^2_{\mathrm{E}})$ (left panel) and $\log(\|\mathbf{P}^a - \hat{\mathbf{P}}^a_{\mathbf{C}}\|^2_{\mathrm{E}})$ (right panel) as a function of the covariance parameter β and the tapering parameter θ (note the log-scales). Here, $q = 40$ and $n = 10$. The contours along large values of θ give the MSEs associated with un-tapered $\hat{\mathbf{P}}^f$ (see Table 1). For both $\hat{\mathbf{P}}^f_{\mathbf{C}}$ and $\hat{\mathbf{P}}^a_{\mathbf{C}}$, the optimal taper values track along the contour $\log(\theta_{\mathrm{opt}}) \approx \log(\beta)$ (see also Fig. 2).

note once more the importance of using a taper to attenuate the effects of sampling variability in $\hat{\mathbf{P}}^f$.

To provide a more detailed picture of the effects of range parameter β and taper length θ on the MSEs, Fig. 8 shows $\|\mathbf{P}^f - \hat{\mathbf{P}}^f_{\mathbf{C}}\|^2_{\mathrm{E}}$ (left panel) and $\|\mathbf{P}^a - \hat{\mathbf{P}}^a_{\mathbf{C}}\|^2_{\mathrm{E}}$ (right panel) for β and θ in (e^{-2}, e^7) (note the log-scales). Here, $\hat{\mathbf{P}}^f$ is tapered using the exponential (covariance) function $c(h) = \exp(-3\theta/h)$ and $q = 40$ and $n = 10$. Because the errors range over several orders of magnitude the MSEs are expressed in log-scale. Contours along large values of θ give the MSEs for un-tapered $\hat{\mathbf{P}}^f$ (see Table 1). The variance minimizing effect of the taper can be seen by following contours along small values of β. However, the undesirable bias-inducing effect of an incorrect taper can be seen for small θ and large β. The error surface is relatively flat when both β and θ are small (or large), but note the stronger dependence on selecting optimal taper values for β values in or near the saddle point. Overall, as given by (22), the optimal taper values track along the contour $\log(\theta_{\mathrm{opt}}) \approx \log(\beta)$ (see also Fig. 2).
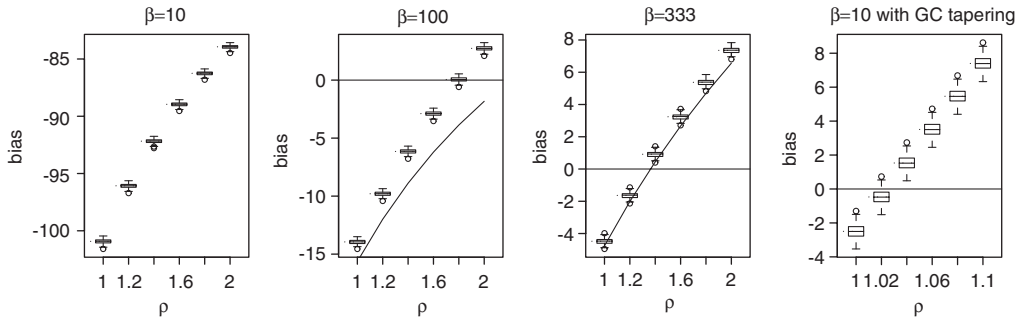
Fig. 9. Boxplots of $\text{tr}(\hat{\mathbf{K}}_\rho - \mathbf{K})$. Left panel: The bias for different inflation factors ρ and parameters β. Here, $\hat{\mathbf{K}}_\rho$ is based on the tapered and boosted forecast covariance $\rho\hat{\mathbf{P}}^f$ and $n = 10$. The superimposed solid line represents the second order approximation given by (26). Right panel: For $\beta = 10$, $\hat{\mathbf{K}}_\rho$ is based on the tapered and boosted forecast covariance $\rho\hat{\mathbf{P}}^f_{\mathbf{C}}$. The means are dotted.

6.3. Covariance boosting

To illustrate the results of Section 5, simulated solutions to $\text{E}\,\text{tr}\left((\mathbf{P}^f + \mathbf{I})^{-1} - (\rho\hat{\mathbf{P}}^f + \mathbf{I})^{-1}\right) = 0$ (see (25)) are compared to numerical solutions to the second order expansion (26). With $\hat{\mathbf{K}}_\rho$ defined using the boosted forecast covariance $\rho\hat{\mathbf{P}}^f$ and with $n = 10$, the three plots of the left panel in Fig. 9 displays the difference $\text{E}\,\text{tr}(\hat{\mathbf{K}}_\rho - \mathbf{K})$ as a function of the covariance parameter β and the boosting factor ρ. For a rapidly decaying spectrum, i.e. $\beta = 333$, the optimal boosting factor $\rho \approx 1.4$ is well approximated by the solution to (26). However, for $\beta = 10$ and $\beta = 100$ the solution to (26) overestimates the necessary boosting. For comparison, the right panel provides results for the case when $\hat{\mathbf{P}}^f$ is tapered using the GC-taper, i.e. with $\hat{\mathbf{K}}_\rho$ evaluated using $\rho\hat{\mathbf{P}}^f_{\mathbf{C}}$. As can be seen, for $\beta = 10$, tapering reduces the bias by more than one order of magnitude and yields a simulated boosting factor of $\rho \approx 1.03$, a value consistent with those reported by Anderson [2] and Whitaker and Hamill [55].

7. Summary and discussion

This work analyzes the MSE properties of the sample forecast (prior) and analysis (posterior) covariance matrices produced by the enKf. In the case of the forecast matrix we obtain exact expressions for the MSE, emphasizing the need of ensemble sizes proportional to the square of the system dimension for bounded error growth. Matrix expansions are required to approximate the MSE of the analysis covariance matrix and generally yield similar sample size requirements. However, for rapidly decaying eigen-spectra relatively few ensemble members are sufficient to obtain bounded error for the analysis covariance. Covariance-shrinking (e.g. tapering) through the use of the Schur product is shown to significantly improve large scale covariance estimation and reduce the MSEs by several orders of magnitude. A non-distance-based taper is presented for general, non-isotropic covariances, and for the exponential forecast covariance we show the optimal taper length to be related to the practical range and to vanish as $\sqrt{n}$ grows. Insight is also provided on optimal boosting factors.

Although the main focus of this paper is of a theoretical nature, practical implementations in high-dimensional applications need further simplifications. For instance, to ease computational

burden in NWP, the following approximations are often made:

$$(\mathbf{C} \circ \mathbf{P}^f)\mathbf{H}' \approx \mathbf{C} \circ (\mathbf{P}^f \mathbf{H}') \quad \text{and} \quad \mathbf{H}(\mathbf{C} \circ \mathbf{P}^f)\mathbf{H}' \approx \mathbf{C} \circ (\mathbf{H}\mathbf{P}^f \mathbf{H}')$$

Ref. [33], where equality holds if and only if $\mathbf{H}$ is diagonal [30]. The condition $\mathbf{H} = h\mathbf{I}$ (some constant h) appears stringent, but note that with an appropriate projection (cf. Corollary 2), it is always possible to obtain a system satisfying this condition. Hence, we believe our results shed light on arbitrary (linear or linearized) observation operators.

Another practical aspect is posed by the need to estimate necessary taper functions from samples. Given a forecast ensemble, we believe this requirement can be met using "plug-in" schemes based on classical covariance or variogram estimators (e.g. [43,13,47]). In particular, by implication of the invariance principle (e.g. [9]), choosing an appropriately parameterized taper function can be thought of as a maximum likelihood problem. However, fitting covariance functions in high-dimensional systems is computationally prohibitive. For a discussion of covariance estimation techniques in the context of NWP, (see, e.g. [28,41,15,16]). Fortunately, the simulation results of Section 6 illustrate that the improvement in MSE due to tapering is robust to the choice of taper function and of taper length.

The presented simulations illustrate the theoretical results for moderate to large dimensional state-spaces. We envision to proceed by applying our results in very large-scale problems. The Data Assimilation Research Testbed (DART) [1] furnishes an excellent computational environment supporting data assimilation research and evaluation for arbitrarily sized problems.

Acknowledgments

Appendix

In the text, the propositions and corollaries are numbered consecutively. However, in this Appendix, to provide technical details, we first present the Lemmas, and follow this with proofs of the propositions and corollaries that appear in the text. We start with a brief summary of the results.

Lemma 7 uses the moment properties of the normal distribution to evaluate expectations necessary to explicitly write $\|\mathbf{P}^f - \hat{\mathbf{P}}^f\|_{\mathrm{E}}^2$ in terms of the eigenvalues of $\mathbf{P}^f$, and Lemma 8 extends these results for the tapered sample covariance $\hat{\mathbf{P}}^f \circ \mathbf{C}$. Lemma 9 then develops a series expansion for matrix inverses that appear in the Kalman gain, and is used in the text to approximate $\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|_{\mathrm{E}}^2$ (again as a function of the eigenvalues of $\mathbf{P}^f$; cf. Proposition 6). Matrix algebra and independence of $\{\mathbf{x}_{t,i}^f\}$ and $\{\boldsymbol{\varepsilon}_{t,i}\}$ are used to prove the two parts of Proposition 1, verifying the decomposition and expectation of $\hat{\mathbf{P}}^a$. Orthogonal matrix rotations, in conjunction with Jensen's inequality, are used to prove Corollaries 2, 3, and 5, as well as Proposition 4, all results describing the effects of sample variability on enKf and $\hat{\mathbf{P}}^a$. Verification of the expansion of $\|\mathbf{P}^a - \hat{\mathbf{P}}^a\|_{\mathrm{E}}^2$ is given by the proof of Proposition 6.

---

[1] URL: ⟨www.image.ucar.edu/DAReS/DART/⟩.

Lemma 7. *Let $x_1^f, \ldots, x_n^f$ be iid Gaussian random q-vectors with mean $\mathbf{0}$ and variance $\mathbf{P}^f = (p_{ij})$. Then, using the notation introduced in Section 3.1, we have*

$$\mathrm{E}(\hat{\lambda}_{ij}) = \begin{cases} 0 & \text{if } i \neq j, \\ \lambda_i & \text{if } i = j, \end{cases}$$

$$\mathrm{E}(\hat{\lambda}_{ij}^2) = \begin{cases} \dfrac{1}{n}\lambda_i\lambda_j & \text{if } i \neq j, \\ \lambda_i^2 + \dfrac{2}{n}\lambda_i^2 & \text{if } i = j. \end{cases}$$

$$\mathrm{E}(\hat{\lambda}_{ij}\hat{\lambda}_{j\ell}\hat{\lambda}_{\ell i}) = \begin{cases} \dfrac{1}{n^2}\lambda_i\lambda_j\lambda_\ell & \text{if } i \neq j, i \neq \ell, \\ \dfrac{1}{n}\lambda_i\lambda_j^2 + \dfrac{2}{n^2}\lambda_i\lambda_j^2 & \text{if } i \neq j = \ell, \\ \dfrac{1}{n}\lambda_i^2\lambda_\ell + \dfrac{2}{n^2}\lambda_i^2\lambda_\ell & \text{if } i = j \neq \ell, \\ \dfrac{1}{n}\lambda_i^2\lambda_j + \dfrac{2}{n^2}\lambda_i^2\lambda_j & \text{if } i = \ell \neq j, \\ \lambda_i^3 + \dfrac{6}{n}\lambda_i^3 + \dfrac{8}{n^2}\lambda_i^3 & \text{if } i = j = \ell. \end{cases} \tag{27}$$

Proof. The first term is

$$\mathrm{E}(\hat{\lambda}_{ij}) = \mathrm{E}((\mathbf{\Gamma}'\hat{\mathbf{P}}^f\mathbf{\Gamma})_{ij}) = (\mathbf{\Gamma}')_i\mathbf{P}^f\mathbf{\Gamma}_j = (\mathbf{\Gamma}')_i\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}'\mathbf{\Gamma}_j = \delta_{ij}\lambda_i,$$

where δ_{ij} is the Kronecker-delta. To evaluate the second term note that

$$\hat{\lambda}_{ij}^2 = (\mathbf{\Gamma}'\hat{\mathbf{P}}^f\mathbf{\Gamma})_{ij}^2 = \gamma_i'\hat{\mathbf{P}}^f\gamma_j\gamma_i'\hat{\mathbf{P}}^f\gamma_j = \sum_{r,s,k,l}\gamma_{ri}\gamma_{sj}\gamma_{ki}\gamma_{lj}\hat{p}_{rs}\hat{p}_{kl}, \tag{28}$$

where $\gamma_i = (\gamma_{1j}, \ldots, \gamma_{qj})'$ is the *i*th column of the matrix $\mathbf{\Gamma}$. With $x_i^f = (u_{1i}, u_{2i}, \ldots, u_{qi})'$, we need

$$\mathrm{E}(\hat{p}_{rs}\hat{p}_{kl}) = \frac{1}{n^2}\mathrm{E}\left(\sum_v^n u_{vr}u_{vs}\sum_w^n u_{wk}u_{wl}\right).$$

Given the assumption of multivariate normality, we have

$$\mathrm{E}(u_{vr}u_{vs}u_{wk}u_{wl}) = \begin{cases} p_{rs}p_{kl} & \text{if } v \neq w, \\ p_{rs}p_{kl} + p_{rk}p_{sl} + p_{rl}p_{sk} & \text{if } w = v, \end{cases} \tag{29}$$

and

$$\mathrm{E}\left(\hat{p}_{rs}\hat{p}_{kl}\right) = p_{rs}p_{kl} + \frac{1}{n}\left(p_{rk}p_{sl} + p_{rl}p_{sk}\right).$$

Substituting the above expression in Eq. (28) yields

$$\mathrm{E}(\hat{\lambda}_{ij}^2) = \sum_{r,s,k,l}\gamma_{ri}p_{rs}\gamma_{sj}\cdot\gamma_{ki}p_{kl}\gamma_{lj} + \frac{1}{n}\sum_{r,s,k,l}\gamma_{ri}p_{rk}\gamma_{ki}\cdot\gamma_{sj}p_{sl}\gamma_{lj}$$

$$+\frac{1}{n}\sum_{r,s,k,l}\gamma_{ri}\,p_{rl}\gamma_{lj}\cdot\gamma_{sj}\,p_{sk}\gamma_{ki}$$

$$=\begin{cases}\dfrac{1}{n}\lambda_i\lambda_j & \text{if } i\neq j,\\[2mm]\lambda_i^2+\dfrac{2}{n}\lambda_i^2 & \text{if } i=j,\end{cases}$$

where we used the fact that $\sum_{r,s}\gamma_{ri}\,p_{rs}\gamma_{sj}=\delta_{ij}\lambda_i$.

To prove the third moment, consider the facts

$$\hat{\lambda}_{ij}\hat{\lambda}_{j\ell}\hat{\lambda}_{\ell i}=\sum_{r,s}\sum_{k,l}\sum_{u,v}\gamma_{ri}\,\hat{p}_{rs}\gamma_{sj}\cdot\gamma_{kj}\,\hat{p}_{kl}\gamma_{l\ell}\cdot\gamma_{u\ell}\,\hat{p}_{uv}\gamma_{vi},$$

and

$$\hat{p}_{rs}\hat{p}_{kl}\hat{p}_{uv}=\frac{1}{n^3}\mathrm{E}\left(\sum_\rho^n u_{\rho r}u_{\rho s}\sum_\zeta^n u_{\zeta k}u_{\zeta l}\sum_\xi^n u_{\xi u}u_{\xi v}\right). \tag{30}$$

The expectation of (30) can be evaluated for different cases of the summation indices. When the indices differ, using (29), we have

$$\mathrm{E}\left(u_{\rho r}u_{\rho s}u_{\zeta k}u_{\zeta l}u_{\xi u}u_{\xi v}\right)=\begin{cases}p_{rs}\,p_{kl}\,p_{uv} & \text{if } \rho\neq\zeta\neq\xi,\\p_{rs}\,p_{kl}\,p_{uv}+p_{rs}(p_{ku}\,p_{lv}+p_{kv}\,p_{ul}) & \text{if } \rho\neq\zeta=\xi,\\p_{kl}\,p_{rs}\,p_{uv}+p_{kl}(p_{ru}\,p_{sv}+p_{rv}\,p_{su}) & \text{if } \rho\neq\xi=\zeta,\\p_{uv}\,p_{rs}\,p_{kl}+p_{uv}(p_{rk}\,p_{sl}+p_{rl}\,p_{sk}) & \text{if } \rho=\zeta\neq\xi.\end{cases} \tag{31}$$

To find the expectation when all three indices are equal (i.e. $\rho=\zeta=\xi$), we evaluate the sixth moment of a multivariate random normal vector with corresponding covariance matrix via the characteristic function:

$$\mathrm{E}\left(u_{\rho r}u_{\rho s}u_{\zeta k}u_{\zeta l}u_{\xi u}u_{\xi v}\right)=p_{rs}\,p_{kl}\,p_{uv}+p_{rs}(p_{ku}\,p_{lv}+p_{kv}\,p_{ul}) \tag{32}$$

$$+p_{kl}(p_{ru}\,p_{sv}+p_{rv}\,p_{su})+p_{uv}(p_{rk}\,p_{sl}+p_{rl}\,p_{sk}) \tag{33}$$

$$+p_{rk}(p_{su}\,p_{lv}+p_{sv}\,p_{lu})+p_{rl}(p_{su}\,p_{kv}+p_{sv}\,p_{ku}) \tag{34}$$

$$+p_{ru}(p_{sk}\,p_{lv}+p_{sl}\,p_{kv})+p_{rv}(p_{sk}\,p_{lu}+p_{sl}\,p_{ku}). \tag{35}$$

We note that the last expression is composed of all the distinct terms found in (31) (first two lines) plus eight additional terms. The 15 terms represent the combination of arranging the indices $\{r,s,k,l,u,v\}$ in three urns. Eq. (30) is now a weighted sum of the terms given in (31) to (35). We note that the term $p_{rs}\,p_{kl}\,p_{uv}$ appears n^3 times, the terms in (33) $(n^2-n)+n$ times and terms in (34) and (35) n times. Hence,

$$\mathrm{E}\left(\hat{p}_{rs}\hat{p}_{kl}\hat{p}_{uv}\right)=p_{rs}\,p_{kl}\,p_{uv}+\frac{1}{n}\Big(p_{rs}(p_{ku}\,p_{lv}+p_{kv}\,p_{ul})+p_{kl}(p_{ru}\,p_{sv}+p_{rv}\,p_{su})$$

$$+p_{uv}(p_{rk}\,p_{sl}+p_{rl}\,p_{sk})\Big)$$

$$+\frac{1}{n^2}\Big(p_{rk}(p_{su}\,p_{lv}+p_{sv}\,p_{lu})+p_{rl}(p_{su}\,p_{kv}+p_{sv}\,p_{ku})$$

$$+p_{ru}(p_{sk}\,p_{lv}+p_{sl}\,p_{kv})+p_{rv}(p_{sk}\,p_{lu}+p_{sl}\,p_{ku})\Big),$$

The last step consists of calculating

$$\gamma_{ri}\gamma_{kj}\gamma_{u\ell}\mathrm{E}\left(\hat{p}_{rs}\hat{p}_{kl}\hat{p}_{uv}\right)\gamma_{sj}\gamma_{l\ell}\gamma_{vi}.$$

Note that for any distinct indices i and j, all the terms in $\gamma_{\alpha i}\mathrm{E}\left(\hat{p}_{rs}\hat{p}_{kl}\hat{p}_{uv}\right)\gamma_{\beta j}$ with $p_{\alpha\beta}$ will be zero, and the result in (27) can be straightforwardly obtained. This completes the proof.

Note that we use the moment properties of the Gaussian distribution in (29) and (31)–(35). In a similar way, Gaussianity allow us to calculate higher-order moments, but the number of terms for such expressions increase exponentially. $\square$

Lemma 8. *Given the same assumptions as in Lemma 7, but with calculations based on* $\hat{\mathbf{P}}^f \circ \mathbf{C}$ *(instead of* $\hat{\mathbf{P}}^f$*), we have*

$$\mathrm{E}(\hat{\lambda}_{ij}) = \gamma_i'(\mathbf{P}^f \circ \mathbf{C})\gamma_j,$$

$$\mathrm{E}(\hat{\lambda}_{ij}^2) = \left(\gamma_i'(\mathbf{P}^f \circ \mathbf{C})\gamma_j\right)^2 + \frac{1}{n}\sum_{r,s,k,l}\gamma_{ri}\gamma_{ki}\gamma_{sj}\gamma_{lj}p_{rk}p_{sl}c_{rs}c_{kl}$$

$$+\frac{1}{n}\sum_{r,s,k,l}\gamma_{ri}\gamma_{lj}\gamma_{sj}\gamma_{ki}p_{rl}p_{sk}c_{rs}c_{kl}.$$

Proof. Note that

$$\sum_{r,s}\gamma_{ri}p_{rs}c_{rs}\gamma_{sj} = \gamma_i'(\mathbf{P}^f \circ \mathbf{C})\gamma_j.$$

Calculations similar to those in the proof of Lemma 7 yield the result. $\square$

Lemma 9. *Let* Λ_1 *and* Λ_2 *be two matrices in* $\mathbb{R}^{q \times q}$ *and let* $\mathbf{D} = (\Lambda_1 + \mathbf{I})^{-1}$. *Suppose* $\rho\left((\Lambda_2 - \Lambda_1)\mathbf{D}\right) < 1$, *then*

$$(\Lambda_1 + \mathbf{I})^{-2} - 2(\Lambda_1 + \mathbf{I})^{-1}(\Lambda_2 + \mathbf{I})^{-1} + (\Lambda_2 + \mathbf{I})^{-2}$$

$$= \sum_{k=2}^{\infty}(k-1)\mathbf{D}^2\left((\Lambda_1 - \Lambda_2)\mathbf{D}\right)^k. \tag{36}$$

Proof. Corollary 5.6.16 of Horn and Johnson [29] implies that $\mathbf{I} + (\Lambda_2 - \Lambda_1)\mathbf{D}$ is invertible and

$$(\mathbf{I} + (\Lambda_2 - \Lambda_1)\mathbf{D})^{-1} = \sum_{k=0}^{\infty}\left((\Lambda_1 - \Lambda_2)\mathbf{D}\right)^k,$$

$$(\mathbf{I} + (\Lambda_2 - \Lambda_1)\mathbf{D})^{-2} = \sum_{k=0}^{\infty}(k+1)\left((\Lambda_1 - \Lambda_2)\mathbf{D}\right)^k.$$

Given the two series above, the right-hand side of (36) can be written as

$$\mathbf{D}^2 - 2\mathbf{D}^2\left(\mathbf{I} + (\Lambda_2 - \Lambda_1)\mathbf{D}\right)^{-1} + \mathbf{D}^2\left(\mathbf{I} + (\Lambda_2 - \Lambda_1)\mathbf{D}\right)^{-2}. \tag{37}$$

Further note that

$$\mathbf{D}^2 \, (\mathbf{I} + (\mathbf{\Lambda}_2 - \mathbf{\Lambda}_1)\mathbf{D})^{-1} = \mathbf{D}(\mathbf{D}^{-1} + \mathbf{\Lambda}_2 - \mathbf{\Lambda}_1)^{-1} = (\mathbf{\Lambda}_1 + \mathbf{I})^{-1}(\mathbf{\Lambda}_2 + \mathbf{I})^{-1}.$$

The same technique is applied to the last term of (37). The proof is concluded by collecting terms. $\square$

Proof of Proposition 1. *Part* (a): To simplify the notation, we omit the time subscript t. We use the definition of the updated sample in Eq. (8) and expand $\hat{\mathbf{P}}^a$:

$$\hat{\mathbf{P}}^a = \frac{1}{n-1} \sum_{i=1}^{n} \left((\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})(x_i^f - \bar{x}^f) + \hat{\mathbf{K}}(\varepsilon_i - \bar{\varepsilon}) \right) \left((\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})(x_i^f - \bar{x}^f) + \hat{\mathbf{K}}(\varepsilon_i - \bar{\varepsilon}) \right)',$$

where $\bar{x}^f$ and $\bar{\varepsilon}$ indicate the sample means. Using the definition of $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$, Eq. (9) follows.

Part (b): The matrix elements of (9) are linear combinations of the observation perturbations $\{\varepsilon_i\}$ and functions of $\{x_i^f\}$, over which the expectation of (10) is taken. As the latter two are independent samples, we can separate the expectation. Then, as $\mathrm{E}(\hat{\mathbf{C}}) = \mathbf{0}$ and $\mathrm{E}(\hat{\mathbf{R}}) = \mathbf{R}$, the result follows. $\square$

Proof of Corollary 2. We omit the time subscript t. First, notice that $\mathrm{tr}(\hat{\mathbf{P}}^a) = \mathrm{tr}(\mathbf{H}\hat{\mathbf{P}}^a\mathbf{H}')$. Further, one can show that

$$\mathbf{H}\mathbf{P}^a\mathbf{H}' = \sigma^2\mathbf{I} - \sigma^2\mathbf{I}(\mathbf{H}\mathbf{P}^f\mathbf{H}' + \sigma^2\mathbf{I})^{-1}\sigma^2\mathbf{I},$$

where the covariance recursion (3) was used along with well-known matrix identities. It suffices to show

$$\mathrm{E}\,\mathrm{tr}\left((\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H} + \sigma^2\mathbf{I})^{-1} \right) > \mathrm{tr}\left((\mathbf{H}\mathbf{P}^f\mathbf{H}' + \sigma^2\mathbf{I})^{-1} \right).$$

Let $\mathbf{G}\mathbf{L}\mathbf{G}' = \mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}'$ be the spectral decomposition of $\mathbf{H}\hat{\mathbf{P}}^f\mathbf{H}'$, that is, $\mathbf{L}$ is a diagonal matrix with entries $\mathbf{L}_{ii}$ and $\mathbf{G}$ is an orthogonal matrix. Then

$$\mathrm{E}\,\mathrm{tr}\left((\mathbf{G}\mathbf{L}\mathbf{G}' + \sigma^2\mathbf{I})^{-1} \right) = \sum_{i=1}^{q} \mathrm{E}\left(\frac{1}{\mathbf{L}_{ii} + \sigma^2} \right) > \sum_{i=1}^{q} \left(\frac{1}{\mathrm{E}(\mathbf{L}_{ii}) + \sigma^2} \right)$$

$$= \mathrm{tr}\left((\mathbf{G}\mathrm{E}(\mathbf{L})\mathbf{G}' + \sigma^2\mathbf{I})^{-1} \right) = \mathrm{tr}\left((\mathbf{H}\mathbf{P}^f\mathbf{H}' + \sigma^2\mathbf{I})^{-1} \right),$$

where we have used Jensen's inequality. The result of the corollary follows. $\square$

Proof of Corollary 3. We omit the time subscript t. The Kalman gain can be written as $\mathbf{K} = \mathbf{P}^f(\mathbf{P}^f + \sigma^2\mathbf{I})^{-1} = \sigma^2\mathbf{I} - \sigma^2(\mathbf{P}^f + \sigma^2\mathbf{I})^{-1}$. To show that $\mathrm{tr}\,\mathrm{E}(\hat{\mathbf{K}}) < \mathrm{tr}(\hat{\mathbf{K}})$, it is sufficient to show $\mathrm{tr}\left((\hat{\mathbf{P}}^f + \sigma^2\mathbf{I})^{-1} \right) > \mathrm{tr}\left((\mathbf{P}^f + \sigma^2\mathbf{I})^{-1} \right)$. We proceed as in the proof of Corollary 2. $\square$

Proof of Proposition 4. We use the same arguments as in the proof of Proposition 1. $\square$

Proof of Corollary 5. Proposition 4 assures the orthogonal decomposition. $\square$

Proof of Proposition 6. With $\mathbf{H} = \mathbf{R} = \mathbf{I}$ and $\mathbf{P}^a = \mathbf{I} - (\mathbf{P} + \mathbf{I})^{-1}$, we evaluate

$$\left\| (\mathbf{P}^f + \mathbf{I})^{-1} - (\hat{\mathbf{P}}^f + \mathbf{I})^{-1} \right\|_{\mathrm{E}}^2 = \mathrm{E}\,\mathrm{tr}\left((\mathbf{\Lambda} + \mathbf{I})^{-2} - 2(\mathbf{\Lambda} + \mathbf{I})^{-1}(\hat{\mathbf{\Lambda}} + \mathbf{I})^{-1} + (\hat{\mathbf{\Lambda}} + \mathbf{I})^{-2} \right).$$

Because $\varrho\left((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D}\right) < 1$, we use Lemma 9, and write the previous expression as

$$\sum_{k=2}^{\infty}(k-1)\operatorname{tr}\operatorname{E}\left(\left(\mathbf{D}^2\left(\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}\right)\mathbf{D}\right)^k\right).$$

Hence,

$$\operatorname{E}\operatorname{tr}\left(((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D})^2\mathbf{D}^2\right) - 2\operatorname{E}\operatorname{tr}\left(((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D})^3\mathbf{D}^2\right) \tag{38}$$

is a third order approximation of $\left\|\mathbf{P}^a - \hat{\mathbf{P}}^a\right\|_{\mathrm{E}}^2$. We will evaluate the two terms of (38) separately. The trace of the first term can be expressed as

$$\operatorname{tr}\left(((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D})^2\mathbf{D}^2\right) = \operatorname{tr}\left(\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}\mathbf{D}^3 - \mathbf{\Lambda}\mathbf{D}\hat{\mathbf{\Lambda}}\mathbf{D}^3 - \hat{\mathbf{\Lambda}}\mathbf{D}\mathbf{\Lambda}\mathbf{D}^3 + \hat{\mathbf{\Lambda}}\mathbf{D}\hat{\mathbf{\Lambda}}\mathbf{D}^3\right),$$

and as the trace is invariant with respect to circular permutation we have

$$\begin{aligned}
\operatorname{E}\operatorname{tr}\left(((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D})^2\mathbf{D}^2\right) &= \sum_{i,j}\operatorname{E}(\hat{\lambda}_{ij}d_j\hat{\lambda}_{ji}d_i^3) - \sum_{i,j}\operatorname{E}(\lambda_i d_i\hat{\lambda}_{ii}d_i^3) \\
&\quad - \sum_{i,j}\operatorname{E}(\hat{\lambda}_{ii}d_i\lambda_i d_i^3) + \sum_{i,j}d_i^4\lambda_i^2 \\
&= \frac{1}{n}\sum_i\frac{\lambda_i^2}{(\lambda_i+1)^4} + \frac{1}{n}\sum_{i,j}\frac{\lambda_i\lambda_j}{(\lambda_i+1)^3(\lambda_j+1)},
\end{aligned} \tag{39}$$

where Lemma 7 was used to evaluate expectations. To evaluate the second term in (38), we simplify the cube and collect identical terms yielding

$$\begin{aligned}
\operatorname{tr}\left(((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D})^3\mathbf{D}^2\right) &= \operatorname{tr}\left(\mathbf{\Lambda}^3\mathbf{D}^5\right) - 3\operatorname{tr}\left(\hat{\mathbf{\Lambda}}\mathbf{\Lambda}^2\mathbf{D}^5\right) + 2\operatorname{tr}\left(\hat{\mathbf{\Lambda}}\mathbf{D}\hat{\mathbf{\Lambda}}\mathbf{D}\mathbf{\Lambda}\mathbf{D}^3\right) \\
&\quad + \operatorname{tr}\left(\hat{\mathbf{\Lambda}}\mathbf{D}\mathbf{\Lambda}\mathbf{D}\hat{\mathbf{\Lambda}}\mathbf{D}^3\right) - \operatorname{tr}\left((\hat{\mathbf{\Lambda}}\mathbf{D})^3\mathbf{D}^2\right).
\end{aligned}$$

Analogously to the development of the first term in (38), the above can be reduced to

$$\begin{aligned}
\operatorname{E}\operatorname{tr}&\left(((\mathbf{\Lambda} - \hat{\mathbf{\Lambda}})\mathbf{D})^3\mathbf{D}^2\right) \\
&= -\sum_i\frac{\lambda_i^3}{(\lambda_i+1)^5} + 2\frac{n+1}{n^2}\left(\sum_i\frac{\lambda_i^3}{(\lambda_i+1)^5} + \sum_{i,j}\frac{\lambda_i^2\lambda_j}{(\lambda_i+1)^4(\lambda_j+1)}\right) \\
&\quad + \frac{n+1}{n^2}\left(\sum_i\frac{\lambda_i^3}{(\lambda_i+1)^5} + \sum_{i,j}\frac{\lambda_i\lambda_j^2}{(\lambda_i+1)^3(\lambda_j+1)^2}\right) \\
&\quad - \sum_{i,j,\ell}\frac{\operatorname{E}\left(\hat{\lambda}_{ij}\hat{\lambda}_{j\ell}\hat{\lambda}_{\ell i}\right)}{(\lambda_i+1)^3(\lambda_j+1)(\lambda_\ell+1)},
\end{aligned} \tag{40}$$

where, again, expectations are evaluated using Lemma 7. Evaluating the expectation of the last term in the previous equation all the terms not involving $1/n^2$ cancel. Combining (39) and (40) gives the third order approximation given in (14) and (15). $\quad\square$

References

[1] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions, Dover, New York, 1970.

[2] J.L. Anderson, An ensemble adjustment Kalman filter for data assimilation, Monthly Weather Rev. 129 (2001) 2884–2903.

[3] J.L. Anderson, Exploring the need for localization in ensemble data assimilation using an hierarchical ensemble filter, Phys. D (2006) in press.

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

[5] T.W. Anderson, Asymtotic theory for principal component analysis, Ann. Math. Statist. 34 (1963) 122–148.

[6] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, second ed., Wiley, New York, 1984.

[7] S. Arnold, Mathematical Statistics, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[8] T. Bengtsson, C. Snyder, D. Nychka, Toward a nonlinear ensemble filter for high-dimensional systems, J. Geophys. Res. 108 (2003) 8775.

[9] P. Bickel, K. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, vol. 1, second ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[10] C. Bishop, B. Etherton, S. Majumdar, Adaptive sampling with the ensemble transform Kalman Filter, Part i: theoretical aspects, Monthly Weather Rev. 129 (2001) 420–436.

[11] G. Burgers, P.J. van Leeuwen, G. Evensen, Analysis scheme in the ensemble Kalman filter, Monthly Weather Rev. 126 (1998) 1719–1724.

[12] P. Courtier, O. Talagrand, Variational assimilation of meteorological observations with the adjoint equation, Part I: numerical results, Quart. J. Roy. Meterol. Soc. 113 (1987) 1329–1347.

[13] N.A.C. Cressie, Fitting variogram models by weighted least squares, J. Internat. Assoc. Math. Geol. 17 (1985) 563–586.

[14] N.A.C. Cressie, Statistics for Spatial Data, Revised reprint, Wiley, New York, 1993.

[15] D. Dee, A. da Silva, Maximum likelihood estimation of forecast and observation error covariance parameters, Part I: methodology, Monthly Weather Rev. 127 (1999) 1822–1834.

[16] D. Dee, G. Gaspari, C. Redder, L. Rukhovets, A. da Silva, Maximum likelihood estimation of forecast and observation error covariance parameters, Part II: applications, Monthly Weather Rev. 127 (1999) 1835–1849.

[17] A. Doucet, N. Freitas, N. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Springer, Berlin, 2001.

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

[19] G. Evensen, P.J. van Leeuwen, Assimilation of geostat altimeter data for the Agulhas current using the ensemble Kalman filter, Monthly Weather Rev. 124 (1996) 85–96.

[20] R. Furrer, Covariance estimation under spatial dependence, J. Multivariate Anal. 94 (2005) 366–381.

[21] R. Furrer, M.G. Genton, D. Nychka, Covariance tapering for interpolation of large spatial datasets, J. Comput. Graphical Statist. 15 (2006) 502–523.

[22] G. Gaspari, S.E. Cohn, Construction of correlation functions in two and three dimensions, Quart. J. Roy. Meteorol. Soc. 125 (1999) 723–757.

[23] T. Gneiting, Correlation functions for atmospheric data analysis, Quart. J. Roy. Meteorol. Soc. 125 (1999) 2449–2464.

[24] T. Gneiting, Radial positive definite functions generated by Euclid's hat, J. Multivariate Anal. 69 (1999) 88–119.

[25] T. Gneiting, Compactly supported correlation functions, J. Multivariate Anal. 83 (2002) 493–508.

[26] T.M. Hamill, C. Snyder, R. Morss, A comparison of probabilistic forecasts from bred, singular vector and perturbed observation ensembles, Monthly Weather Rev. 128 (2000) 1835–1851.

[27] T.M. Hamill, J.S. Whitaker, C. Snyder, Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Monthly Weather Rev. 129 (2001) 2776–2790.

[28] A. Hollingsworth, P. Lonnberg, The statistical structure of short-range forecast errors as determined from radiosonde data, Part I: the wind field, Tellus 38A (1986) 111–136.

[29] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1990.

[30] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1994.

[31] P.L. Houtekamer, L. Lefaivre, J. Derome, H. Ritchie, H.L. Mitchell, A system simulation approach to ensemble prediction, Monthly Weather Rev. 124 (1996) 1225–1242.

[32] P.L. Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman filter technique, Monthly Weather Rev. 126 (1998) 796–811.

[33] P.L. Houtekamer, H.L. Mitchell, A sequential ensemble Kalman filter for atmospheric data assimilation, Monthly Weather Rev. 129 (2001) 123–137.

[34] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295–327.

[35] R.E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (1960) 34–45.

[36] E. Klinker, F. Rabier, G. Kelly, J. Mahfouf, The ECMWF operational implementation of four-dimensional variational assimilation, III: experimental results and diagnostics with operational configuration, Quart. J. Roy. Meterol. Soc. 126 (2000) 1191–1215.

[37] H.R. Künsch, State space and hidden markov models, in: O.E. Barndorff-Nielsen, D.R. Cox, C. Klüppelberg (Eds.), Complex Stochastic Systems, Monographs on Statistics and Applied Probability, vol. 87, Chapman & Hall, London, 2001, pp. 109–173, (Chapter 3).

[38] P. Leeuwen, Comment on "Data assimilation using an ensemble Kalman filter", Monthly Weather Rev. 127 (1999) 1374–1379.

[39] B. Li, T. Bengtsson, P. Bickel, Curse-of-dimensionality revisited: collapse of importance sampling in very large scale systems, Technical Report #696, Department of Statistics, UC-Berkeley, 2005.

[40] J.S. Liu, Monte Carlo Strategies in Scientific Computing, Springer Series in Statistics, Springer, New York, 2001.

[41] P. Lonnberg, A. Hollingsworth, The statistical structure of short-range forecast errors as determined from radiosonde data, Part II: the covaraince of height and wind errors, Tellus 38A (1986) 137–161.

[42] A.C. Lorenc, Analysis methods for numerical weather prediction, Quart. J. Roy. Meterol. Soc. 112 (1986) 1177–1194.

[43] K.V. Mardia, R.J. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71 (1984) 135–146.

[44] F. Molteni, R. Buizza, T.N. Palmer, T. Petroliagis, The ECMWF ensemble prediction system: methodology and validation, Quart. J. Roy. Meteorol. Soc. 122 (1996) 73–119.

[45] G. Pellerin, L. Lefaivre, P.L. Houtekamer, C. Girard, Increasing the horizontal resolution of ensemble forecasts at CMC, Nonlinear Process. Geophys. 10 (2003) 463–468.

[46] D. Richardson, Skill and relative economic value of the ECMWF ensemble prediction system, Quart. J. Roy. Meteorol. Soc. 126 (2000) 649–667.

[47] M.L. Stein, Minimum norm quadratic estimation of spatial variograms, J. Amer. Statist. Assoc. 82 (1987) 765–772.

[48] M. Tippet, S. Cohn, Adjoints and low-rank covariance representation, Nonlinear Process. Geophys. 8 (2001) 331–340.

[49] M.K. Tippett, J.L. Anderson, C.H. Bishop, T.M. Hamill, J.S. Whitaker, Ensemble square-root filters, Monthly Weather Rev. 131 (2003) 1485–1490.

[50] Z. Toth, E. Kalnay, Ensemble forecasting at NMC: the generation of perturbations, Bull. Amer. Meteorol. Soc. 74 (1993) 2317–2330.

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

[52] C.A. Tracy, H. Widom, The distribution of the largest eigenvalue in the Gaussian ensembles, in: J. van Diejen, L. Vinet (Eds.), Calogero—Moser—Sutherland Models, CRM Series in Mathematical Physics, vol. 4, Springer, New York, 2000, pp. 461–472.

[53] H. Wendland, Piecewise polynomial positive definite and compactly supported radial functions of minimal degree, Adv. Comput. Math. 4 (1995) 389–396.

[54] H. Wendland, Error estimates for interpolation by compactly supported radial basis functions of minimal degree, J. Approx. Theory 93 (1998) 258–272.

[55] J.S. Whitaker, T.M. Hamill, Ensemble data assimilation without perturbed observations, Monthly Weather Rev. 130 (2002) 1913–1924.

[56] Z.M. Wu, Compactly supported positive definite radial functions, Adv. Comput. Math. 4 (1995) 283–292.