Adjoints and low-rank covariance representation

Michael K. Tippett[1] and Stephen E. Cohn[2]

[1]International Research Institute for climate prediction, LDEO, Palisades, NY, USA.
[2]Data Assimilation Office, NASA/GSFC, Greenbelt, MD, USA.

Manuscript submitted to

Nonlinear Processes in Geophysics

Abstract

Quantitative measures of the uncertainty of Earth System estimates can be as important as the estimates themselves. Second moments of estimation errors are described by the covariance matrix, whose direct calculation is impractical when the number of degrees of freedom of the system state is large. Ensemble and reduced-state approaches to prediction and data assimilation replace full estimation error covariance matrices by low-rank approximations. The appropriateness of such approximations depends on the spectrum of the full error covariance matrix, whose calculation is also often impractical. Here we examine the situation where the error covariance is a linear transformation of a forcing error covariance. We use operator norms and adjoints to relate the appropriateness of low-rank representations to the conditioning of this transformation. The analysis is used to investigate low-rank representations of the steady-state response to random forcing of an idealized discrete-time dynamical system.

Correspondence to: ff

1 Introduction

The value of an estimate of the Earth system state is increased when accompanied by a measure of its uncertainty. Sophisticated users of weather and climate predictions base decisions on both forecast and forecast uncertainty (Changnon et al., 1999). Data assimilation systems combine information from different sources in a manner that depends upon its presumed uncertainty (Cohn, 1997). Ensemble prediction systems generate ensembles of initial conditions with statistics that attempt to reflect analysis errors (Barkmeijer et al., 1998).

Uncertainty can be modeled as a random variable which is completely described by a probability density function (pdf). A partial description of the pdf is given by its mean and covariance. With some simplifying assumptions, equations can be obtained for the evolution of the mean and covariance. However, for realistic dynamical systems with a large number of degrees of freedom, the direct solution of these equations, particularly those for the covariance, is impractical due to computational costs and to incomplete knowledge of the sources of uncertainty. Therefore, in many situations the covariance must be modeled.

One approach to covariance modeling is to specify analytically parameters such as the variances and correlation lengths (Dee and da Silva, 1998; Rabier et al., 1998). Complex features such as flow dependence may also be modeled through appropriate parameterization (Riishøjgaard, 1998). A second approach is to assume that the error variability is described well by just a few dominant structures and hence that the covariance matrix is approximated by a low-rank matrix. Low-dimensional covariance representations are often directly connected to the dynamics, as in reduced-state Kalman filter data assimilation and ensemble prediction (Cane et al., 1996; Cohn and Todling, 1996; Houtekamer and Mitchell, 1998; Molteni et al., 1996).

Comparison of low-dimensional covariance representations with the full covariance is only possible in idealized models where the number of degrees of freedom is small and the sources of uncertainty are specified (Cohn and Todling, 1996; Kleeman and Moore, 1997; Whitaker and Sardeshmukh, 1998). Here our approach is to examine the equation satisfied by the covariance and to determine which of its properties control the covariance spectrum. We consider the situation where the error covariance is a linear transformation of a forcing error covariance and seek the properties of the linear transformation that affect the covariance spectrum.

The paper is organized as follows. Section 2 introduces low-rank representations and the linear covariance equation. Section 3 analyzes this equation using first the singular value decomposition and then operator norms. Section 4 applies the analysis to the discrete-time algebraic Lyapunov equation and illustrates the results with an example using the dynamics of a generalized advection equation. Conclusions are given in Section 5.

2 Mathematical setting

2.1 Covariance

We suppose that our system state ϵ, a real vector of length n, is a mean-zero random variable. In typical problems, ϵ might be forecast or analysis error, or climatological anomaly. Considerable information about the system state is contained in the $n \times n$ covariance matrix P defined by

$$\mathsf{P} \equiv \left\langle \epsilon\epsilon^T \right\rangle \tag{1}$$

where $()^T$ denotes transpose and $\langle\cdot\rangle$ denotes expectation. The total variance of ϵ is given by $\operatorname{tr}\mathsf{P}$ where tr denotes trace. The eigenvectors of P order state-space directions according to the variance they explain

A measure of how well a rank-k matrix $\mathsf{P}^{(k)}$ approximates the covariance matrix P is the quantity $||\mathsf{P} - \mathsf{P}^{(k)}||$; $||\cdot||$ is a matrix norm, for instance the Schatten p-norm defined by

$$||\mathsf{X}||_p = \left(\sum_{i=1}^{n} \sigma_i^p(\mathsf{X})\right)^{1/p}, \quad 1 \leq p \leq \infty\,, \tag{2}$$

where $\sigma_i(\mathsf{X})$ is the i-th singular value of X. This family of matrix norms includes several common matrix norms; for covariance matrices, $||\mathsf{P}||_1 = \operatorname{tr}\mathsf{P}$ and $||\mathsf{P}||_\infty = \lambda_1(\mathsf{P})$ where $\lambda_i(\mathsf{P})$ is the i-th eigenvalue of P ordered in decreasing modulus.

The optimal rank-k approximation $\mathsf{P}^{(k)}$ of P in the Schatten p-norm is its projection onto its leading k eigenvectors given by

$$\mathsf{P}^{(k)} = \sum_{i=1}^{k} \lambda_i(\mathsf{P}) w_i w_i^T \tag{3}$$

where w_i is the orthonormal eigenvector of P corresponding to the eigenvalue $\lambda_i(\mathsf{P})$. In this case, the error of the approximation is

$$\|\mathsf{P} - \mathsf{P}^{(k)}\|_p = \left\| \sum_{i=k+1}^{n} \lambda_i^p(\mathsf{P}) \right\|^{1/p} , \tag{4}$$

and the appropriateness of low-rank representations of P depends on its spectrum. However, direct calculation of the eigenvalues of P is often not practical.

2.2 Linear covariance equation

Our approach is to to identify properties of the equation satisfied by P that control the spectrum of P. We consider the situation where the covariance matrix P satisfies an equation of the form

$$\mathsf{P} = \mathcal{L}\mathsf{P}_0 \, ; \tag{5}$$

$\mathcal{L}$ is a linear operator on matrices and P_0 is a forcing covariance matrix. Many systems where the covariance evolution is linear and the forcing is stationary can be written in this form. An example of a system with this form comes from the stochastically-forced discrete-time linear system

$$\epsilon_{k+1} = \mathsf{A}_k \epsilon_k + b_{k+1} \, , \quad \left\langle b_i b_j^T \right\rangle = \delta_{ij} \mathsf{P}_0 \, ; \tag{6}$$

ϵ_k is the system state at time-step k, A_k is the dynamics, b_k is the random forcing and δ_{ij} is the Kronecker delta. The covariance P_k defined by

$$\mathsf{P}_k \equiv \left\langle \epsilon_k \epsilon_k^T \right\rangle \, . \tag{7}$$

satisfies the discrete-time covariance evolution equation

$$\mathsf{P}_{k+1} = \mathsf{A}_k \mathsf{P}_k \mathsf{A}_k^T + \mathsf{P}_0 = \sum_{i=0}^{k} \mathsf{A}^k \mathsf{P}_0 (\mathsf{A}^k)^T \equiv \mathcal{L}\mathsf{P}_0 \, , \tag{8}$$

which has the form of (5); we use the notation $\mathsf{A}^k = \mathsf{A}_k\mathsf{A}_{k-1}\cdots\mathsf{A}_2\mathsf{A}_1$.

In the special case of stable[1] time-independent dynamics $\mathsf{A}_k = \mathsf{A}$, the equation for the steady-state covariance matrix P is

$$\mathcal{L}_{\mathsf{A}}^{-1}\mathsf{P} \equiv \mathsf{P} - \mathsf{A}\mathsf{P}\mathsf{A}^T = \mathsf{P}_0\,. \tag{9}$$

We will examine in some detail equation (9) known as the discrete-time algebraic Lyapunov equation (DALE) (Kwon et al., 1996; Tippett and Marchesin, 1999a,b; Tippett et al., 2000a). We refer to $\mathcal{L}_{\mathsf{A}}$ as the DALE operator; this nomenclature and notation is differs from that of Tippett et al. (2000a). Covariance evolution in a continuous-time system can also be expressed in the form of (5) (Byers and Nash, 1987).

We next examine what properties of (5) cause its solution P to have a good low-rank approximation. Since the initial or forcing covariance P_0 is often poorly known, results that require knowing details of P_0 are unsatisfactory. Results that depend mostly on $\mathcal{L}$ and require limited knowledge of P_0 are desirable.

3 Analysis of linear matrix equations

3.1 Inner products

Linear algebra methods such as the eigenvalue and singular value decompositions can be used to analyze (5), treating the covariance matrices as vectors of length n^2 and $\mathcal{L}$ as an $n^2 \times n^2$ matrix (Byers and Nash, 1987; Ghavimi and Laub, 1995). Eigenmode analysis is appropriate when the linear operator $\mathcal{L}$ is *normal*, that is when $\mathcal{L}$ has a complete set of orthogonal eigenvectors. The normality of $\mathcal{L}$ depends on the definition of orthogonality and hence on the choice of inner product on matrices. Any inner product $(\cdot,\cdot)$ on vectors can be written in the form

$$(\epsilon_1, \epsilon_2) = \epsilon_1^\dagger \mathsf{M} \epsilon_2 \tag{10}$$

where M is a Hermitian positive definite matrix and $()^\dagger$ denotes conjugate transpose. The adjoint A^* of A with respect to this inner product is $\mathsf{A}^* = \mathsf{M}^{-1}\mathsf{A}^\dagger\mathsf{M}$. We may assume

[1] The discrete-time dynamics A is stable if and only if the eigenvalues of A lie inside the unit circle.

$\mathsf{M} = \mathsf{I}$ without loss of generality since a new state variable $\hat{\epsilon} = \mathsf{M}^{1/2}\epsilon$ can be introduced with $(\epsilon_1, \epsilon_2) = \hat{\epsilon}_1^\dagger \hat{\epsilon}_2$.

A natural matrix inner product, though not the most general, is defined by

$$(\mathsf{X}, \mathsf{Y}) = \operatorname{tr} \mathsf{X}^\dagger \mathsf{M} \mathsf{Y} \,. \tag{11}$$

The matrix inner product (11) is compatible with the vector inner product (10) in the sense that the orthogonality of two rank-1 matrices $\epsilon_1\epsilon_1^\dagger$ and $\epsilon_2\epsilon_2^\dagger$ is equivalent to the orthogonality of ϵ_1 and ϵ_2. The DALE operator $\mathcal{L}_\mathsf{A}$ has the pleasant properties with respect to this inner product that $\mathcal{L}_\mathsf{A}^* = \mathcal{L}_{\mathsf{A}^*}$ and that $\mathcal{L}_\mathsf{A}$ being normal is equivalent to A being normal. We mention finally that for $\mathsf{M} = \mathsf{I}$ the matrix inner product (11) is the Euclidean inner product on vectors of length n^2 since

$$\operatorname{tr} \mathsf{X}^\dagger \mathsf{Y} = \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{\mathsf{X}}_{ij} \mathsf{Y}_{ij} \,. \tag{12}$$

3.2 Eigenvectors and singular vectors

The singular value decomposition of a linear operator gives its rank, range, null space, 2-norm and optimal low-rank approximations. Approximations of $\mathcal{L}$ that are low-rank in the space of $n \times n$ matrices are obtained from its singular value decomposition; when $\mathcal{L}$ is normal it is sufficient to use the eigenvalue decomposition. Suppose that $\sigma_i(\mathcal{L})$, U_i, and V_i are respectively the i-th singular value, left singular vector and right singular vector of $\mathcal{L}$. That is, the $n \times n$ matrices U_i and V_i satisfy $\mathcal{L}\mathsf{V}_i = \sigma_i(\mathcal{L})\mathsf{U}_i$ and $(\mathsf{U}_i, \mathsf{U}_j) = (\mathsf{V}_i, \mathsf{V}_j) = \delta_{ij}$ where $(\cdot, \cdot)$ is the matrix inner product. Then the decomposition of P in the left singular vectors of $\mathcal{L}$ is

$$\mathsf{P} = \mathcal{L}\mathsf{P}_0 = \sum_{i=1}^{n^2} \sigma_i(\mathcal{L}) \; (\mathsf{V}_i, \mathsf{P}_0) \; \mathsf{U}_i \,. \tag{13}$$

Approximate solutions of (5) are found using low-rank approximations of $\mathcal{L}$. If $\mathcal{L}$ has some large singular values, for instance, suppose that $\sigma_k(\mathcal{L}) \gg \sigma_{k+1}(\mathcal{L})$, then a natural approximation $\tilde{\mathsf{P}}$ of P is obtained using the optimal (in the space of $n \times n$ matrices) rank-k approximation $\mathcal{L}^{(k)}$ of $\mathcal{L}$ to obtain

$$\tilde{\mathsf{P}} = \mathcal{L}^{(k)}\mathsf{P}_0 = \sum_{i=1}^{k} \sigma_i(\mathcal{L}) \; (\mathsf{V}_i, \mathsf{P}_0) \; \mathsf{U}_i \,. \tag{14}$$

with error

$$\|\mathsf{P} - \tilde{\mathsf{P}}\|_2 = \mathcal{L}\mathsf{P}_0 = \sum_{i=k+1}^{n^2} \sigma_i(\mathcal{L})\, |(\mathsf{V}_i, \mathsf{P}_0)| \ . \tag{15}$$

So if the projection of P_0 onto the first k right singular vectors of $\mathcal{L}$ is not too small, P shares with $\tilde{\mathsf{P}}$ a large component in the directions spanned by the first k left singular vectors of $\mathcal{L}$. However, there is no guarantee that the approximation $\tilde{\mathsf{P}}$ is low-rank unless there is some information about the rank of the left singular vectors of $\mathcal{L}$. For instance, if the singular vectors of $\mathcal{L}$ are rank-1 matrices, then $\tilde{\mathsf{P}}$ is at most rank-k, though not, in general, the optimal rank-k approximation $\mathsf{P}^{(k)}$. In general the singular vectors of $\mathcal{L}$ are not rank-1 matrices.

An example where $\mathcal{L}$ is normal and its eigenvectors are rank-1 matrices is the normal DALE operator $\mathcal{L}_{\mathsf{A}}$. The eigenvalues and eigenvectors of $\mathcal{L}_{\mathsf{A}}$ are $(1 - \lambda_i(\mathsf{A})\overline{\lambda_j(\mathsf{A})})^{-1}$ and $z_i z_j^\dagger$ respectively where $\lambda_i(\mathsf{A})$ and z_i are respectively the eigenvalues and eigenvectors of A (Lancaster, 1970). The DALE operator $\mathcal{L}_{\mathsf{A}}$ is normal if and only if A is normal, in which case the eigenvectors z_i of A form a complete and orthonormal basis, and P_0 can be decomposed in this basis as

$$\mathsf{P}_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(z_i^\dagger \mathsf{P}_0 z_j\right) z_i z_j^\dagger \, . \tag{16}$$

The representation

$$\mathsf{P} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{z_i^\dagger \mathsf{P}_0 z_j}{1 - \lambda_i(\mathsf{A})\overline{\lambda_j(\mathsf{A})}} z_i z_j^\dagger \, , \tag{17}$$

of the solution of (9) shows that there are two extreme situations where the covariance matrix P has a dominant low-rank part; in one case the spectrum of P is determined by that of A, in the other by P_0.

In the first situation A is nearly unstable with some eigenvalues near the unit circle. In this case, P has large components that project onto the leading eigenvectors of A, if the projection of P_0 onto the leading eigenvectors of A does not happen to be small. For example, for $\mathsf{P}_0 = \mathsf{I}$ and normal A, the solution of (9) is

$$\mathsf{P} = \sum_{i=1}^{n} \frac{1}{1 - |\lambda_i(\mathsf{A})|^2} z_i z_i^\dagger \, , \tag{18}$$

and the eigenvectors of A are eigenvectors of P. In this case, P has a good low-rank representation if A has a few, but not all of its eigenvalues near the unit circle. The second situation occurs when P_0 has some large eigenvalues and the spectrum of A is fairly flat. Then P has large components on the space spanned by the leading eigenvectors of P_0. For instance, when $\mathsf{A} = c\mathsf{I}$, $0 \leq c < 1$, P given by

$$\mathsf{P} = \frac{1}{1-c^2}\mathsf{P}_0 \,, \tag{19}$$

has a good low-rank representation if P_0 does. These two mechanisms can interfere with each other and examples can be constructed where P_0 has large eigenvalues and A is nearly unstable but the spectrum of P is flat.

There is no correspondingly simple analysis of the DALE for nonnormal dynamics. However, in the next section we show that features seen in the analysis of the normal DALE operator hold both for the nonnormal DALE operator and for general $\mathcal{L}$. Namely for the DALE we show the relation between the stability of A and the spectrum of P and for general $\mathcal{L}$ the connection between the conditioning of $\mathcal{L}$ and the spectrum of P.

3.3 Operator norms and positive maps

Examining the eigenvalue and singular value decompositions of the linear operator $\mathcal{L}$ does not use the fact that $\mathcal{L}$ is an operator on matrices and hence does not necessarily give information about how $\mathcal{L}$ relates the matrix properties of P_0 and P. This kind of information can be found in the operator norm of $\mathcal{L}$ defined by

$$||\mathcal{L}||_p = \max_{\mathsf{X}} \frac{||\mathcal{L}\mathsf{X}||_p}{||\mathsf{X}||_p} \,. \tag{20}$$

The operator norm $||\mathcal{L}||_p$ relates the matrix properties of P and P_0, though, for general operators there is not a simple expression for its value.

A linear operator $\mathcal{L}$ with the property that it maps covariance matrices to covariance matrices is *positive map* and hence has the property that its ∞-norm is simple to characterize and that its conditioning is related to the spectrum of P (Bhatia, 1997; Tippett et al., 2000a). Many properties of solutions of the DALE and of the continuous time algebraic Lyapunov equation (CALE) are direct consequences of the solutions being given by positive maps (Bhatia, 1997). For instance, the fact that positive maps have their maximum response in the $p = \infty$ norm to uncorrelated forcing

$$\|\mathcal{L}\|_\infty = \max_{\mathsf{P}_0} \frac{\lambda_1(\mathcal{L}\mathsf{P}_0)}{\lambda_1(\mathsf{P}_0)} = \max_{\mathsf{P}_0} \frac{\lambda_1(\mathsf{P})}{\lambda_1(\mathsf{P}_0)} = \lambda_1(\mathcal{L}\mathsf{I}), \tag{21}$$

has consequences for solutions of the DALE and CALE (Hewer and Kenney, 1988; Gahinet et al., 1990; Kenney and Hewer, 1990). The eigenvectors of the "bound matrix" $\mathsf{B} \equiv \mathcal{L}\mathsf{I}$ order the state space directions according to the possible response of the system there and can be used in estimates (Tippett and Marchesin, 1999a).

Similarly, results relating the forcing P_0 that produces maximum variance to the response of the adjoint system to uncorrelated error forcing (Kenney and Hewer, 1995; Farrell and Ioannou, 1996; Kleeman and Moore, 1997) also hold for any positive map. Since the 1-norm measuring variance response is dual to the ∞-norm, $\|\mathcal{L}\|_1 = \|\mathcal{L}^*\|_\infty = \|\mathcal{L}^*\mathsf{I}\|_\infty$. Additionally, the linear operator $\mathcal{L}$ achieves its 1-norm on the rank-1 matrix ww^T where w is the leading eigenvector of $\mathcal{L}^*\mathsf{I}$ since

$$\|\mathcal{L}ww^T\|_1 = (\mathsf{I}, \mathcal{L}ww^T) = w^T(\mathcal{L}^*\mathsf{I})w = \lambda_1(\mathcal{L}^*\mathsf{I}). \tag{22}$$

The eigenvectors of the $\mathsf{B}_T \equiv \mathcal{L}^*\mathsf{I}$ are the stochastic optimals, ordering state space according to how much variance is excited by forcing there as shown by the relation $\operatorname{tr}\mathsf{P} = \operatorname{tr}\mathcal{L}\mathsf{P}_0 = (\mathsf{P}_0, \mathsf{B}_T)$ (Bhatia, 1997).

The operator norm of $\mathcal{L}$ is also related to low-rank representations of P. A necessary condition for P to admit a low-rank representation is that it be ill-conditioned, that is, for the condition number $\kappa_\infty(\mathsf{P})$ to be large where $\kappa_\infty(\mathsf{P}) = \|\mathsf{P}\|_\infty \|\mathsf{P}^{-1}\|_\infty$ The conditioning of P depends on that of $\mathcal{L}$ and of P_0 as shown by the relation

$$\kappa_\infty(\mathsf{P}) \le \kappa_\infty(\mathcal{L})\kappa_\infty(\mathsf{P}_0); \tag{23}$$

more general results are available for $\mathcal{L}_\mathsf{A}$ (Tippett et al., 2000a). Therefore, for the special case $\mathsf{P}_0 = \mathsf{I}$, P can be ill-conditioned and have a good low-rank approximation only if $\mathcal{L}$ is ill-conditioned.

Another connection between low-rank representations of P and the operator norm of $\mathcal{L}$ is seen in the relation

$$\frac{\lambda_1(\mathsf{P})/\operatorname{tr}\mathsf{P}}{\lambda_1(\mathsf{P}_0)/\operatorname{tr}\mathsf{P}_0} \le \|\mathcal{L}\|_\infty \|\mathcal{L}^{-1}\|_1, \tag{24}$$

which follows directly from the operator norm definition and bounds the fraction of the total variance explained by the first eigenmode of P. A significant fraction of the total variance of P can be in its first eigenmode if the same is true of P_0, or if the quantity $\|\mathcal{L}\|_\infty \|\mathcal{L}^{-1}\|_1$ is large.

An additional useful property of positive maps is that they preserve ordering[2]. That is, if one has upper and lower bounds $\mathsf{P}_0^- \leq \mathsf{P}_0 \leq \mathsf{P}_0^+$, then

$$\mathcal{L}\mathsf{P}_0^- \leq \mathsf{P} \leq \mathcal{L}\mathsf{P}_0^+ . \tag{25}$$

These bounds give bounds for the eigenvalues, diagonal and total variance of P

$$\lambda_i(\mathcal{L}\mathsf{P}_0^-) \leq \lambda_i(\mathsf{P}) \leq \lambda_i(\mathcal{L}\mathsf{P}_0^+) , \tag{26}$$

$$\mathrm{diag}(\mathcal{L}\mathsf{P}_0^-) \leq \mathrm{diag}(\mathsf{P}) \leq \mathrm{diag}(\mathcal{L}\mathsf{P}_0^+) , \tag{27}$$

$$\mathrm{tr}\, \mathcal{L}\mathsf{P}_0^- \leq \mathrm{tr}\, \mathsf{P} \leq \mathrm{tr}\, \mathcal{L}\mathsf{P}_0^+ . \tag{28}$$

Using the upper and lower bounds with $\mathsf{P}_0^- = \lambda_n(\mathsf{P}_0)\mathsf{I}$ and $\mathsf{P}_0^- = \lambda_1(\mathsf{P}_0)\mathsf{I}$ gives bounds that depend on the bound matrix:

$$\lambda_n(\mathsf{P}_0)\mathsf{B} \leq \mathsf{P} \leq \lambda_1(\mathsf{P}_0)\mathsf{B} . \tag{29}$$

When the bounds in (29) are tight and P has a well-separated set of leading eigenvalues, the leading eigenvectors of P and B span approximately the same subspaces (Golub and Van Loan, 1996, Theorem 7.2.4). However, if P does not have a well-separated set of leading eigenvalues, $\mathsf{P} - \lambda_1(\mathsf{P}_0)\mathsf{B}$ may be small without the leading eigenvectors of P and B spanning approximately the same subspaces.

4 Analysis of the DALE

The preceding analysis can be used to determine when the solution of the DALE (9) admits a low-rank representation. The condition number of the DALE operator requires calculating the norm of $\mathcal{L}_\mathsf{A}$ and $\mathcal{L}_\mathsf{A}^{-1}$. For the normal DALE operator, $\|\mathcal{L}_\mathsf{A}\|_p = \lambda_1(\mathcal{L}_\mathsf{A}) = (1 - |\lambda_1(\mathsf{A})|^2)^{-1}$ and (Tippett et al., 2000a).

$$1 - |\lambda_n(\mathsf{A})|^2 \leq \|\mathcal{L}_\mathsf{A}^{-1}\|_p \leq 1 + |\lambda_1(\mathsf{A})|^2 , \tag{30}$$

[2]For two symmetric matrices X and Y, $\mathsf{X} \leq \mathsf{Y}$ means that $\mathsf{Y} - \mathsf{X}$ is positive semidefinite.

showing that $\mathcal{L}_{\mathsf{A}}$ is ill-conditioned when A has some but not all of its eigenvalues near the unit circle.

In the case of nonnormal dynamics A, $\|\mathcal{L}_{\mathsf{A}}^{-1}\|_p$ can be estimated in terms of the singular values of A,

$$\max_j |1-\sigma_i^2(\mathsf{A})| \leq \|\mathcal{L}_{\mathsf{A}}^{-1}\|_p \leq 1+\sigma_1^2(\mathsf{A}) \tag{31}$$

and $\|\mathcal{L}_{\mathsf{A}}\|_p$ can be bounded by

$$\frac{1}{2r(\mathsf{A})+r^2(\mathsf{A})} \leq \|\mathcal{L}_{\mathsf{A}}\|_p \leq \frac{1}{r(\mathsf{A})^2} \tag{32}$$

where the *radius of stability* $r(\mathsf{A})$ is the distance from A to the closest unstable matrix (Mori, 1990; Tippett et al., 2000a)[3]. These estimates show that the DALE operator $\mathcal{L}_{\mathsf{A}}$ is ill-conditioned when A is nearly unstable and the A has at least one singular value that is not near unity.

An example of stable nonnormal dynamics that produces a highly ill-conditioned DALE is the dynamics coming from the generalized one-dimensional advection equation

$$\mu_t + a\mu_x + c'(x)\mu = 0\,, \qquad 0 \leq x \leq 1 \tag{33}$$

with $a > 0$ and initial and boundary conditions

$$\mu(x, t=0) = \mu_0(x), \quad \mu(x=0, t) = 0\,, \tag{34}$$

respectively. Nonnormality is due to the undifferentiated term $c'(x)\mu$ and the boundary conditions. A similar model with periodic boundary conditions is described in detail in (Tippett et al., 2000b). Disturbances move from left to right with speed a. As disturbances pass through regions where $c'(x) < 0$ they grow and where $c'(x) > 0$ they decay. The zero boundary condition at the left boundary forces the solution to be identically zero after the time a^{-1} required to cross the domain. As a consequence, the operator A_τ that advances the solution τ time units is nilpotent and all its eigenvalues are identically zero; there is no modal growth.

[3]For normal stable matrices $r(\mathsf{A}) = 1 - |\lambda_1(\mathsf{A})|$, the distance from its largest eigenvalue to the unit circle. For nonnormal dynamics the radius of stability depends on the pseudospectrum of A (Trefethen, 1997).Eigenvalues near the unit circle, large singular values and sensitive eigenvalues cause the radius of stability to be small.

The discrete-time dynamics A_τ that advances the solution τ time units is given by (see Appendix)

$$\mathsf{A}_\tau \mu(x,t) = \begin{cases} 0 & 0 \le x \le a\tau \\ s(x)\mu(x-a\tau,t) & a\tau < x \le 1\,, \end{cases} \tag{35}$$

where $s(x) = \exp\left[(c(x-a\tau) - c(x))/a\right]$; note that $s(x) = \exp -\tau c'(x)$ in the limit $a\tau \ll 1$. In this example we take $a = 1/12$, $\tau = 1$ and $c(x)$ given by

$$c(x) = 1 + \frac{1}{16}\frac{2}{\pi}\arctan(16(0.5-x))\,. \tag{36}$$

In most parts of the domain $c'(x) \approx 0$ so that there is little growth. Around $x = 0.5$, $c'(x) < 0$ and there is amplification as shown in the plots of $c(x)$ and $s(x)$ in Fig. 1.

Nonmodal transient growth is found from the singular values and singular vectors of A_τ. The singular values and left singular vectors of A_τ are the square roots of the eigenvalues and the eigenvectors of $\mathsf{A}_\tau\mathsf{A}_\tau^T$ (see Appendix)

$$\mathsf{A}\mathsf{A}^T \mu(x,t) = \begin{cases} 0 & 0 \le x \le a\tau \\ s^2(x)\mu(x,t) & a\tau < x \le 1\,. \end{cases} \tag{37}$$

The singular values of A are zero and the values taken on by $s(x)$. The maximum growth in one time step is $\sigma_1(\mathsf{A}) = 1.37$ is sensitive to $c'(x)$. Where $s(x) > 1$ there is local amplification. In the spatially discrete case, the singular vectors of A are columns of the identity matrix. The leading singular vectors are associated with the region where $s(x)$ has its maximum.

We add mean-zero Gaussian distributed random noise b_k at each time-step:

$$\epsilon_{k+1} = \mathsf{A}_\tau \epsilon_k + b_k\,, \quad \left\langle bb^T \right\rangle = \mathsf{P}_0 \tag{38}$$

with covariance $\mathsf{P}_0 = \mathsf{I} + \mathsf{G}$ where G is a Gaussian correlation model with correlation length 0.25 and normalized so that $\operatorname{tr}\mathsf{G} = 1$. A sense of the time behavior of the system is seen by looking at the mean with respect to x of the forcing b_k and of the response y_k in Fig. 2 which show that the dynamics amplify the forcing and increase the time coherence.

The maximum possible amplification is found from the diagonal matrices $\mathsf{B} = \mathcal{L}\mathsf{I}$ and $\mathsf{B}_T = \mathcal{L}_{\mathsf{A}^T}\mathsf{I}$ to give that $\|\mathcal{L}_\mathsf{A}\|_\infty = 1209.6$ and $\|\mathcal{L}_\mathsf{A}\|_1 = 1157.6$. The norm of $\mathcal{L}_\mathsf{A}$ is large because though the eigenvalues of A are far from the unit circle, they are sensitive to perturbations. The maximum of B is within $a\tau$ of the right boundary and the maximum of B_T occurs within $a\tau$ of the left boundary, in contrast to the leading singular vectors which depending on $c(x)$ can be located anywhere in the domain (see Appendix). Since left singular vectors will always be to the left of their corresponding right singular vectors, left (right) singular vectors will always be a better approximation for B (B_T) than right (left). The norm of $\mathcal{L}^{-1}$ is bounded by $1 \leq \|\mathcal{L}_{\mathsf{A}}^{-1}\|_p \leq 2.88$ from (31). So $\mathcal{L}_{\mathsf{A}}$ is ill-conditioned and we expect the steady-state covariance P to have a low-rank approximation.

We calculate P using exact dynamics and a spatial descretization with 48 grid points. Figure 3a shows eigenvalues $\lambda_i(\mathsf{P})$ of P along with the upper and lower bounds obtained from the bound matrix B and (29). Much of the variance of P is contained in a few modes as suggested by the ill-conditioning of $\mathcal{L}_{\mathsf{A}}$. The spread in the bounds is due to the spectrum of P_0 not being flat. Figure 3b shows the diagonal of P and bounds obtained from the bound matrix and (29).

The relation between k and the fraction of variance explained by $\mathsf{P}^{(k)}$ is seen in Fig. 4; about 10 modes explain half of the variance. Also shown is the fraction of variance explained when P is projected onto the eigenvectors of B, the left singular vectors of A and the right singular vectors of A. The eigenvectors of B efficiently explain the variance of P while the left singular vectors of A do not do as well but are much better than the right singular vectors.

5 Conclusions

Ensemble and reduced-state approaches to prediction and data assimilation have shown low-rank covariance representations to be useful covariance models. How appropriate such approximation are in a given problem depends on the spectrum of the full covariance, which is often not available. We have examined the situation when the covariance P is a linear transform $\mathcal{L}$ of a forcing covariance P_0.

The singular value decomposition of the transformation $\mathcal{L}$ only gives useful information in special cases. More useful information is obtained from operator norms of the transformation. Since the transformation is a positive map, mapping covariance matrices to covariance matrices, there are simple expressions for its norm. Ill-conditioning is a necessary condition for the covariance matrix to admit a low-rank approximation. We show that the covariance P can be ill-conditioned only when the forcing covariance P_0 or the transformation $\mathcal{L}$ are ill-conditioned. A similar result shows that the fraction of variance explained by the first eigenmode of the covariance can be large only when the same is true of the forcing covariance or when $\mathcal{L}$ is ill-conditioned.

The numerical cost of calculating these norms can be comparable to calculating the full covariance, though the issue of poorly known error sources is avoided. Lanczos methods can be used if it is possible to apply $\mathcal{L}\mathsf{P}_0$ to a vector. Theoretical insight can also be gained as demonstrated in the case of the discrete-time algebraic Lypunov equation where the usefulness of low-rank representation is directly related to stability properties of the dynamics.

Appendix

The operator A_τ that advances the solution of (33) in time is easily calculated by making a linear change of variables such that in the new variables (33) is the usual advection equation. The new dependent variable $\nu(x,t) \equiv \mathsf{L}(x)\mu(x,t)$ where $\mathsf{L}(x) \equiv \exp\left(c(x)/a\right)$ satisfies

$$\nu_t + a\nu_x = 0\,, \tag{A1}$$

with initial and boundary conditions

$$\nu(x, t=0) = \mathsf{L}(x)\mu_0(x), \quad \nu(x=0,t) = 0\,. \tag{A2}$$

The operator $\tilde{\mathsf{A}}_\tau$ that advances (A1) in time is $\tilde{\mathsf{A}}_\tau \nu(x,t) = \nu(x - a\tau, t)$ and is related to A_τ by $\mathsf{A}_\tau = (\mathsf{L}(x))^{-1}\tilde{\mathsf{A}}_\tau \mathsf{L}(x)$.Therefore

$$\begin{aligned}
\mathsf{A}_\tau \mu(x,t) &= (\mathsf{L}(x))^{-1}\tilde{\mathsf{A}}_\tau \mathsf{L}(x)\mu(x,t) \\
&= \mathsf{L}(x - a\tau)(\mathsf{L}(x))^{-1}\mu(x - a\tau, t) \\
&= \exp\left(c(x - a\tau) - c(x)/a\right)\, \mu(x - a\tau, t)\,.
\end{aligned} \tag{A3}$$

The singular values and vectors of A_τ are the eigenvalues and eigenvectors of $\mathsf{A}_\tau \mathsf{A}_\tau^T$.

$$\begin{aligned}
\mathsf{A}_\tau \mathsf{A}_\tau^T \mu(x,t) &= (\mathsf{L}(x))^{-1}\tilde{\mathsf{A}}(\mathsf{L}(x))^2\tilde{\mathsf{A}}^T(\mathsf{L}(x))^{-1}\mu(x,t) \\
&= (\mathsf{L}(x))^{-1}\tilde{\mathsf{A}}(\mathsf{L}(x))^2(\mathsf{L}(x+a\tau))^{-1}\mu(x+a\tau,t) \\
&= (\mathsf{L}(x))^{-2}(\mathsf{L}(x-a\tau))^2\mu(x,t) \\
&= \exp((2c(x-a\tau)-2c(x))/a)\mu(x,t)
\end{aligned} \tag{A4}$$

except for $0 \le x \le a\tau$ where $\mathsf{A}_\tau \mathsf{A}_\tau^T$ is identically zero. $\mathsf{A}_\tau^T \mathsf{A}_\tau$ is a multiplication operator. Its point spectrum contains zero and its continuous spectrum contains the range of $s^2(x)$ for $0 \le x \le 1$ (Halmos, 1967, Problem 66). For an eigenvalue λ in the continuous spectrum there is no eigenfunction but there is a sequence of approximate eigenfunctions f_j such that $\|\mathcal{A}f_j - \lambda f_j\| \to 0$. Therefore, for non-constant $c(x)$ the operator $\mathsf{A}_\tau \mathsf{A}_\tau^T$ has no eigenfunctions but does have approximate eigenfunctions, for example those given by the eigenvectors of a spatially discrete approximation. For instance, suppose the n spatial grid points are $x_i = ia\tau/n$ for $0 \le i \le n-1$. Then, the left singular vector associated with the singular value $s(x_i)$ is the i-the column of the identity matrix and the right singular vector is the $i-1$-st column of the identity matrix.

The bound matrix B satisfies $\mathsf{B} = \mathsf{A}\mathsf{B}\mathsf{A}^T + \mathsf{I}$ and is found by solving

$$\tilde{\mathsf{P}} = \tilde{\mathsf{A}}\tilde{\mathsf{P}}\tilde{\mathsf{A}}^T + (\mathsf{L}(x))^2 \tag{A5}$$

where $\mathsf{B} = \mathsf{L}(x)^{-1}\tilde{\mathsf{P}}\mathsf{L}(x)^{-1}$. $\tilde{\mathsf{P}}$ is diagonal and given by

$$\tilde{\mathsf{P}}(x) = \sum_{k=0}^{\mathrm{int}(x/a\tau)} (\mathsf{L}(x-ka\tau))^{-2}\,, \tag{A6}$$

as can be verified using the relation

$$\begin{aligned}
\tilde{\mathsf{A}}\tilde{\mathsf{P}}(x)\tilde{\mathsf{A}}^T = \tilde{\mathsf{P}}(x-a\tau) &= \sum_{k=1}^{\mathrm{int}(x/a\tau)} (\mathsf{L}(x-ka\tau))^{-2} \\
&= \tilde{\mathsf{P}}(x) - (\mathsf{L}(x))^{-2}\,.
\end{aligned} \tag{A7}$$

Therefore,

$$
\begin{aligned}
\mathsf{B}(x) &= (\mathsf{L}(x))^{-2} \sum_{k=0}^{\operatorname{int}(x/a\tau)} (\mathsf{L}(x-ka\tau))^2 \\
&= \exp -2a^{-1}c(x) \sum_{k=0}^{\operatorname{int}(x/a\tau)} \exp 2a^{-1}c(x-ka\tau) \\
&= \sum_{k=0}^{\operatorname{int}(x/a\tau)} \exp -2a^{-1}(c(x)-c(x-ka\tau))
\end{aligned} \tag{A8}
$$

Similarly, the solution of $\mathsf{B}_T = \mathsf{A}^\dagger \mathsf{B}_T \mathsf{A} + \mathsf{I}$ is given by

$$
\mathsf{B}(x) = \sum_{k=0}^{\operatorname{int}(x/a\tau)} \exp 2a^{-1}(c(x)-c(x+ka\tau)) \tag{A9}
$$

The relation

$$
\begin{aligned}
\mathsf{B}(x-a\tau) &= \sum_{k=1}^{\operatorname{int}(x/a\tau)} \exp -2a^{-1}(c(x-a\tau)-c(x-ka\tau)) \\
&= \mathsf{B}(x) \exp -2a^{-1}(c(x-a\tau)-c(x)) \\
&- \exp -2a^{-1}(c(x)-c(x-a\tau))
\end{aligned} \tag{A10}
$$

shows that $\mathsf{B}(x) > \mathsf{B}(x-a\tau)$ if $c(x)$ is decreasing. As a consequence the maxima of B and B_T occur within $a\tau$ of 1 and of 0 respectively.

Acknowledgements. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grants 91.0029/95-4, 381737/97-7 and 30.0204/83-3, Financiadora de Estudos e Projetos (FINEP) Grant 77.97.0315.00, and the NASA EOS Interdisciplinary Project on Data Assimilation.

References

Barkmeijer, J., Gijzen, M. V., and Bouttier, F., Singular vectors and estimates of the analysis-error covariance matrix, *Q. J. R. Meteorol. Soc.*, *124*, 1695–1713, 1998.

Bhatia, R., A Note on the Lyapunov Equation, *Linear Algebra Appl.*, *259*, 71–76, 1997.

Byers, R. and Nash, S., On the singular "vectors" of the Lyapunov operator, *SIAM J. Alg. Disc. Math*, *8*, 59–66, 1987.

Cane, M. A., Kaplan, A., Miller, R., Tang, B., Hackert, E., and Busalacchi, A. J., Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter, *J. Geophys. Res.*, *101*, 22,599–22,617, 1996.

Changnon, D., Creech, T., and Marsili, N., Interactions with a weather-sensitive decision maker: A case study incorporating ENSO information into a strategy for purchasing natural gas, *Bull. Amer. Meteor. Soc.*, *80*, 1117–1126, 1999.

Cohn, S. E., An introduction to estimation theory, *J. Meteor. Soc. Japan*, *75*, 257–288, 1997.

Cohn, S. E. and Todling, R., Approximate data assimilation schemes for stable and unstable dynamics, *J. Meteor. Soc. Japan*, *74*, 63–75, 1996.

Dee, D. P. and da Silva, A. M., Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology., *Mon. Wea. Rev.*, 1998.

Farrell, B. F. and Ioannou, P. J., Generalized stability theory. Part I: Autonomous operators, *J. Atmos. Sci.*, *53*, 2025–2040, 1996.

Gahinet, P. M., Laub, A. J., Kenney, C. S., and Hewer, G. A., Sensitivity of the stable discrete-time Lyapunov equation, *IEEE Trans. Automat. Contr.*, *35*, 1209–1217, 1990.

Ghavimi, A. R. and Laub, A. J., Computation of approximate null vectors of Sylvester and Lyapunov operators, *IEEE Trans. Automat. Contr.*, *40*, 387–391, 1995.

Golub, G. H. and Van Loan, C. F., *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Third edn., 694 pp., 1996.

Halmos, P. R., *A Hilbert Space Problem Book*, Van Nostrand-Reinhold, New York, 365 pp., 1967.

Hewer, G. and Kenney, C., The sensitivity of the stable Lyapunov equation, *SIAM J. Control and Optimization*, *26*, 321–344, 1988.

Houtekamer, P. L. and Mitchell, H. L., Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Wea. Rev.*, *126*, 796–811, 1998.

Kenney, C. and Hewer, G., The senitivity of the algebraic and differential Riccati equations, *SIAM J. Control Optim.*, *28*, 50–69, 1990.

Kenney, C. and Hewer, G., Trace norm bounds for stable Lyapunov operators, *Linear Algebra Appl.*, *221*, 1–18, 1995.

Kleeman, R. and Moore, A. M., A theory for the limitation of ENSO predictability due to stochastic atmospheric transients, *J. Atmos. Sci.*, *54*, 753–767, 1997.

Kwon, W. H., Moon, Y. S., and Ahn, S. C., Bounds in algebraic Riccati and Lyapunov equations: A survey and some new results, *Int. J. Control*, *64*, 377–389, 1996.

Lancaster, P., Explicit solutions of linear matrix equations, *SIAM Review*, *12*, 544–566, 1970.

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T., The ECMWF ensemble prediction system: Methodology and validation, *Q. J. R. Meteorol. Soc.*, *122*, 73–119, 1996.

Mori, T., On the relationship between the spectral radius and the stability radius of discrete systems, *IEEE Trans. Automat. Contr.*, *35*, 835, 1990.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Unden, P., Eyre, J., Hollingsworth, A., and Bouttier, F., The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions, *Q. J. R. Meteorol. Soc.*, *124*, 1809–1829, 1998.

Riishøjgaard, L. P., A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus*, *50A*, 42–57, 1998.

Tippett, M. K. and Marchesin, D., Bounds for solutions of the discrete algebraic Lyapunov equation, *IEEE Trans. Automat. Contr.*, *44*, 214–218, 1999a.

Tippett, M. K. and Marchesin, D., Upper bounds for the solution of the discrete algebraic Lyapunov equation, *Automatica*, *35*, 1485–1489, 1999b.

Tippett, M. K., Cohn, S. E., Todling, R., and Marchesin, D., Conditioning of the stable, discrete-time Lyapunov operator, *SIAM J. Matrix Anal. Appl.*, *22*, 56–65, 2000a.

Tippett, M. K., Cohn, S. E., Todling, R., and Marchesin, D., Low-dimensional representation of error covariance, *Tellus*, in press, 2000b.

Trefethen, L. N., Pseudospectra of linear operators, *SIAM Review*, *39*, 383 – 406, 1997.

Whitaker, J. S. and Sardeshmukh, P. D., A linear theory of extratropical synoptic eddy statistics, *J. Atmos. Sci.*, *55*, 237–258, 1998.

Figure Captions

Fig. 1. $c(x)$ (solid line) and $s(x)$ (dash-dot line)

Fig. 2. (a) Mean with respect to x of the (a) forcing b_k and (b) response ϵ_k plotted with respect to the advection time $T = k\tau a$.

Fig. 3. (a) Eigenvalues $\lambda_i(\mathsf{P})$ of P (solid line) and their upper and lower bounds (dotted lines) obtained using (26) with $\mathsf{P}_0^- = \lambda_n(\mathsf{P}_0)\mathsf{I}$ and $\mathsf{P}_0^- = \lambda_1(\mathsf{P}_0)\mathsf{I}$. (b) Diagonal of P (solid line) and its bounds (dotted lines) obtained using (27) with $\mathsf{P}_0^- = \lambda_n(\mathsf{P}_0)\mathsf{I}$ and $\mathsf{P}_0^- = \lambda_1(\mathsf{P}_0)\mathsf{I}$

Fig. 4. Fraction of the variance explained by eigenvectors of P (solid line) eigenvectors of B (dotted dashed line), left singular vectors of A (dashed line) and right singular vectors of A (dotted line).

Figures

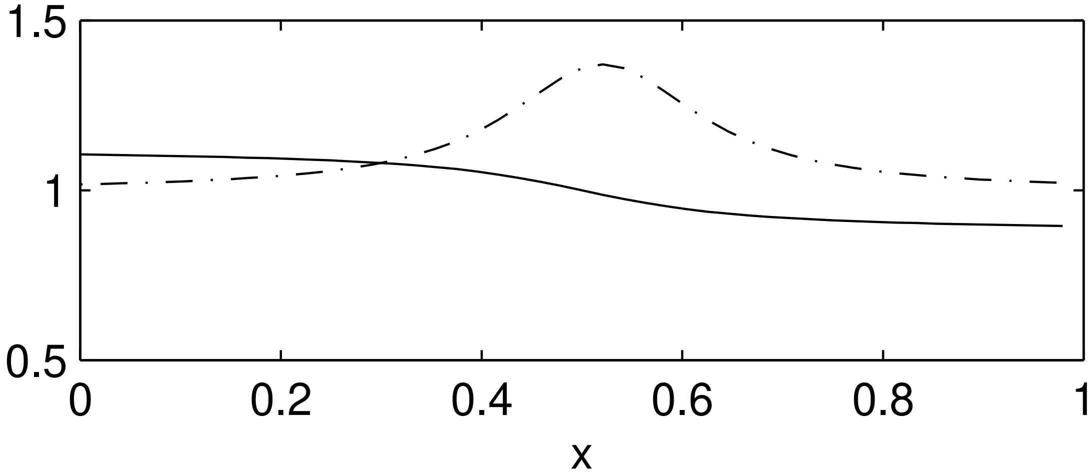

Fig. 1. $c(x)$ (solid line) and $s(x)$ (dash-dot line)

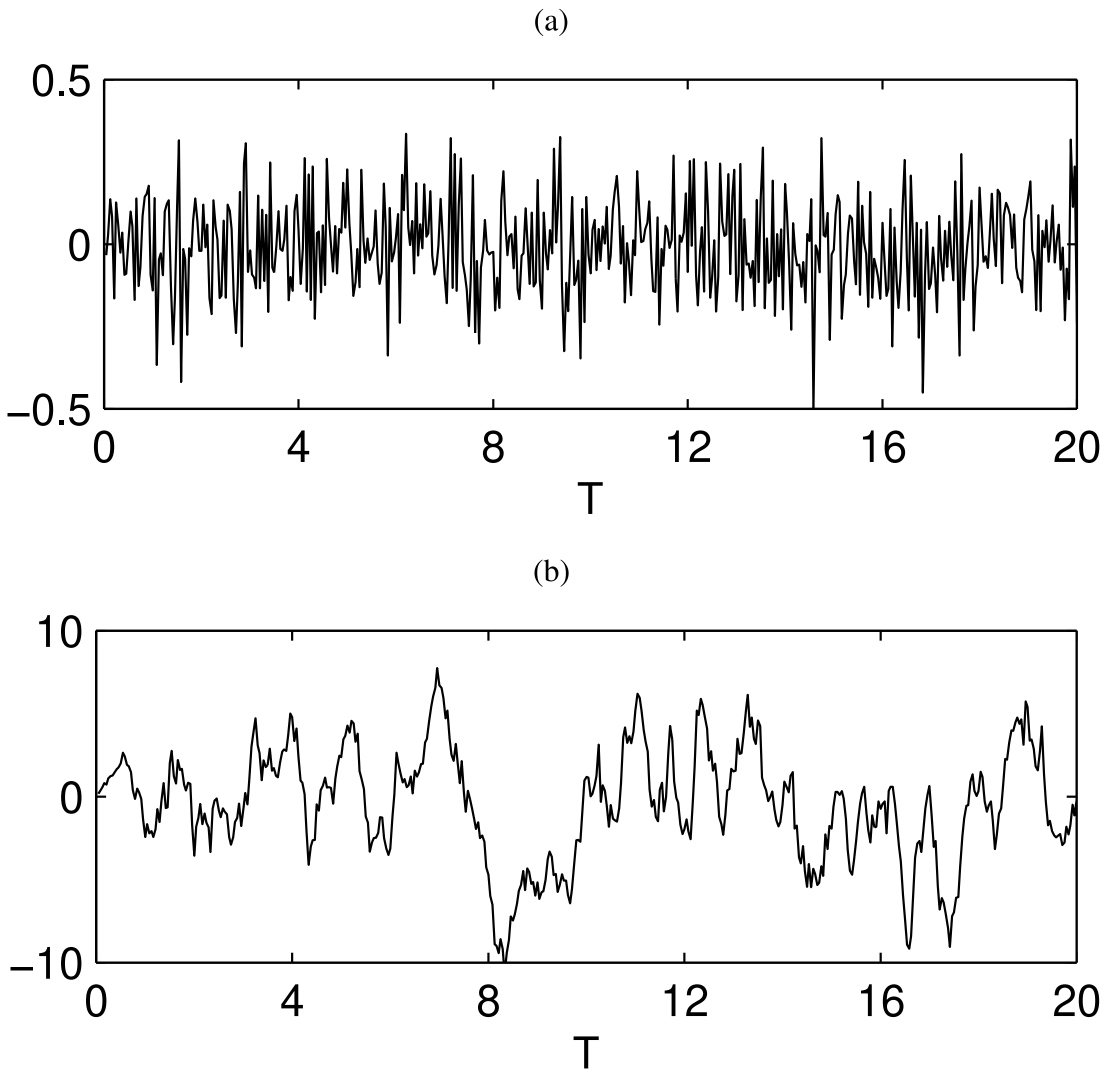

Fig. 2. (a) Mean with respect to x of the (a) forcing b_k and (b) response ϵ_k plotted with respect to the advection time $T = k\tau a$.

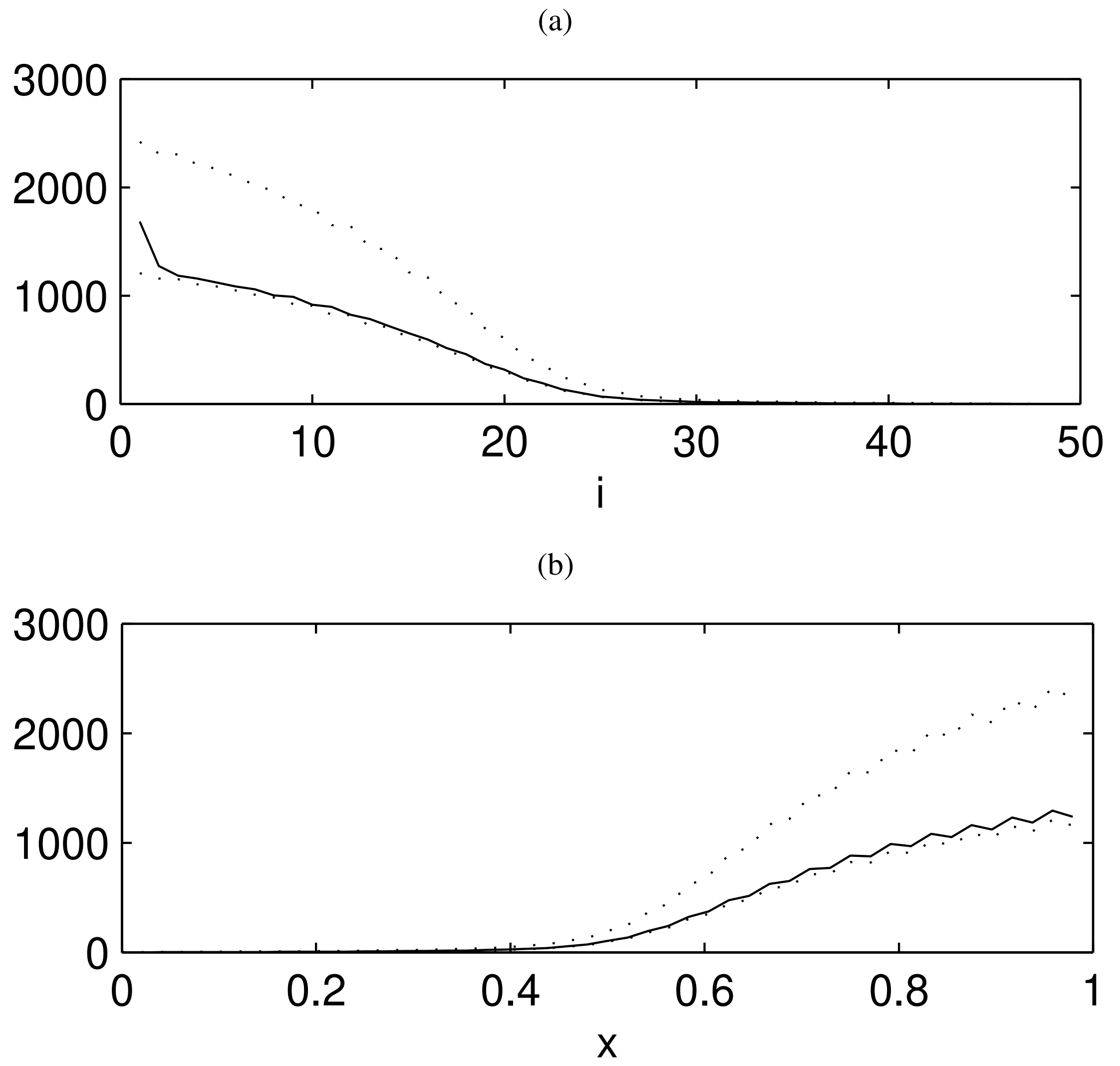

Fig. 3. (a) Eigenvalues $\lambda_i(\mathsf{P})$ of P (solid line) and their upper and lower bounds (dotted lines) obtained using (26) with $\mathsf{P}_0^- = \lambda_n(\mathsf{P}_0)\mathsf{I}$ and $\mathsf{P}_0^- = \lambda_1(\mathsf{P}_0)\mathsf{I}$. (b) Diagonal of P (solid line) and its bounds (dotted lines) obtained using (27) with $\mathsf{P}_0^- = \lambda_n(\mathsf{P}_0)\mathsf{I}$ and $\mathsf{P}_0^- = \lambda_1(\mathsf{P}_0)\mathsf{I}$

.

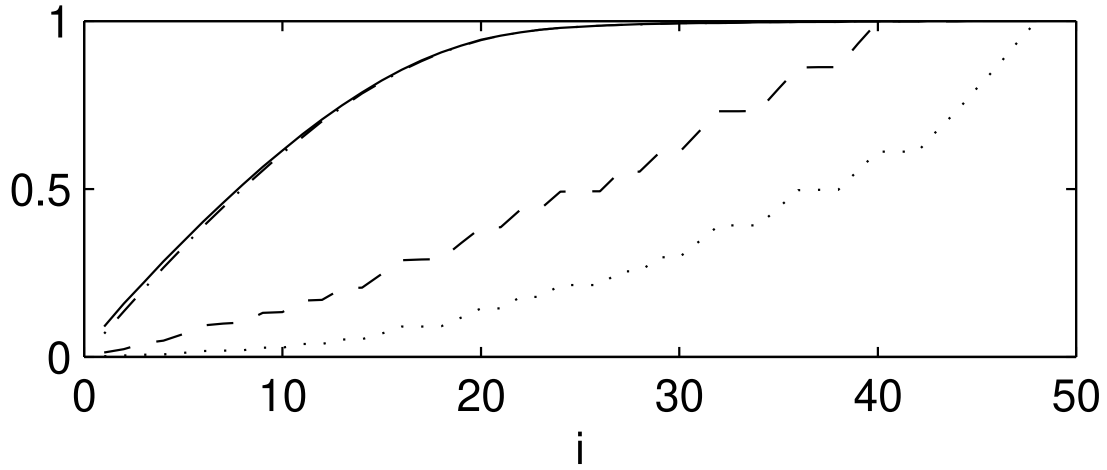

Fig. 4. Fraction of the variance explained by eigenvectors of P (solid line) eigenvectors of B (dotted dashed line), left singular vectors of A (dashed line) and right singular vectors of A (dotted line).