Covariance modeling by means of eigenfunctions of Laplace operator

Marie Turčičová* Jan Mandel† Kryštof Eben‡

Abstract

Large dimension of both state vector and data is a well-known challenge in environmental modeling (e.g., numerical weather forecast) and, in particular, in data assimilation. The Ensemble Kalman Filter addresses this problem by estimating the current system state and its uncertainty via their sample counterparts. However, sample covariance matrix based on a small ensemble is not a sufficiently good estimate of the true covariance. In this paper, we deal with techniques relying on transformation of the state to the spectral space and assuming a particular covariance structure based on the Laplace operator. Parameters, which this special structure depends on, are estimated by a least squares method and a maximum likelihood method. The behavior of both estimators is illustrated by a simulation. Both methods have a smaller error in Frobenius norm than the sample covariance, moreover, the latter method performs better than the former one, which corresponds to its stronger assumptions.

Key Words: Data assimilation, discrete Fourier transform, spectral, diagonal covariance, maximum likelihood

1. Data assimilation into numerical weather model

In numerical weather models, the state of the atmosphere is understood as a collection of several (generally 3D) random fields of variables, such as pressure, temperature, wind vectors, etc. In practice, the model domain is represented as a discrete grid, with a vector of values of meteorological variables at each gridpoint. If we let the model evolve one time step, say, from $t-1$ to t, it provides a forecast for time t, $\mathbf{X}_t^{\mathrm{f}}$, which is a prior estimate of the true state $\mathbf{X}_t^*$ of the atmosphere. Now at time t, some variables $\mathbf{y}_t$ are observed, often not directly and with an error. There exist a large number of mathematical and statistical techniques for incorporating the observations into the model [4].

Data assimilation as a mathematical discipline encompasses these techniques together with a number of supporting methods. From the point of view of statistics, assimilation of data means finding a posterior estimate of the state. This posterior estimate is called "analysis" in meteorology. One of the classical schemes of data assimilation is the update step of the Kalman filter [3]

$$\begin{aligned} \mathbf{X}_t^{\mathrm{a}} &= \mathbf{X}_t^{\mathrm{f}} + G(\mathbf{y}_t - H\mathbf{X}_t^{\mathrm{f}}) \\ C^{\mathrm{a}} &= \mathrm{cov}\{\mathbf{X}_t^{\mathrm{a}}\} = (I - GH)C^{\mathrm{f}} \end{aligned}$$

*Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 271/2, 182 07 Praha 8, Czech Republic, and Charles University in Prague, Faculty of Mathematics and Physics, Sokolovská 83, Prague 8, 186 75, Czech republic

†University of Colorado Denver, Denver, CO 80217-3364, USA, and Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 271/2, 182 07 Praha 8, Czech Republic

‡Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 271/2, 182 07 Praha 8, Czech Republic

where $\mathbf{X}_t^a$ is the analysis, the matrix $G = C^{\mathrm{f}}H^\top(HP^{\mathrm{f}}H^\top + R)^{-1}$ is called Kalman gain, C^{f}, resp. R, stand for the covariance matrix of forecast errors, resp. observation errors, and H denotes an observation operator ($H\mathbf{X}_t^{\mathrm{f}}$ is a model counterpart of $\mathbf{y}_t$). The Kalman filter relies on many assumptions like linearity of the observation operator and unbiasedness of the forecast. $\mathbf{X}_t^{\mathrm{a}}$ is the best linear unbiased estimator of $\mathbf{X}_t^*$ [9].

In meteorology, the dimensions of the 3D grid are tens to hundreds. As a result, the dimension of the state vector is about 10^6 and the matrices C^{f} and C^{a} cannot be stored in the memory of a computer and treated in a usual manner. The aim is to estimate them cheaply, but with a sufficient accuracy.

In the rest of the paper we will consider the system state in one particular time so we will omit the subscript t.

2. Ensemble Kalman Filter

For the sake of simplicity, we shall deal with one variable on a 2D grid in a spatial domain $\mathcal{D}$. Thus, let $\mathbf{X}$ be a centered random field describing spatial distribution of some meteorological variable (e.g. temperature) on a horizontal grid with dimensions $M \times N$ which can be ordered to a vector of length MN if necessary.

One version of the Kalman filter that addresses the "dimension problem" is the Ensemble Kalman filter [1], where the distribution of system state is represented by an ensemble $\mathbb{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_S]$, $S << MN$. The ensemble represents a sample of stochastically independent draws from the distribution of $\mathbf{X}$.

The forecast system state is estimated through the ensemble mean

$$\bar{\mathbf{X}}^{\mathrm{f}} = \frac{1}{S}\sum_{s=1}^{S}\mathbf{X}_s^{\mathrm{f}}$$

and the covariance matrix is estimated by the sample covariance

$$\begin{aligned}\hat{C}^{\mathrm{f}} = \frac{1}{S-1}\sum_{s=1}^{S}(\mathbf{X}_s^{\mathrm{f}} - \bar{\mathbf{X}}^{\mathrm{f}})(\mathbf{X}_s^{\mathrm{f}} - \bar{\mathbf{X}}^{\mathrm{f}})^\top &= \frac{1}{S-1}\mathbb{X}^{\mathrm{f}}(I - \frac{1}{S}\mathbf{1}\mathbf{1}^\top)(\mathbb{X}^{\mathrm{f}})^\top \\ &= \mathbb{Z}^{\mathrm{f}}(\mathbb{Z}^{\mathrm{f}})^\top,\end{aligned}$$

where $\mathbb{Z}^{\mathrm{f}}$ denotes the matrix of forecast ensemble perturbations (its size is $MN \times S$).

Because of high computational cost, the size of the ensemble is small in real tasks (usually a few tens of members). It is well known that the sample covariance matrix suffers from low rank and so called spurious correlations arise, causing artefacts in the meteorological analysis. Methods like localization and covariance inflation (e.g., [2]) have been developed to deal with this problem, however, they always contain some heuristics (choice of the localization length, choice of the inflation factor). More robust and less heuristic estimates can be designed after introducing some further assumptions.

3. Laplace operator as a base for construction of covariance operators

The key idea is to model C^{f} through an operator. From now, we will focus mainly on a forecast ensemble and its covariance matrix, so we will omit the index f.

The covariance matrix C is only a discretization of covariance function. Covariance function is defined as

$$C(z, z') = \mathrm{E}\{\mathbf{X}(z)\mathbf{X}(z')\}$$

for $z, z' \in \mathcal{D}$. We will use C for both the covariance matrix and function. The covariance operator $\mathcal{C}$ associated with the covariance function C is defined as

$$(\mathcal{C}h)(z) = \int_{\mathcal{D}} C(z', z)h(z')dz'$$

for every $h \in L_2(\mathcal{D})$.

In [8, 10], the covariance operator $\mathcal{C} = e^{\Delta}$ was considered and suggested as suitable for weather prediction applications. Here Δ stands for the Laplace operator, which in two dimensions has the form

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

In these papers, it has been shown that this choice of $\mathcal{C}$ results in a Gaussian covariance function, which is homogeneous and isotropic. Moreover, the action of covariance operator e^{Δ} is equivalent to solving the classical diffusion equation, which brings many computational benefits.

These facts may serve as a motivation for a natural generalization, namely, to take $\mathcal{C}$ equal to various functions of Laplace operator Δ(or better, equal to functions of operator $(-\Delta)$, which has positive eigenvalues). In order to produce smooth random fields, we require such a function f of $(-\Delta)$ to be decreasing and positive on $\mathbb{R}^+$ and $\lim_{y\to\infty} f(y) = 0$. However, how to find the associated covariance function?

Suppose that $\lambda_1^* \geq \lambda_2^* \geq \ldots$ and $\psi_1, \psi_2, \ldots$ are the eigenvalues and normalized eigenfunctions of the operator $\mathcal{C}$, i.e.,

$$\lambda_k^* \psi_k(z) = \int_{\mathcal{D}} C(z', z)\psi_k(z')dz', \quad k = 1, 2, \ldots,$$

$$\int_{\mathcal{D}} \psi_j(z)\psi_k(z)dz = \begin{cases} 1, & j = k \\ 0, & j \neq k. \end{cases}$$

Having the eigenvalues and eigenfunctions of operator $\mathcal{C}$, we can express the covariance function C using the Mercer expansion [7]

$$C(z, z') = \sum_{k=1}^{\infty} \lambda_k^* \psi_k(z)\psi_k(z').$$

The eigenfunctions of $f(-\Delta)$ are the same as those of the Laplace operator and its eigenvalues are $\lambda_k^* = f(\lambda_k)$, where λ_k denote the eigenvalues of $(-\Delta)$, by the spectral mapping theorem. This gives us a possibility of widening the class of available covariance operators in this way.

Apart from global models, a typical weather prediction task works on a rectangular domain $\mathcal{D}$, say, $[0, a] \times [0, b]$. It is well known that for the operator $(-\Delta)$ on the rectangle it holds

$$\begin{aligned} \lambda_{mn} &= \left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2\right] \\ \psi_{mn}(x, y) &= \sin\left(\frac{x}{a}m\pi\right)\sin\left(\frac{y}{b}n\pi\right), \end{aligned}$$

where $(x, y) \in \mathcal{D}$ and $m, n \in \mathbb{N}$.

We have studied two particular choices of $\mathcal{C}$: $(-\Delta)^{-\alpha}$ and $ce^{-\alpha(-\Delta)^p}$. The first one gives the covariance function

$$\begin{aligned} C((x,y),(x',y')) &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\left[\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2\right]^{-\alpha}\cdot \\ &\quad \sin\left(\frac{x}{a}m\pi\right)\sin\left(\frac{y}{b}n\pi\right)\sin\left(\frac{x'}{a}m\pi\right)\sin\left(\frac{y'}{b}n\pi\right) \end{aligned}$$

and the second one the covariance function

$$\begin{aligned} C((x,y),(x',y')) &= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} ce^{-\alpha\left[\left(\frac{m\pi}{a}\right)^2+\left(\frac{n\pi}{b}\right)^2\right]^p}\cdot \\ &\quad \sin\left(\frac{x}{a}m\pi\right)\sin\left(\frac{y}{b}n\pi\right)\sin\left(\frac{x'}{a}m\pi\right)\sin\left(\frac{y'}{b}n\pi\right), \end{aligned}$$

where $(x, y), (x', y') \in \mathcal{D}$ and $m, n \in \mathbb{N}$.

In practice, the random field $\mathbf{X}$ is represented by an $M \times N$ grid and its covariance function by a matrix. The Mercer expansion, which determines the covariance between two nodes $[i, j]$ and $[i', j']$ of the grid, has the discrete form

$$\begin{aligned} C([i,j],[i',j']) &= \frac{4}{(M+1)(N+1)}\sum_{m=1}^{M}\sum_{n=1}^{N} d_{mn}\cdot \\ &\quad \sin\left(\frac{\pi}{M+1}im\right)\sin\left(\frac{\pi}{N+1}jn\right)\cdot \\ &\quad \sin\left(\frac{\pi}{M+1}i'm\right)\sin\left(\frac{\pi}{N+1}j'n\right), \end{aligned} \tag{1}$$

where $i, i' = 1, \ldots, M$ and $j, j' = 1, \ldots, N$. In this discrete situation, it is necessary to use eigenvalues of the discrete version of the chosen operator $\mathcal{C}$. We denoted these eigenvalues by d_{mn} without giving their explicit forms because these are not important for what follows.

Expression (1) can be written in a matrix form as

$$C = \mathrm{Cov}(\mathbf{X}) = F^\top D F,$$

where F is a matrix formed from the eigenvectors of $\mathcal{C}$ and D is a diagonal matrix with eigenvalues d_{mn} on its diagonal. The reader has probably noticed that eigenvectors of Laplace operator represent the sine transform, therefore matrix F can be viewed as an orthonormal transformation matrix, which transform objects to spectral space. Since $FF^\top = I$, we have

$$\mathrm{Cov}(F\mathbf{X}) = FCF^\top = D,$$

so the covariance matrix in the spectral space is diagonal. This gives us a possibility to truncate the sample covariance matrix as it is shown in [6] and [5]. The method consists of computing the sample covariance matrix in spectral space (after applying F to each member $\mathbf{X}_s$, $s = 1, \ldots, S$) and setting all off-diagonal terms equal to zero. This diagonalized matrix $\tilde{D}$ is then transformed back to the original space (using operator $F^\top$), which results in an improved estimate $F^\top \tilde{D} F$. However, we can follow this idea further and try to model the diagonal terms of matrix $\tilde{D}$ in order to reduce sampling noise.

4. Fitting parametric models for diagonal elements of the covariance matrix

In modeling the spectral diagonal covariance, we may easily generalize the base form $\mathcal{C} = e^{\Delta}$ towards faster or slower decay of eigenvalues by adopting a parametric model for the eigenvalues. In particular, we have studied the cases $\mathcal{C} = (-\Delta)^{-\alpha}$ (a random field less smooth than the base case $e^{-(-\Delta)}$) and $\mathcal{C} = ce^{-\alpha(-\Delta)^p}$ for $p > 1$ (a smoother random field). As stated above, the eigenvectors are known while the eigenvalues are known up to the parameters α, resp. (c,α). We can estimate them and use them to model the diagonal of matrix $\tilde{D}$. The resulting matrix will be denoted as $\hat{D}$.

Consider the smoother case $\mathcal{C} = ce^{-\alpha(-\Delta)^p}$ and look at the estimation in more detail. The first method used for estimating the parameters is the weighted least squares method applied to the logarithms of the diagonal terms. This method does not require any additional assumptions and its estimation criterion has the form

$$\min_{c,\alpha} \sum_{k=1}^{MN} (\log c - \alpha\lambda_k^p - \log \tilde{d}_k)^2 w_k^2 \tag{2}$$

where $\tilde{d}_k$ are the diagonal entries of $\tilde{D}$, λ_k are the eigenvalues of $(-\Delta)$, $p > 0$ is a fixed parameter, and w_k are chosen weights, $k = 1, \ldots, MN$.

For ease of notation, we now drop the summation bounds and write only $\sum_k$. Estimates of parameters c and α can be found in a standard manner by differentiating (2) and setting the derivative equal to zero:

$$\begin{aligned}
\log \hat{c} &= \frac{\sum_k w_k^2\lambda_k^{2p} \sum_k w_k^2 \log \tilde{d}_k - \sum_k w_k^2\lambda_k^p \sum_k w_k^2\lambda_k^p \log \tilde{d}_k}{\sum_k w_k^2 \sum_k w_k^2\lambda_k^{2p} - (\sum_k w_k^2\lambda_k^p)^2} \\
\hat{\alpha} &= \frac{\sum_k w_k^2\lambda_k^p \sum_k w_k^2 \log \tilde{d}_k - \sum_k w_k^2 \sum_k w_k^2\lambda_k^p \log \tilde{d}_k}{\sum_k w_k^2 \sum_k w_k^2\lambda_k^{2p} - (\sum_k w_k^2\lambda_k^p)^2}.
\end{aligned}$$

Hence, the j-th diagonal entry of the smoothed covariance matrix in spectral space equals to

$$\begin{aligned}
\hat{d}_j = \exp\Bigg(& \frac{\sum_k w_k^2\lambda_k^{2p} - \lambda_j^p \sum_k w_k^2\lambda_k^p}{\sum_k w_k^2 \cdot \sum_k w_k^2\lambda_k^{2p} - \left(\sum_k w_k^2\lambda_k^p\right)^2} \sum_k w_k^2 \log \tilde{d}_k - \\
& \frac{\sum_k w_k^2\lambda_k^p - \lambda_j^p \sum_k w_k^2}{\sum_k w_k^2 \cdot \sum_k w_k^2\lambda_k^{2p} - \left(\sum_k w_k^2\lambda_k^p\right)^2} \sum_k w_k^2\lambda_k^p \log \tilde{d}_k \Bigg),
\end{aligned}$$

where all summations are carried over the index $k = 1, \ldots, MN$.

If we do not know how to choose the weights we can set all weights equal to 1 and compute classical least squares estimates. Then

$$\begin{aligned}
\hat{d}_j = \exp\Bigg(& \frac{\sum_k \lambda_k^{2p} - \lambda_j^p \sum_k \lambda_k^p}{MN \sum_k \lambda_k^{2p} - \left(\sum_k \lambda_k^p\right)^2} \sum_k \log \tilde{d}_k - \\
& \frac{\sum_k \lambda_k^p - MN\lambda_j^p}{MN \sum_k \lambda_k^{2p} - \left(\sum_k \lambda_k^p\right)^2} \sum_k \lambda_k^p \log \tilde{d}_k \Bigg),
\end{aligned} \tag{3}$$

where all summations are again carried over the index $k = 1, \ldots, MN$.

If we can further assume that the underlying random field is Gaussian, we can estimate the parameters c and α via maximizing the likelihood. Assume that the original field is normally distributed: $\mathbf{X}_i \sim N_K(\mathbf{0}, C)$, $i = 1, \ldots, S$. Hence, each ensemble member in spectral space

$$F\mathbf{X}_i = \mathbf{U}_i = [U_{i1}, \ldots, U_{iK}], \quad i = 1, \ldots, S,\ K = MN$$

comes from the distribution $N_K(\mathbf{0}, D)$, and the likelihood has the form

$$\begin{aligned} L &= \prod_{i=1}^{S} g(\mathbf{U}_i) = \left[(2\pi)^K |D|\right]^{-\frac{S}{2}} e^{-\frac{1}{2}\sum_{i=1}^{S} \mathbf{U}_i^\top D^{-1} \mathbf{U}_i} \\ &= \left[(2\pi)^K \prod_{k=1}^{K} d_k\right]^{-\frac{S}{2}} e^{-\frac{1}{2}\sum_{i=1}^{S}\sum_{k=1}^{K} d_k^{-1} U_{ik}^2}, \end{aligned}$$

where g denotes the density of $N_K(\mathbf{0}, D)$. The log-likelihood is

$$\begin{aligned} l &= \log L \\ &= -\frac{KS}{2}\log(2\pi) - \frac{S}{2}\sum_{k=1}^{K}\log d_k - \frac{1}{2}\sum_{k=1}^{K} d_k^{-1}\sum_{i=1}^{S} U_{ik}^2 \\ &= -\frac{KS}{2}\log(2\pi) - \frac{S}{2}\sum_{k=1}^{K}(\log c - \alpha\lambda_k^p) - \frac{1}{2}\sum_{k=1}^{K} e^{-\log c + \alpha\lambda_k^p} Q_k^2, \end{aligned}$$

where d_k were replaced by $\exp(\log c - \alpha\lambda_k^p)$ and Q_k^2 denotes the sum of squares $\sum_{i=1}^{S} U_{ik}^2$ (a sufficient statistic for variance).

Now, differentiate the log-likelihood with respect to c and α and form the likelihood equations

$$\begin{aligned} \frac{\partial l}{\partial c} &= -\frac{SK}{2}\frac{1}{c} + \frac{1}{2}\sum_{k=1}^{K} e^{-\log c + \alpha\lambda_k^p}\frac{1}{c} Q_k^2 \stackrel{!}{=} 0 \\ \frac{\partial l}{\partial \alpha} &= \frac{S}{2}\sum_{k=1}^{K}\lambda_k^p - \frac{1}{2}\sum_{k=1}^{K} e^{-\log c + \alpha\lambda_k^p}\lambda_k^p Q_k^2 \stackrel{!}{=} 0. \end{aligned}$$

These equations lead us to the implicit form of parameters' estimates

$$\hat{c} = \frac{1}{SK}\sum_{k=1}^{K} e^{\hat{\alpha}\lambda_k^p} Q_k^2 \tag{4}$$

$$0 = \sum_{k=1}^{K} e^{\hat{\alpha}\lambda_k^p} Q_k^2 (\lambda_k^p - \bar{\lambda^p}), \tag{5}$$

where $\bar{\lambda^p} = \frac{1}{K}\sum_{k=1}^{K}\lambda_k^p$. The values of $\hat{c}$ and $\hat{\alpha}$ can be computed from (4,5) numerically.

The j-th diagonal entry of the smoothed covariance matrix in spectral space equals to

$$\hat{d}_j = \hat{c} e^{-\hat{\alpha}\lambda_j^p}, \tag{6}$$

where $\hat{c}$ and $\hat{\alpha}$ are given by (4) and (5).

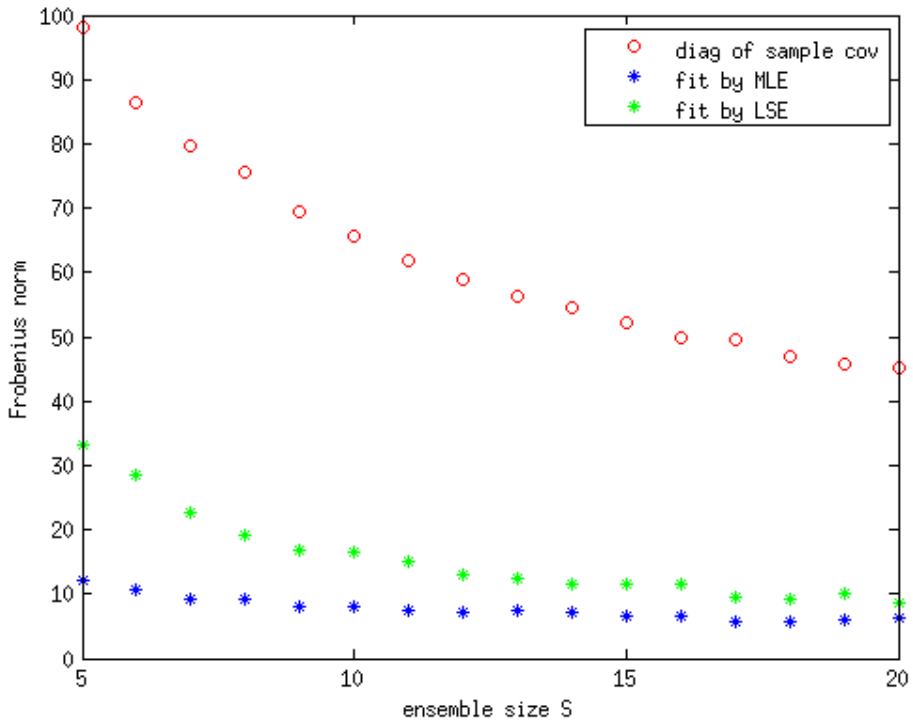

Figure 1: Comparison of the error matrix $\hat{D} - D$ in the Frobenius norm. The field had dimension 10×10. Exponential decay of eigenvalues (i.e. $ce^{-\alpha\lambda}$) was used with parameters $c = 30$ and $\alpha = 0.002$.

4.1 Comparing both estimates

We illustrate the behaviour of both described estimates - least squares estimate (LSE) and maximum likelihood estimate (MLE) - on a set of simulated Gaussian fields with the covariance operator $\mathcal{C} = ce^{\alpha\Delta}$ with $c = 30$ and $\alpha = 0.002$.

In Figure 1, it can be seen how the Frobenius norm of the error matrix $||\hat{D} - D||_{\text{Frob}}$ depends on the ensemble size. Entries of matrix $\hat{D}$ have been estimated using the formula (3), resp. (6), with $p = 1$. It is evident that in both cases, the fit is better than the diagonal of sample covariance matrix, especially for small ensembles. Apparently, the MLE is better than the LSE, which is not surprising because maximum likelihood estimates use the additional information about the original distribution. It should be noted that for both MLE and LSE, the error is small already with very small ensembles, which is due to the fact that the covariance of the Gaussian field is actually of the parametric form given, and we are only using the samples to identify the two parameters.

Acknowledgements

This research was partially supported by the Czech Science Foundation under the grant GA13-34856S and the US National Science Foundation under the grant DMS-1216481. A part of this research was done when Marie Turčičová was visiting the University of Colorado Denver.

References

[1] Evensen, G. (1994), "Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, 99 (C5), 143–162.

[2] Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007), "Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter," *Physica D: Nonlinear Phenomena*, 230, 112–126.

[3] Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME – Journal of Basic Engineering, Series D*, 82, 35–45.

[4] Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press.

[5] Kasanický, I., Mandel, J., and Vejmelka, M. (2015), "Spectral diagonal ensemble Kalman filters," *Nonlinear Processes in Geophysics*, 22, 485 – 497.

[6] Mandel, J., Beezley, J. D., and Kondratenko, V. Y. (2010), "Fast Fourier Transform Ensemble Kalman Filter with Application to a Coupled Atmosphere-Wildland Fire Model," in *Computational Intelligence in Business and Economics, Proceedings of MS'10*, eds. Gil-Lafuente, A. M. and Merigo, J. M., World Scientific, 777–784.

[7] Mercer, J. (1909), "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 209, 415–446.

[8] Mirouze, I. and Weaver, A. T. (2010), "Representation of correlation functions in variational assimilation using an implicit diffusion operator," *Quarterly Journal of the Royal Meteorological Society*, 136, 1421–1443.

[9] Talagrand, O. (1997), "Assimilation of observations, an Introduction," *Journal of the Meteorological Society of Japan*, 75 (1B), 191–209.

[10] Weaver, A. and Courtier, P. (2001), "Correlation modelling on the sphere using a generalized diffusion equation," *Quarterly Journal of the Royal Meteorological Society*, 127, 1815–1846.