A doubly stochastic advection-diffusion-decay model for testing data assimilation methodologies

Michael Tsyrulnikov and Alexander Rakitko

(michael.tsyrulnikov@gmail.com)

August 24, 2017

Abstract

A doubly stochastic advection-diffusion-decay model defined on the 1D spatial domain (the circle) is proposed. The model is intended to be used in testing and developing data assimilation methodologies as well as in a broader context of non-stationary spatio-temporal random field modeling. The model is hierarchical: it is a stochastic (forced by the white noise) partial differential equation whose coefficients are spatio-temporal random fields by themselves satisfying their own stochastic partial differential equations with constant coefficients. The solution to the model is a random field that has complex spatio-temporal covariances with the tunable degree of non-stationarity. Another important feature of the model is that it allows the estimation of "true" space and time specific spatio-temporal covariances: both signal covariances and error covariances for any filter in question. This provides the researcher not only with the true state, as in the Observing Systems Simulation Experiments (OSSE) methodology, but also with the true error "covariances of the day". The capabilities of the doubly stochastic model are illustrated in numerical experiments with the Ensemble Kalman Filter (EnKF). It is shown that the accuracy of the forecast-ensemble sample covariance matrix can be substantially improved by its temporal smoothing.

Keywords: *Non-stationary spatio-temporal random field, Hierarchical modeling, Advection-diffusion-decay model, EnKF, Data assimilation.*

Running head: A doubly stochastic advection-diffusion-decay model

1 Introduction

In-depth analysis of existing and emerging data assimilation methodologies is greatly facilitated if a "system–model" pair is available. This allows the researcher to compute the "synthetic truth", which is then used to create "synthetic" observations (with exactly known observation operator and error statistics) and to compute analysis and forecast

errors. This is the OSSE methodology. With this methodology, knowing the forecast and analysis errors, one can reliably estimate their biases and spatial covariances by temporal and, possibly, spatial averaging (Errico and Privé, 2014). These statistics are relevant for variational data assimilation schemes, in which the background error covariances are *static*. State-of-the-art ensemble data assimilation schemes, however, produce and utilize estimates of the "statistics of the day", which cannot be verified with this approach. To validate and thoroughly examine EnKF and other ensemble (Monte Carlo) based data assimilation techniques, not only the "truth" itself but also the true time and space specific error statistics would be helpful. Knowing the true error statistics provides a very valuable opportunity to examine and, hopefully, to improve the filter. In this paper, we present a model on a one-dimensional spatial domain that allows the estimation of the true non-stationary in space and time error covariances.

Bishop and Satterfield (2013) were the first who assessed the "true" EnKF's background error variances in a setting with a deterministic system (model of truth) and stochastic forecast model. Their idea was to run multiple assimilation replicates (in an OSSE setting) and estimate pointwise time and space specific error variances by averaging over those replicates. We adopt this idea but argue that in Bayesian filtering (which is the most widely used approach to data assimilation, e.g. Jazwinski (1970); Särkkä (2013)), the system state is assumed to be random whilst the forecast models is deterministic. Therefore, to be consistent with the Bayesian data assimilation paradigm, the truth should be governed by a stochastic equation. For this reason, Tsyrulnikov and Rakitko (2017) set up a system in which the scalar (one-variable) truth was governed by a *doubly stochastic* model (e.g. Tjøstheim, 1986), whereas the forecast model was deterministic. More specifically, the true state ξ_k (where k labels the time instant) satisfied the equation

$$\xi_k = F_k \xi_{k-1} + \sigma_k \varepsilon_k, \tag{1}$$

in which not only the white noise forcing $\varepsilon_k \sim \mathcal{N}(0,1)$ was random, the coefficients F_k and σ_k were random sequences by themselves, each satisfying its own linear stochastic model similar to Eq.(1) but with constant coefficients. In this formulation, *conditionally on the coefficient processes* F_k and σ_k (that is, after their realizations are computed and kept fixed), the forecast model $\xi_k = F_k \xi_{k-1}$ becomes deterministic whilst the truth is, obviously, a non-stationary random process.

The idea of introducing variability in coefficients of a stochastic differential equation as a means of achieving non-stationary behavior of its solution is not new, for references see section 6 below. A novelty in the model Eq.(1) is in specifying stochastic models for the coefficient processes with the intention to achieve conditional non-stationarity in the distribution of the process ξ_k. Parameters of the models for the coefficient processes F_k and σ_k (i.e. hyperparameters) define the degree of the non-stationarity and its structure (say, a large variance of the process σ_k implies a large variability in $\mathsf{Var}\,\xi_k$, whereas a large variance of the process F_k implies a high variability in both the variance and the temporal length scale of the process ξ_k). We note that only under non-stationarity in time

and space, ensemble based data assimilation schemes can outperform variational schemes. Therefore, a thorough examination of an ensemble data assimilation technique requires a model of truth that produces non-stationary spatio-temporal fields (processes). In this research we extend the model Eq.(1) to a spatio-temporal random field on the circle by replacing F_k with a sum of advection, diffusion, and decay operators whose coefficients are spatio-temporal random fields by themselves. The result is the new doubly stochastic advection-diffusion-decay model (DSADM).

Besides the non-stationarity, the scalar model Eq.(1) allows, importantly, the estimation of the true time-specific variances: both of the signal (i.e. ξ_k itself) and of errors of any filter. The estimation can be made as accurate as needed and is achieved by averaging over a number of truths (and random observation errors) while keeping the sequences F_k and σ_k fixed. This feature of the doubly stochastic model Eq.(1) was exploited by Tsyrulnikov and Rakitko (2017) in the development of their Hierarchical Bayes Ensemble Filter. Here, we set up the DSADM to have realistic spatial and temporal scales, run an OSSE with the stochastic EnKF, and study time and space specific *errors* in different candidate background error *covariances*, including ensemble sample covariances, time mean covariances, time smoothed covariances, and some of their hybrids.

The model we propose here (DSADM) belongs to the class of so-called "toy models" intended, on the one hand, to serve as quick testbeds, and on the other hand, to facilitate thorough exploration of fundamental features of data assimilation and related techniques. A number of evolutionary toy models of increasing complexity has been proposed. The simplest models have just a few variables, like the logistic map (used by Mitchell and Houtekamer, 2009), the Ikeda model (used by Hansen and Smith, 2001), or the three-variable Lorenz'63 model (proposed by Lorenz, 1963). Models defined on a 1D spatial domain include, among others, the linear advection-diffusion equation (used by Daley and Ménard, 1993), the viscous Burgers equation (used by Apte et al., 2010), and the most popular Lorenz'96 model (proposed by Lorenz, 2005). Most of these models are nonlinear. The model we propose in this paper is linear but it can exhibit intermittent instability and allows studying true signal and filtering error covariances (and, more generally, probability distributions) in flow regimes with significant (and highly tunable) spatio-temporal non-stationarity.

The rest of the paper is organized as follows. We start with examining the stationary spatio-temporal statistics of a stochastic advection-diffusion-decay model with constant coefficients. Then, we introduce the doubly stochastic advection-diffusion-decay model, in which the model parameters become random fields by themselves governed by their own stochastic models with constant coefficients. After that, we study the model's behavior in numerical experiments. Finally, we show that a simple temporal smoothing of EnKF's background error covariances leads to a significant improvement in their accuracy. The R code of the model and the R scripts that produced this paper's numerical results are available from https://github.com/rakitko/NoStRa.

2 Stochastic advection-diffusion-decay model with constant coefficients

In this section, we examine spatio-temporal statistics of solutions to a stochastic advection-diffusion-decay model with non-stochastic and constant coefficients (parameters).

2.1 Model

The model is the following stochastic partial differential equation (Whittle (1986, Ch. 20, Sec. 3), Lindgren et al. (2011, Eq.(17)), Sigrist et al. (2015)):

$$\frac{\partial \xi}{\partial t} + U\frac{\partial \xi}{\partial s} + \rho\xi - \nu\frac{\partial^2 \xi}{\partial s^2} = \sigma\alpha, \tag{2}$$

where t is time, s is the spatial coordinate on the circle $\mathbb{S}^1(R)$ of radius R, U is the advection velocity, ρ is the decay (damping) parameter, ν is the diffusion parameter, $\alpha(t,s)$ is the standard white in time and space noise, and σ is the intensity of the forcing. The four parameters $\boldsymbol{\theta} = (U, \rho, \nu, \sigma)$ are constant in space and time.

2.2 Stationary spatio-temporal statistics

We start with rewriting Eq.(2) using the material time derivative (i.e. along the characteristic $s = s_0 + Ut$), or, equivalently, switching to the Lagrangian frame of reference by making the change of variables $(t, s) \mapsto (t, s - Ut)$:

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} + \rho\xi - \nu\frac{\partial^2 \xi}{\partial s^2} = \sigma\alpha. \tag{3}$$

Next, we employ the spectral expansion in space,

$$\xi(t,s) = \sum_{m=-n/2}^{n/2} \tilde{\xi}_m(t)\, \mathrm{e}^{\mathsf{i}ms/R} \tag{4}$$

and

$$\alpha(t,s) = \sum_{m=-n/2}^{n/2} \tilde{\alpha}_m(t)\, \mathrm{e}^{\mathsf{i}ms/R}, \tag{5}$$

where n denotes the size of the (uniform) grid on the circle, i is the imaginary unit, and $\tilde{\xi}_m(t)$ and $\tilde{\alpha}_m(t)$ are the (complex) spectral coefficients. It can be shown (e.g. Tsyrulnikov and Gayfulin, 2016, Appendix A.4) that $\tilde{\alpha}_m(t)$ are independent Gaussian complex standard white noise processes $\omega_m(t)$ with the common intensity $a = 1/\sqrt{2\pi R}$:

$$\tilde{\alpha}_m(t) = a\,\omega_m(t). \tag{6}$$

Now, we substitute Eqs.(4)–(6) into Eq.(3), getting

$$\frac{\mathrm{d}\tilde{\xi}_m}{\mathrm{d}t} + (\rho + \frac{\nu}{R^2}m^2)\tilde{\xi}_m = a\sigma\omega_m(t). \tag{7}$$

This is the spectral space form of the model Eq.(3). It is easily seen that with $\rho > 0$ and $\nu > 0$, the solutions to Eq.(7) for different m become, after an initial transient, mutually independent stationary zero-mean random processes. This implies that the physical-space solution $\xi(t, s)$ becomes a zero-mean stationary in time and space[1] random field. Each elementary stochastic process $\tilde{\xi}_m(t)$ is an Ornstein-Uhlenbeck process (e.g. Arnold, 1974, sec. 8.3) with the stationary covariance function

$$B_m(t) = b_m \cdot \mathrm{e}^{-|t|/\tau_m}, \tag{8}$$

where the spectral variances b_m are

$$b_m = \frac{a^2\sigma^2}{2} \cdot \frac{1}{\rho + \frac{\nu}{R^2}m^2} \tag{9}$$

and the spectral temporal length scales τ_m are

$$\tau_m = \frac{1}{\rho + \frac{\nu}{R^2}m^2}. \tag{10}$$

Note that Eq.(10) provides the motivation for the inclusion of the decay term in the model. Indeed, with $\rho = 0$, the time scale τ_0 would be infinitely large, which, on the finite sphere, is unphysical.

The stationary in space-time covariance function of $\xi(t, s)$ can be easily derived as $\mathbb{E}\,\xi(t_1, s_1)\,\xi(t_1 + t, s_1 + s)$ (where t_1 and s_1 are any admissible time and space coordinates, respectively) from Eq.(4) while utilizing independence of the spectral processes $\tilde{\xi}_m(t)$, Eq.(8), and returning to the Eulerian frame of reference:

$$B(t, s) = \sum_{m=-n/2}^{n/2} b_m \,\mathrm{e}^{-|t|/\tau_m}\,\mathrm{e}^{\mathrm{i}m(s-Ut)/R}. \tag{11}$$

This equation implies that the space-time correlations are *non-separable*, i.e. they cannot be represented as a product of purely spatial and purely temporal correlations. Moreover, according to Eq.(10), smaller spatial scales (i.e. larger wavenumbers m) correspond to smaller temporal scales τ_m. This feature of space-time correlations ("proportionality of scales") is physically reasonable—as opposed to the simplistic and unrealistic separability of space-time correlations—and widespread in the real world, see Tsyroulnikov (2001) and references therein.

Finally, from Eq.(11), the stationary (steady-state) variance of $\xi(t, s)$ is

$$\mathsf{Var}\,\xi \equiv (\mathrm{SD}(\xi))^2 = \sum_{m=-n/2}^{n/2} b_m = \frac{a^2\sigma^2}{2} \sum_{m=-n/2}^{n/2} \frac{1}{\rho + \frac{\nu}{R^2}m^2}, \tag{12}$$

where SD stands for the standard deviation.

[1] By definition, a zero-mean space-time random field (process) is (second-order) stationary (homogeneous) if its spatio-temporal covariances are invariant under translations: $\mathbb{E}\,\xi(t_1, s_1)\,\xi(t_2, s_2) = \mathbb{E}\,\xi(t_1 + u, s_1 + v)\,\xi(t_2 + u, s_2 + v)$.

2.3 Roles of model parameters

Firstly, we note that U does not impact the variance spectrum and the spectral temporal length scales of ξ (see Eqs.(9) and (10)), its role is just to rotate the solution with the constant angular velocity U/R.

Secondly, Eq.(9) implies that the *shape* of the spatial spectrum is

$$b_m \propto \frac{1}{\rho + \frac{\nu}{R^2} m^2} \propto \frac{1}{1 + (\frac{m}{m_0})^2}, \tag{13}$$

where $m_0 = R\sqrt{\rho/\nu}$ is the characteristic non-dimensional wavenumber, which defines the *width* of the spectrum and thus the spatial length scale. Therefore, the latter can be defined [2] as the inverse dimensional wavenumber m_0/R:

$$L = \sqrt{\frac{\nu}{\rho}}. \tag{14}$$

Thus, the ratio ν/ρ controls the spatial length scale L. In addition, ν/ρ impacts the temporal correlations. Indeed, a higher L implies a redistribution of the variance towards larger spatial scales (i.e. lower wavenumbers m). But as we noted, in the model Eq.(2), larger spatial scales correspond to larger temporal scales τ_m. As a result, a higher ν/ρ leads to a larger temporal length scale as well as the spatial length scale L.

Thirdly, using Eq.(14), we can rewrite Eq.(10) as

$$\tau_m = \frac{1}{\rho} \cdot \frac{1}{1 + (\frac{Lm}{R})^2} \tag{15}$$

This equation implies that with L being fixed, all spectral temporal length scales τ_m are inversely proportional to ρ, which, thus, determines the physical-space temporal Lagrangian length scale T of the spatio-temporal random field ξ. We define T as the macroscale (e.g. Yaglom, 1987, Eq.(2.88)) along the characteristic,

$$T = \frac{1}{2\mathsf{Var}\,\xi} \int_{-\infty}^{\infty} B(t, Ut)\mathrm{d}t = \frac{\sum b_m \tau_m}{\sum b_m} = \frac{1}{\rho} \frac{\sum [1 + (\frac{Lm}{R})^2]^{-2}}{\sum [1 + (\frac{Lm}{R})^2]^{-1}}, \tag{16}$$

where the second equality is due to Eq.(11), the third equality is due to Eqs.(9) and (15), and the summations are over m from $-n/2$ to $n/2$.

Technically, Eqs.(14), (16), and (12) allow us to compute the *internal* model parameters ρ, ν, and σ from the *externally* specified parameters L, T, and SD(ξ). Conceptually, we summarize the above conclusions as follows.

- U does not affect the Lagrangian spatio-temporal covariances. It tilts the Eulerian spatio-temporal correlations towards the direction $\mathrm{d}s = U\mathrm{d}t$ in space-time.
- The ratio ν/ρ determines the spatial length scale L and impacts the temporal length scale T.

[2] One can show that for a dense enough spatial grid, thus defined length scale L almost coincides with the macroscale Λ_ξ defined below in Eq.(27).

- With the ratio ν/ρ being fixed, ρ controls the temporal length scale T.
- The triple ρ, ν, σ determines the resulting process variance $\mathsf{Var}\,\xi$.

These relationships provide us with guidance about which *local* properties of the spatio-temporal statistics are going to be impacted if the parameters $\boldsymbol{\theta} = (U, \rho, \nu, \sigma)$ become variable in space and time.

3 Doubly stochastic advection-diffusion-decay model (DSADM)

Here, we allow the parameters $\boldsymbol{\theta} = (U, \rho, \nu, \sigma)$ of the model Eq.(2) to be spatio-temporal random fields by themselves each satisfying the model Eq.(2) but with constant coefficients. The resulting model becomes, thus, a three-level hierarchical model (Wikle et al., 1998; Banerjee et al., 2015). At the first level is the random field in question $\xi(t, s)$ modeled conditionally on second-level fields $\boldsymbol{\theta}(t, s)$. At the second level are the random fields $\boldsymbol{\theta}(t, s) = (U(t, s), \rho(t, s), \nu(t, s), \sigma(t, s))$ controlled by the hyperparameters $\boldsymbol{\gamma}$ (which are on the third level). So, to compute a realization of the pseudo-random field $\xi(t, s)$, we, first, specify the hyperparameters $\boldsymbol{\gamma}$. Then, we compute realizations of the second-level fields $\boldsymbol{\theta}(t, s)$ (which we will also refer to as secondary or parameter fields). Finally, we substitute the secondary fields for the respective parameters in Eq.(2) and solve the resulting spatio-temporal model for the primary field $\xi(t, s)$.

The idea behind this extension of the basic model Eq.(2) is the following. If the secondary fields $\boldsymbol{\theta}(t, s)$ vary *smoothly* in space and time, then, locally, in a vicinity of some point in space-time (t_0, s_0), the statistics of the field $\xi(t, s)$ will resemble that for the stationary model Eq.(2) with constant parameters equal to $\boldsymbol{\theta}(t, s)$ *frozen* at the point (t_0, s_0) (see also Lindgren et al., 2011, sec. 3.2). As the statistics of the model Eq.(2) with constant parameters do depend on the parameters $\boldsymbol{\theta}$ (see section 2), the resulting solution $\xi(t, s)$ to the model Eq.(2) with variable parameters becomes non-stationary in space-time, with the degree of the non-stationarity controlled by the variability in the secondary fields $\boldsymbol{\theta}(t, s)$.

3.1 First level of the hierarchy: the field in question ξ

At the first level, $\xi(t, s)$ satisfies the basic Eq.(2) with variable in space and time coefficients,

$$\frac{\partial \xi}{\partial t} + U(t, s)\,\frac{\partial \xi}{\partial s} + \rho(t, s)\,\xi - \nu(t, s)\,\frac{\partial^2 \xi}{\partial s^2} = \sigma(t, s)\,\alpha(t, s). \tag{17}$$

3.2 Second level of the hierarchy: the parameter fields $\boldsymbol{\theta}$

The secondary fields $\boldsymbol{\theta}(t, s)$, that is, the coefficients of the first-level stochastic partial differential equation, Eq.(17), are modeled by introducing Gaussian *perturbation fields*

(denoted in what follows by the asterisk) θ^* each satisfying its own stochastic advection-diffusion-decay model Eq.(2). A secondary field θ (i.e. U, ρ, ν, or σ) is controlled by the respective set of *hyperparameters*, $\boldsymbol{\gamma}_\theta$. This set includes the unperturbed values of $\boldsymbol{\theta}$ denoted in what follows by the overbar, $\bar{\boldsymbol{\theta}} = (\bar{U}, \bar{\rho}, \bar{\nu}, \bar{\sigma})$, and secondary-field-specific hyperparameters as described below.

3.2.1 $U(t,s)$

$$U(t,s) = \bar{U} + U^*(t,s), \tag{18}$$

where $U^*(t,s)$ is the perturbation field that satisfies our basic stochastic model Eq.(2):

$$\frac{\partial U^*}{\partial t} + U_U \frac{\partial U^*}{\partial s} + \rho_U U^* - \nu_U \frac{\partial^2 U^*}{\partial s^2} = \sigma_U \, \alpha_U(t,s), \tag{19}$$

$\boldsymbol{\gamma}_U = (\bar{U}, U_U, \rho_U, \nu_U, \sigma_U)$ are constant and non-random hyperparameters, and α_U is the independent from α white noise. Apparently, $U(t,s)$ is a Gaussian random field.

3.2.2 $\sigma(t,s)$

We require the field $\sigma(t,s)$ to be positive and having zero probability density at $\sigma = 0$. For this reason, $\sigma(t,s)$ is modeled as a transformed Gaussian process:

$$\sigma(t,s) = \bar{\sigma} \cdot g(\sigma^*(t,s)), \tag{20}$$

where $\sigma^*(t,s)$ is the Gaussian perturbation field and $g(z)$ is the transformation function selected to be the scaled and shifted logistic function:

$$g(z) := (1 + \mathrm{e}^b) \frac{\mathrm{e}^{z-b}}{1 + \mathrm{e}^{z-b}}, \tag{21}$$

where b is the constant. The function $g(z)$ has the following property: it behaves like the ordinary exponential function everywhere except for $z \gg b$. Indeed, it tends to zero as $z \to -\infty$ in the same way the exponential does. Moreover, like $\exp(z)$, it is equal to 1 at $z = 0$. With $b > 0$, $g(z)$ saturates for $z \gg 0$ at the level $1 + \mathrm{e}^b$; this is the major difference of g from the exponential function and the reason why we replace $\exp(z)$ by $g(z)$: to avoid too large values in $\sigma(t,s)$, which can give rise to unrealistically large spikes in ξ. For $b = 1$, the function $g(z)$ in plotted in Fig.1 alongside the exponential function.

The perturbation Gaussian random field $\sigma^*(t,s)$ is postulated to satisfy the stochastic model Eq.(2):

$$\frac{\partial \sigma^*}{\partial t} + U_\sigma \frac{\partial \sigma^*}{\partial s} + \rho_\sigma \sigma^* - \nu_\sigma \frac{\partial^2 \sigma^*}{\partial s^2} = \sigma_\sigma \, \alpha_\sigma(t,s). \tag{22}$$

where $\boldsymbol{\gamma}_\sigma = (\bar{\sigma}, U_\sigma, \rho_\sigma, \nu_\sigma, \sigma_\sigma, b)$ are the hyperparameters and α_σ is the independent white noise field. Due to the nonlinearity of the transformation function g, the field $\sigma(t,s)$ is non-Gaussian.

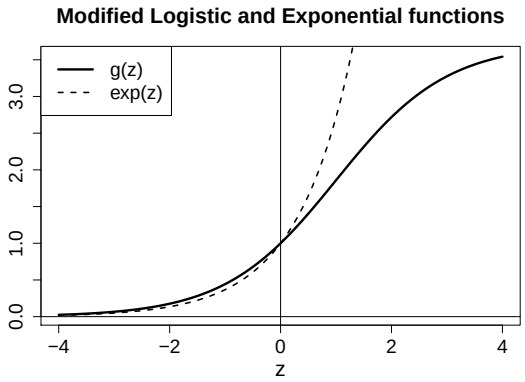

Figure 1: Modified logistic function with $b = 1$ and exponential function

3.2.3 $\rho(t, s)$ and $\nu(t, s)$

The remaining two secondary fields $\rho(t, s)$ and $\nu(t, s)$ (denoted here generically by ψ) are modeled as

$$\psi(t, s) = \bar{\psi} \cdot [(1 + \varepsilon_\psi) \cdot g(\psi^*(t, s)) - \varepsilon_\psi] , \tag{23}$$

where ε_ψ is the small positive constant, the parameter b in the definition of the function g is the same as above, and

$$\frac{\partial \psi^*}{\partial t} + U_\psi \frac{\partial \psi^*}{\partial s} + \rho_\psi \psi^* - \nu_\psi \frac{\partial^2 \psi^*}{\partial s^2} = \sigma_\psi \, \alpha_\psi(t, s). \tag{24}$$

Equation (24) is analogous to Eq.(22), whereas Eq.(23) differs from Eq.(20) in that we allow for small negative values of ψ (that is, for ρ and ν), note the subtraction of ε_ψ in Eq.(23). Allowing for negative values of ρ and ν is motivated by the desire to introduce an intermittent instability into the model.

Thus, for each of the fields $\psi = \rho, \nu$, the respective hyperparameters are $\boldsymbol{\gamma}_\psi = (\bar{\psi}, U_\psi, \rho_\psi, \nu_\psi, \sigma_\psi, \varepsilon_\psi)$. Like $\sigma(t, s)$, the fields $\rho(t, s)$ and $\nu(t, s)$ are non-Gaussian.

3.3 Third level of the hierarchy: hyperparameters $\boldsymbol{\gamma}$

In contrast to the Lorenz'96 model, which has only one parameter F, the DSADM has many parameters as discussed above. In total, there are 23 hyperparameters. It is convenient to specify the hyperparameters using a set of more sensible *external* parameters, see Appendix A.

3.4 Properties and capabilities of the DSADM

Here, we list features of the DSADM that can be useful for data assimilation applications. In passing, we also compare the DSADM with existing nonlinear toy models.

3.4.1 Non-stationary and knowable field (signal) covariances

It is worth stressing that the DSADM Eq.(17), like the scalar doubly stochastic model Eq.(1), is intended to be used as a model of truth with random but fixed (in an experiment) secondary fields. In other words, we consider the *conditional* probability distribution of the random field ξ given the fields $\boldsymbol{\theta}$. In this setting, the space-time covariances (and, more generally, the probability distribution) of $\xi(t,s)$ are, first, knowable, and second, non-stationary in space-time.

Indeed, given the fields $\boldsymbol{\theta}(t,s)$, Eq.(17) becomes a linear evolutionary state-space stochastic model,

$$\frac{d\boldsymbol{\xi}}{\mathrm{d}t} = \boldsymbol{\Phi}(t)\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \tag{25}$$

where the boldface $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}$ stand for the vectors of the spatially gridded fields $\xi(t,s)$ and $\sigma(t,s)\,\alpha(t,s)$, respectively, and $\boldsymbol{\Phi}(t)$ is the linear spatial operator dependent on the fields $\boldsymbol{\theta}(t,s)$. Discretizing Eq.(25) in time yields the equation for the signal $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}_k = \mathbf{F}_k\,\boldsymbol{\xi}_{k-1} + \boldsymbol{\varepsilon}_k \tag{26}$$

(where $k = 1, 2, \dots$). Now we note that knowing the secondary fields $\boldsymbol{\theta}$, we know the operator $\mathbf{F}_k$ and the probability distribution of the forcing $\boldsymbol{\varepsilon}_k$. This implies that, starting from a random zero-mean initial condition with known covariance matrix, we can compute the covariance matrix of $\boldsymbol{\xi}_k$ at any time instant k. So, the covariances of $\boldsymbol{\xi}_k$ are *knowable*.

The dependence of the operator $\mathbf{F}_k$ and the forcing $\boldsymbol{\varepsilon}_k$ on the random fields $\boldsymbol{\theta}$ makes the covariances of $\boldsymbol{\xi}_k$ *non-stationary* in space-time.

The spatio-temporal covariance function of the non-stationary field ξ is a function of four arguments, $B(t_1,t_2;s_1,s_2) = \mathbb{E}\,\xi(t_1,s_1)\xi(t_2,s_2)$. We will be particularly interested in the following two aspects of $B(t_1,t_2;s_1,s_2)$: the time and space specific *field variance* $\mathsf{Var}\,\xi(t,s) = B(t,t;s,s)$ and a time and space specific *spatial length scale* defined to be, say, the local macroscale:

$$\Lambda_\xi(t,s) = \frac{1}{2\,\mathsf{Var}\,\xi(t,s)} \int_{\mathbb{S}^1(R)} B(t,t;s,s')\mathrm{d}s'. \tag{27}$$

As it follows from Eq.(26), the spatio-temporal field covariances deterministically depend on the secondary fields $\boldsymbol{\theta}(t,s)$ and the initial covariances at the start of the time integration of Eq.(17). For large enough t, the impact of the initial conditions fades out and it appears that a spatio-temporal realization of the random fields $\boldsymbol{\theta}$ gives rise to a specific pattern in the spatio-temporal field covariances, in particular, to specific fields $\mathsf{Var}\,\xi(t,s)$ and $\Lambda_\xi(t,s)$. Moreover, since the fields $\boldsymbol{\theta}$ are stationary in space-time (we recall that these are governed by their advection-diffusion-decay models with constant coefficients), so are the fields $\mathsf{Var}\,\xi(t,s)$ and $\Lambda_\xi(t,s)$, whose variances and length scales are thus determined by the variances and the length scales of the secondary fields $\boldsymbol{\theta}$.

3.4.2 Estimation of the true covariances

Instead of the recursive computation of the field covariances using Eq.(26), one can *estimate* them as accurately as needed by running the DSADM Eq.(17) M times with independent realizations of the forcing $\alpha(t,s)$ and with the fields $\boldsymbol{\theta}(t,s)$ held fixed. The spatial covariance $B(t,t;s_1,s_2)$ can then be estimated from the resulting sample $\xi^{(m)}(t,s)$, $m=1,\ldots,M$, as the usual sample covariance.

Finally, we stress that knowing the above *signal covariances* can be useful, but it is more important for us that the above estimation procedure can be used to estimate the *true* time and space specific *error covariances* for any filter in question. To this end, we generate M assimilation runs in which the secondary fields $\boldsymbol{\theta}(t,s)$, the observation operators, and the observation-error covariance matrix $\mathbf{R}$ are held fixed, whilst the white noise α, observation errors, and random sources in the filter are all independent of each other. Then, the *true* background error covariance matrix $\mathbf{B}(t)$ can be estimated as the sample covariance matrix given M forecast-minus-truth differences $\mathbf{x}^{\mathrm{f}(m)}(t)-\mathbf{x}^{(m)}(t)$ (where $\mathbf{x}^{\mathrm{f}(m)}$ is the background, $\mathbf{x}^{(m)}$ is the truth, and m runs from 1 to M).

3.4.3 Instability

The nonlinear deterministic models mentioned in the Introduction are chaotic, that is, having unstable modes (positive Lyapunov exponents), whereas the DSADM is stochastic but experiencing intermittent instabilities due to the possibility for ρ and ν to attain negative values. In the deterministic models instabilities are curbed by the nonlinearity, whereas in the DSADM, these are limited by the time the random fields ρ and ν remain negative.

3.4.4 Gaussianity

Conditionally on $\boldsymbol{\theta}$, the DSADM is linear, hence $\xi(t,s)$ is conditionally Gaussian. Unconditionally, $\xi(t,s)$ is a continuous mixture of zero-mean Gaussian distributions and so should have a non-Gaussian distribution with heavy tails[3] (see section 4.4 for experimental evidence).

3.4.5 Unbeatable benchmark filter

The conditional linearity of the DSADM makes it possible to use the exact Kalman filter and thus to know how far from the optimal performance the filter in question is, which is often not possible with nonlinear deterministic models of truth.

[3] Indeed, one can prove using the Jensen inequality that the *kurtosis* of a non-degenerate mixture of this kind is always greater than 3 (the Gaussian kurtosis). A positive excess kurtosis means, normally, more probability mass in the tails of the distribution than in the tails of the Gaussian distribution with the same mean and variance.

4 Numerical experiments: behavior of the model

In this section we examine the spatio-temporal non-stationarity of solutions to the DSADM. The non-stationarity in the space-time covariances is examined conditionally on the secondary fields $\bar{\boldsymbol{\theta}}$. The *unconditional* probability distribution of ξ is considered only when we analyze non-Gaussianity of $\xi(t,s)$.

The model's differential equations were solved numerically using an unconditionally stable implicit upwind finite difference scheme. The solver was written in C++ (using the R package Rcpp), which led to a great speedup w.r.t. the R only code.

4.1 Setup

The experiments were carried out with the spatial grid size $n = 60$ on the circle of radius $R = 6370$km. The model integration time step was $\Delta t = 6$h. The time span was 600 time steps. The true field statistics were estimated from the sample of size $M = 5000$ (see section 3.4.2). The default external parameters were the following: $\bar{U} = 10$ ms^{-1}, $\bar{L} = 5\Delta s = 3300$km (where $\Delta s = 2\pi R/n$ is the spatial grid spacing), $L^* = \bar{L}$, $U_{\text{char}} = 5\text{ms}^{-1}$, $\overline{\text{SD}(\xi)} = 5$, $\text{SD}(U^*) = 10\text{ms}^{-1}$, $b = 1$, $\kappa_\rho = \kappa_\mu = \kappa_\sigma = 2$, $\pi_\rho = 0.1$, and $\pi_\nu = 0.05$.

4.2 $\xi(t,s)$ plots

Figure 2 compares typical spatio-temporal segments of the solutions to: (a) the stationary stochastic model Eq.(2) (i.e. the DSADM with the constant secondary fields) and (b) the DSADM Eq.(17). For comparison, a solution to the 60-variable Lorenz'96 model (with the popular choice of its parameter, $F = 8$) is presented in Fig.3. One can see that, first, the solutions to our stochastic and doubly stochastic models look much more realistic and resembling a fragment of a meteorological field than the solution to the Lorenz'96 model (which is unrealistically oscillatory and quasi-regular). Second, comparing the stationary and the non-stationary fields in Fig.2 reveals that, indeed, the non-stationary field shown in Fig.2(b) was substantially less regular than the stationary field displayed in Fig.2(a). Specifically, one can spot areas where the non-stationary field experienced more small-scale and more large-scale fluctuation than in the rest of the plot. The small-scale spots appear to be related to low and negative values of $\nu(t,s)$ and/or large values of $\sigma(t,s)$ (not shown). In contrast, areas where the non-stationary field was smoother than the stationary field are related to small and negative values of $\rho(t,s)$ and small values of $\sigma(t,s)$ (not shown). The dependencies of the spatial length scale of ξ on ρ and ν can be explained using Eq.(14).

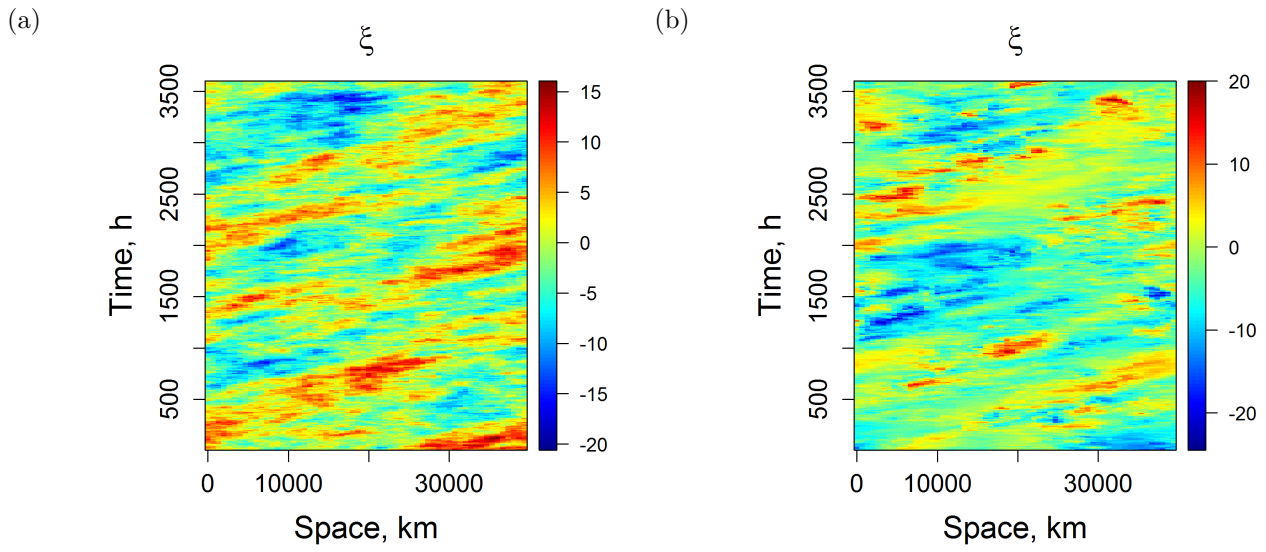

Figure 2: Space-time plots of the solution to the models: (a) stationary model Eq.(2), (b) DSADM Eq.(17).

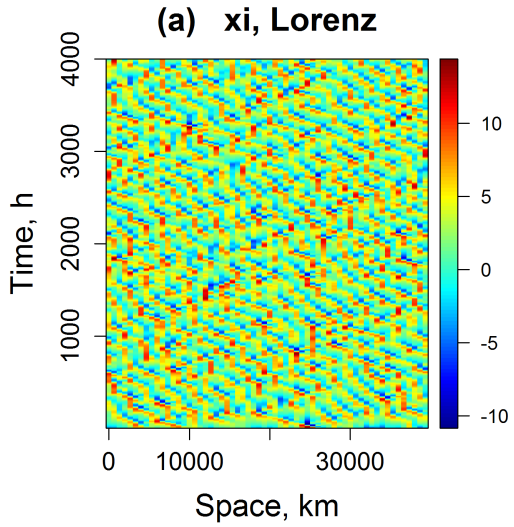

Figure 3: Space-time plot of the solution to the Lorenz'96 model.

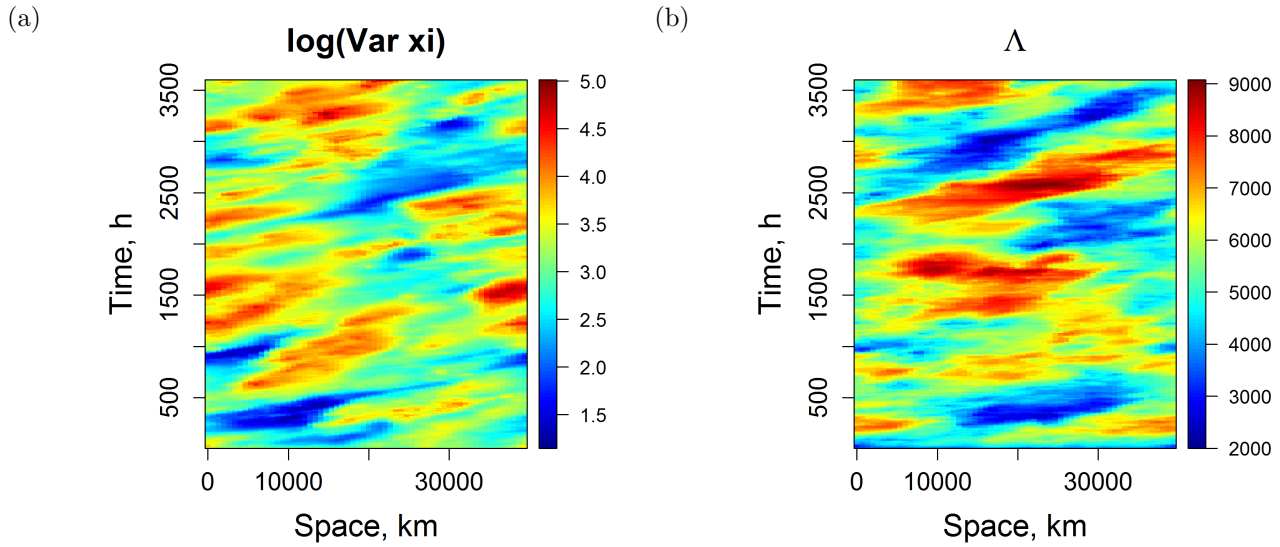

Figure 4: Space-time plots of estimated (a) field variance $\log(\mathsf{Var}\,\xi(t,s))$, (b) spatial macroscale $\Lambda_\xi(t,s)$.

4.3 Non-stationarity

Figure 4 shows the non-stationary field characteristics, $\log(\mathsf{Var}\,\xi(t,s))$ and $\Lambda_\xi(t,s)$, estimated following section 3.4.2 with the same realizations of the secondary fields that were used to compute ξ depicted in Fig.2(b). One can see the substantial degree of the non-stationarity (note that in the stationary case both $\mathsf{Var}\,\xi$ and Λ_ξ are constant). Specifically, in Fig.4, the ratio of the maximum to the minimum field variance $\mathsf{Var}\,\xi(t,s)$ was as large as about 50. The same ratio for the spatial length scale $\Lambda_\xi(t,s)$ was about 5, which also indicates a significant degree of variation. The spatio-temporal variation in both $\mathsf{Var}\,\xi(t,s)$ and $\Lambda_\xi(t,s)$ (and thus the degree of the non-stationarity in ξ) can be conveniently tuned by changing the (constant in space and time) variances and the spatial length scales of the secondary fields, respectively (not shown).

Figure 5 illustrates how diverse the true spatial correlations were. Note, however, that the diversity was, in a sense, limited. In particular, there were no negative lobes in the correlations. This point is discussed in section 6.

Thus, the DSADM generates significantly non-stationary random fields. The degree of the spatio-temporal non-stationarity is highly tunable.

4.4 Non-Gaussianity

Figure 6 displays the Gaussian quantile-quantile plot for the values of the field ξ (to get stable estimates, field values for as many as 60,000 time steps were used). One can see that the plot is far from the straight line, which means that the *unconditional* distribution of ξ was substantially non-Gaussian. Sample quantiles at the tails of the distributions are seen to be greater in modulus than they were for the matching Gaussian distribution (the

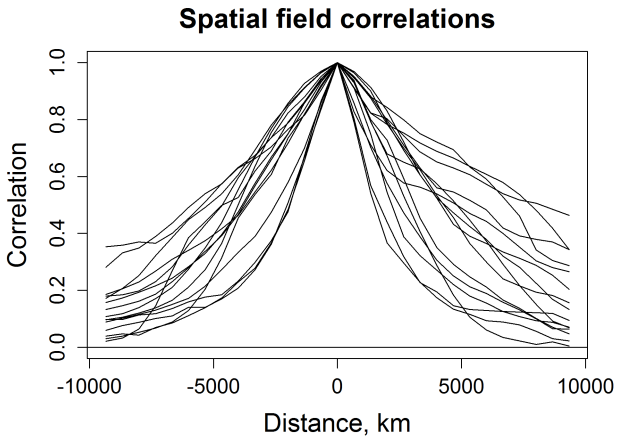

Figure 5: Spatial correlations of the field ξ w.r.t. 15 randomly selected points in space-time.

straight line). So, the unconditional distribution of $\xi(t, s)$ was indeed heavy-tailed as we anticipated in section 3.4.4. The higher the magnitudes of the secondary fields were, the higher the resulting non-Gaussianity in ξ became (not shown).

5 Application to studying temporal smoothing of EnKF background error covariances

Here, we give an example of how the DSADM can be used in practically oriented research aimed at the improvement of data assimilation (filtering) schemes. Specifically, we exploit the capability of the DSADM to allow the estimation of the *true* background error covariances for a filter (section 3.4.2). This capability implies that any tentative estimate of the time-dependent error covariance matrix produced by the filter (e.g. the sample covariance matrix, or a time mean sample covariance matrix, or their linear combination, etc.) can be verified against the respective true covariance matrix. The stochastic EnKF is studied. We show that a temporal smoothing of the EnKF's forecast-ensemble sample covariances yields a significant improvement in their accuracy.

5.1 Experimental methodology

Like in section 4 and according to our hierarchical paradigm, the hyperparameters $\boldsymbol{\zeta}$ were first selected (and kept constant in all experiments). Second, given $\boldsymbol{\zeta}$, the secondary fields $\boldsymbol{\theta}(t, s)$ were generated following section 3.2. Once computed, $\boldsymbol{\theta}(t, s)$ were kept fixed in an experiment. Third, given $\boldsymbol{\theta}(t, s) = (U(t, s), \rho(t, s), \nu(t, s), \sigma(t, s))$, a sample of M independent "truths" was generated in an experiment using Eq.(17).

Then, in each of the M runs, observations were generated by adding random noise

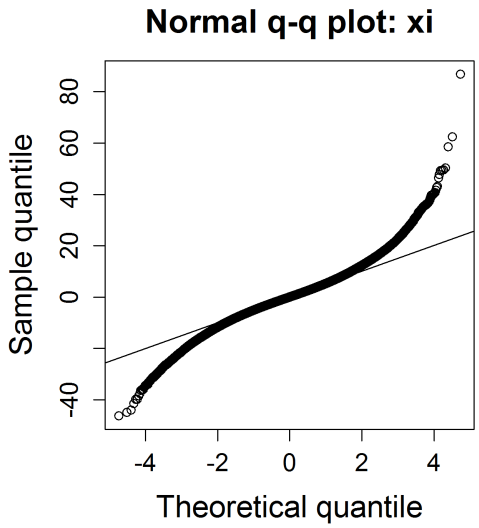

Figure 6: Gaussian q-q plot for the field ξ.

(with the diagonal covariance matrix $\mathbf{R}$) to the truth. These observations were assimilated by M stochastic EnKFs (with independent simulated model and observation errors). The length of an assimilation run was n_{time} (the number of the analyses in the run). Next, we estimated the actual (true) background error covariance matrices $\mathbf{B}_k$ (following section 3.4.2) for all time instants k (excluding the first 30 cycles to skip the initial transient period). Different experiments were carried out by specifying different random seeds to generate independent realizations of the secondary fields $\boldsymbol{\theta}(t, s)$.

5.2 Setup

The model setup was as in section 4.1. In the EnKF, the popular function by Gaspari and Cohn (1999, Eq.(4.10)) was used to specify the localization matrix $\mathbf{C}(R_{\text{loc}})$ (with R_{loc} being the cut-off radius beyond which the function vanishes). The optimally tuned R_{loc} was equal to 6000km. The tuned inflation factor (defined as a factor by which the forecast ensemble perturbations were multiplied) was 1.08. Observations were located at every 6th grid point in space. $\mathbf{R}$ was a multiple of the identity matrix; $\sqrt{\mathbf{R}_{ii}} = 4$ was selected to ensure that the mean forecast error variance reduction in an analysis was close to the value of 10% reported by Errico and Privé (2014) for a realistic data assimilation system. The default ensemble size was $N = 10$, $n_{time} = 300$, and the number of the experiments was 7.

The time interval between the consecutive analyses, T_a, was selected to mimic realistic temporal correlations of meteorological fields at the 6-h time shift (the most widely used in operational practice assimilation cycle). The time correlations (averaged over space and time) for the DSADM with the above external parameters were estimated to be 0.92, 0.86, and 0.77 for the time shifts 6h, 12h, and 24h, respectively. On the other hand, the

mid-tropospheric 6-h time correlations for the wind fields were estimated using radiosonde data to be about 0.8 (Fig.1(b) Seaman, 1975) and can be interpolated to the value of about 0.95 for geopotential (Olevskaya, 1968). Thus, we chose $T_a = 12$h in our system here, which implies that the time correlations between the consecutive analyses were roughly the same as in practical global data assimilation schemes. Both time and space correlation length scales in our system were roughly twice as large as their atmospheric counterparts.

5.3 Temporal covariance smoothing

In the Hierarchical Bayes Ensemble Filter (HBEF) proposed by Tsyrulnikov and Rakitko (2017), the predictive (analysis prior) covariances are cycled in a full-fledged secondary filter (with persistence at the forecast step). With the one-variable model reproduced here as Eq.(1), cycling the covariances was shown in (Tsyrulnikov and Rakitko, 2017) to be beneficial for the filter's performance. In the simplest version of the HBEF, the secondary filter reduces to the temporal smoothing of the (non-localized) forecast-ensemble sample covariances $\mathbf{S}_k$:

$$\mathbf{T}_k = \mu \mathbf{T}_{k-1} + (1-\mu)\mathbf{S}_k, \tag{28}$$

where $\mathbf{T}_k$ is the time-smoothed sample covariance and μ is the scalar ensemble-size dependent memory parameter.

Here, we verify that with the DSADM used as a model of truth and set up to have realistic temporal and spatial scale, the temporal smoothing of the prior covariances, Eq.(28), can make them significantly closer to the true prior covariances.

5.4 Results

We tested several candidate covariance matrices against the estimated truth $\mathbf{B}_k$ (the time index k is dropped below to simplify the notation):

- $\mathbf{B}_{\text{clim}}$: the time mean covariance matrix ("climatology"),
- $\mathbf{S}_{\text{loc}} = \mathbf{S} \circ \mathbf{C}(R_{\text{loc}})$: the ordinary localized ($\circ$ stands for element-wise matrix multiplication) sample covariance matrix with the empirically tuned localization cut-off radius R_{loc},
- $\mathbf{T}_{\text{loc}} = \mathbf{T} \circ \mathbf{C}(R^T_{\text{loc}})$: the localized time smoothed covariances (with the tuned cut-off radius R^T_{loc}),
- $\mathbf{B}_{\text{hybr}} = w\mathbf{S}_{\text{loc}} + (1-w)\mathbf{B}_{\text{clim}}$: the linear combination of $\mathbf{S}_{\text{loc}}$ with $\mathbf{B}_{\text{clim}}$ with the tuned weight w,
- $\mathbf{B}_{\text{hybr-T}} = w_T\mathbf{T}_{\text{loc}} + (1-w_T)\mathbf{B}_{\text{clim}}$ where w_T was also tuned to get the best results.

The "climatology" was calculated by averaging the true $\mathbf{B}_k$ over all time instants k and over all 7 experiments (i.e. over about 2100 assimilation cycles in total).

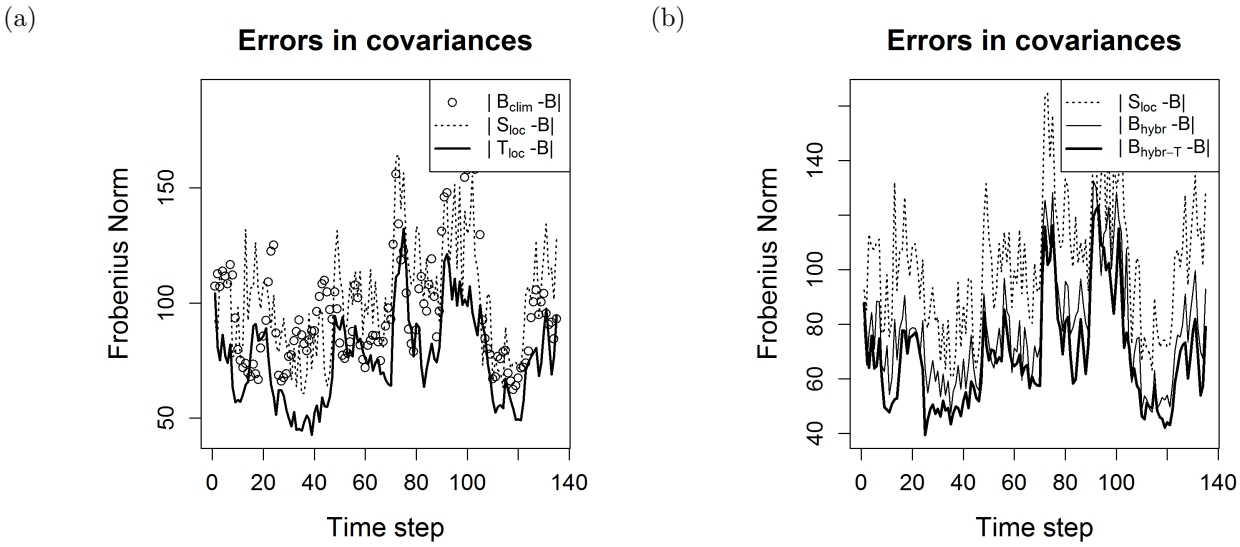

Figure 7: Errors in various candidate background error covariance matrices: (a) Climatology $\mathbf{B}_{\text{clim}}$, localized forecast ensemble sample covariances $\mathbf{S}_{\text{loc}}$, and temporally smoothed and then localized covariances $\mathbf{T}_{\text{loc}}$, (b) $\mathbf{S}_{\text{loc}}$, the tuned linear combination of $\mathbf{S}_{\text{loc}}$ with $\mathbf{B}_{\text{clim}}$ (denoted by $\mathbf{B}_{\text{hybr}}$), and the tuned linear combination of $\mathbf{T}_{\text{loc}}$ with $\mathbf{B}_{\text{clim}}$ (denoted by $\mathbf{B}_{\text{hybr-T}}$).

The error in each of the above candidate covariance matrices was measured by the Frobenius norm $\|.\|_F$ (whose suitability for data assimilation problems was advocated by Chorin and Morzfeld (2013)).

In general, the results were stable and not significantly different from one experiment to the other. Figure 7 shows the time series of errors for the same realization of the secondary fields as was used to demonstrate the model's performance in section 4. One can see that the sample covariances were indeed significantly improved by the temporal smoothing, both *per se* (panel a) and after the hybridization with $\mathbf{B}_{\text{clim}}$ (panel b).

It is interesting to note that the optimally tuned localization cut-off radius for the time smoothed covariances $\mathbf{T}$ appeared to be 1.7 times larger than that for the non-smoothed ones $\mathbf{S}$. So, the time smoothed covariances required a significantly lesser degree of localization, which indicates that these are less noisy than the sample covariances. The tuned hybridization weight was also a bit larger for the time smoothed covariances: $w_T = 0.7$ vs. $w = 0.6$. This implies the greater confidence in $\mathbf{T}_{\text{loc}}$ as compared with $\mathbf{S}_{\text{loc}}$.

The Frobenius norms of the errors in the covariances averaged over time and over the 7 experiments are shown in Table 1. Again, the effect of the temporal covariance smoothing is evident (compare column 4 vs. 3 and 6 vs. 5).

The effect of the temporal smoothing persisted in all different settings we tried: for the ensemble size $N = 20$, for weaker and stronger spatio-temporal non-stationarities of the DSADM generated fields, and for the assimilation cycle $T_a = 6$h (not shown).

Table 1: **Errors in various background error covariance matrices**

$\|\mathbf{B}\|_F$	$\|\mathbf{B}_{\rm clim}-\mathbf{B}\|_F$	$\|\mathbf{S}_{\rm loc}-\mathbf{B}\|_F$	$\|\mathbf{T}_{\rm loc}-\mathbf{B}\|_F$	$\|\mathbf{B}_{\rm hybr}-\mathbf{B}\|_F$	$\|\mathbf{B}_{\rm hybr\text{-}T}-\mathbf{B}\|_F$
209	117	112	86	89	79

We also tried to replace μ in Eq.(28) with a more sophisticated smoothing operator $\mathbf{\Phi T}_{k-1}\mathbf{\Phi}$, where $\mathbf{\Phi}$ is a symmetric spatial smoothing kernel. This led to a further improvement, but the effect was rather small (not shown) so in a practical scheme designing such a smoother would not be worth effort.

6 Discussion

The DSADM is shown to be capable of producing tunable non-stationarity in the conditional (given the secondary fields $\boldsymbol{\theta}$) probability distribution of the random field ξ in space-time. But there are limitations. In section 4.3 we noticed that the *shape* of the spatial correlations of ξ varies only weakly. For example, negative correlations are rare. This is because the *shape* of the local (i.e. with the "frozen" ρ and ν) spatial spectrum, see Eq.(13), remains the same whatever the local values of the fields ρ and ν (as well as U and σ) are: effectively, only m_0 (i.e. the inverse length scale) changes. To obtain more diverse shapes of the correlations, one can extend the DSADM by introducing to it higher-order spatial derivatives ($\partial^3/\partial s^3, \partial^4/\partial s^4$ etc.). Specifically, the terms $\rho\,\xi + U\,\partial\xi/\partial s - \nu\,\partial^2\xi/\partial s^2$ in the model Eq.(2) can be regarded as a second-order polynomial of the derivative operator $\partial/\partial s$, $P(\partial/\partial s) = \rho + U\partial/\partial s - \nu\,(\partial/\partial s)^2$. So, one can increase the order of the polynomial. This will give rise to a much richer class of correlation functions. Even more diverse correlations can be obtained if one introduces another polynomial, $Q(\partial/\partial s)$, and applies it to the right-hand side of the model.

The general idea of non-stationary random field modeling by assuming that parameters of a spatial or spatio-temporal model are random fields themselves was discussed by several authors, including Lindgren et al. (2011); Katzfuss (2013); Cressie and Wikle (2015). Banerjee et al. (2015, sec. 11.6) mentioned possibilities of formulating a stochastic differential equation for parameters of another (stochastic) differential equation, with the intention to add randomness to the process that satisfies the latter equation. Our innovation is the conditionally non-stationary hierarchical model with spatio-temporal stochastic partial differential equations at two levels in the hierarchy. The particular pattern of the non-stationarity is random, its spatio-temporal structure is highly tunable with the model's hyperparameters.

Concerning the temporal smoothing of ensemble covariances produced by a filter, we note that related ideas and techniques are already in use in practical systems. Gustafsson et al. (2014) use time-lagged ensemble members, Berre et al. (2015) make use of ensemble members from the past 4 days to increase the ensemble size, Bonavita et al. (2016) use en-

semble covariances from previous 12 days to estimate their parametric covariance model, Lorenc (2017) found that using time-lagged and time-shifted perturbations increases the effective ensemble size. Comparing those techniques with our recursive temporal smoothing, Eq.(28), would be interesting but this is beyond the scope of this study.

7 Conclusions

In this paper we have presented a new doubly stochastic advection-diffusion-decay model (DSADM) on the circle. Doubly stochastic means that not only the model's forcing is stochastic, the model's coefficients (parameters) are random as well. Moreover, the parameters are specified to be random fields satisfying their own stochastic advection-diffusion-decay models with constant coefficients. Thus, the DSADM is hierarchical, built of linear evolutionary stochastic partial differential equations at two levels in the hierarchy.

The model has two main advantages. First, conditionally on its parameter fields, the DSADM is capable of generating Gaussian spatio-temporal random fields with the tunable degree of non-stationarity. Second, if used as a model of "truth" in a data assimilation experiment, it allows the estimation of "true" time-specific signal and filtering covariances and, more broadly, probability distributions. Note that knowing the true prior/posterior covariances for a filter's analysis at each assimilation cycle implies that errors in these covariances are known, which can suggest ways to correct these errors.

Again conditionally, the DSADM is linear. The linearity is both a disadvantage (because real-world models are often nonlinear) and an advantage because an unbeatable benchmark, the Kalman filter, can be implemented with it and because some EnKF's issues like treatment of the sampling noise might be more easily tackled with a linear model. Unconditionally, the DSADM is nonlinear and the generated fields are non-Gaussian. The model exhibits tunable intermittent instability.

In numerical experiments, the DSADM is shown to produce non-stationary spatio-temporal fields with conveniently tunable structure. The capabilities of the model are illustrated in numerical experiments with the stochastic EnKF. It is experimentally shown that in the system with realistic spatial and temporal scales, the forecast-ensemble sample covariances can be significantly improved by their temporal smoothing.

What we did in this work can be called a *forward* hierarchical modeling. That is, we formulated a (hopefully) reasonable hierarchical model and gave an algorithm to compute realizations of the first-level random field $\xi(t, s)$ given the third-level hyperparameters $\boldsymbol{\gamma}$. A harder problem is the *inverse* modeling, that is, the inference about the model parameters—in our case the parameter fields $\boldsymbol{\theta}(t, s)$—from a number of realizations (an ensemble) of the field ξ. This is the classical hierarchical Bayesian problem, which is beyond the scope of this research but can be relevant in a broader context of non-stationary spatial and spatio-temporal field modeling. In particular, it can be used—with a conceptually similar hierarchical model—for Bayesian estimation of flow dependent background

error covariances from the ensemble, e.g. in the hierarchical filter by Tsyrulnikov and Rakitko (2017).

Appendices

A External model parameters

Here, we introduce external parameters of the DSADM and relate them to its hyperparameters.

The advection velocities $\bar{U}$ and U_θ (for $\theta = U, \rho, \nu, \sigma$) are specified directly as they have a clear physical meaning.

The unperturbed hyperparameters $\bar{\theta} = \bar{\rho}, \bar{\nu}, \bar{\sigma}$ are specified from the unperturbed external parameters $\bar{L}$, $\bar{T}$, and $\overline{\mathrm{SD}(\xi)}$ using Eqs.(14), (16), and (12). For any secondary field $\theta = U, \sigma, \rho$, or ν, the unperturbed value $\bar{\theta}$ can be shown to be the *median* of the pointwise probability distribution of $\theta(t, s)$.

Likewise, for the secondary perturbation fields $\theta^* = U^*, \rho^*, \nu^*, \sigma^*$, the hyperparameters ρ_θ and ν_θ are calculated from the respective external parameters L_θ and T_θ using Eqs.(14) and (16). $L_{\boldsymbol{\theta}}$ and $T_{\boldsymbol{\theta}}$ determine the respective *length scales of variation* in $\mathsf{Var}\,\xi(t, s)$, $\Lambda_\xi(t, s)$, etc. Say, if L_ν and L_ρ are large, then, due to Eq.(14), the random field $\Lambda_\xi(t, s)$ will vary smoothly in space, etc.

The hyperparameters σ_θ are specified as follows. σ_U is calculated from the external parameter $\mathrm{SD}(U^*)$ using Eq.(12). For each of the other three secondary fields, $\theta = \rho, \nu, \sigma$, we select $\kappa_\theta = \exp(\mathrm{SD}(\theta^*))$ as the respective external parameter. This choice is motivated by the fact that these three latter fields are nonlinearly transformed Gaussian fields. As $g(z)$ defined by Eq.(21) and shown in Fig.1 is a "tempered" exponential function, it is worth measuring the standard deviation of, say, the σ^* field on the log scale: $\mathrm{SD}(\sigma^*) = \log \kappa_\sigma$, so that the typical deviation of the transformed field σ from its unperturbed value $\bar{\sigma}$ is κ_σ *times*.

Finally, for the fields $\psi = \rho, \nu$, we parameterize ε_ψ (which, we recall, controls the occurrence of *negative* field values) using the probability π_ψ that $\psi(t, s)$ is negative:

$$P(\psi(t, s) < 0) = \pi_\psi. \tag{29}$$

Substituting ψ from Eq.(23) to Eq.(29) and utilizing the monotonicity of the transformation function g (see Eq.(21)) and the Gaussianity of the field ψ^*, we easily come up with a relation between π_ψ and ε_ψ (the elementary formulas are omitted). The external parameters π_ρ and π_ν determine how often and how strong *local instabilities* can be.

For some applications, the above 23 external parameters can be too many, so we introduced a reduced set as follows.

(i) By default, the unperturbed advection velocity $\bar{U}$ and the advection velocities that enter the models for the secondary perturbation fields, $U_{\boldsymbol{\theta}}$, are all equal to each other.

(ii) The spatial length scale hyperparameters $L_{\boldsymbol{\theta}}$ for all secondary perturbation fields $\boldsymbol{\theta}^*$ are selected to be equal to the common value L^*.

(iii) The temporal length scale hyperparameters $\bar{T}$ and $T_{\boldsymbol{\theta}}$ are specified to be equal to $\bar{L}/U_{\text{char}}$ and $L_{\boldsymbol{\theta}}/U_{\text{char}}$, respectively, where U_{char} is the *characteristic velocity* postulated to be the same for both $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^*$.

Thus, the reduced set of external parameters contains $\overline{\text{SD}(\xi)}$, $\bar{U}$, $\bar{L}$, L^*, U_{char}, $\text{SD}(U^*)$, κ_σ, κ_ρ, κ_ν, b, π_ρ, and π_ν, i.e. 12 parameters.

References

A. Apte, D. Auroux, and M. Ramaswamy. Variational data assimilation for discrete Burgers equation. *Electron. J. Diff. Equ.*, 19:15–30, 2010.

L. Arnold. *Stochastic differential equations*. Wiley, 1974.

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC Press, 2015.

L. Berre, H. Varella, and G. Desroziers. Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quart. J. Roy. Meteor. Soc.*, 141(692):2803–2812, 2015.

C. H. Bishop and E. A. Satterfield. Hidden error variance theory. Part I: Exposition and analytic model. *Mon. Weather Rev.*, 141(5):1454–1468, 2013.

M. Bonavita, E. Hólm, L. Isaksen, and M. Fisher. The evolution of the ECMWF hybrid data assimilation system. *Quart. J. Roy. Meteor. Soc.*, 142(694):287–303, 2016.

A. J. Chorin and M. Morzfeld. Conditions for successful data assimilation. *J. Geophys. Res.: Atmos.*, 118(20), 2013.

N. Cressie and C. K. Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.

R. Daley and R. Ménard. Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Weather Rev.*, 121(5):1554–1565, 1993.

R. M. Errico and N. C. Privé. An estimate of some analysis-error statistics using the Global Modeling and Assimilation Office observing-system simulation framework. *Quart. J. Roy. Meteor. Soc.*, 140(680):1005–1012, 2014.

G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, 125(554):723–757, 1999.

N. Gustafsson, J. Bojarova, and O. Vignes. A hybrid variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM). *Nonlin. Proc. Geophys.*, 21(1):303–323, 2014.

J. A. Hansen and L. A. Smith. Probabilistic noise reduction. *Tellus A*, 53(5):585–598, 2001.

A. Jazwinski. *Stochastic processes and filtering theory*. Academic Press, 1970.

M. Katzfuss. Bayesian nonstationary spatial modeling for very large datasets. *Environmetrics*, 24(3):189–200, 2013.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. Roy. Statist. Soc. B*, 73(4):423–498, 2011.

A. C. Lorenc. Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Quart. J. Roy. Meteor. Soc.*, 143(703):1062–1072, 2017.

E. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20(2):130–141, 1963.

E. Lorenz. Designing chaotic models. *J. Atmos. Sci.*, 62(5):1574–1587, 2005.

H. L. Mitchell and P. Houtekamer. Ensemble Kalman filter configurations and their performance with the logistic map. *Mon. Weather Rev.*, 137(12):4325–4343, 2009.

S. M. Olevskaya. The use of the canonical-correlation method in analyzing the geopotential field. *Izv. Atmos. Ocean. Phys.*, 4:1149–1159, 1968.

S. Särkkä. *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

R. Seaman. Distance-time autocorrelation functions for winds in the Australian region. *Austral. Meteorol. Mag.*, 23(2):27–40, 1975.

F. Sigrist, H. R. Künsch, and W. A. Stahel. Stochastic partial differential equation based modelling of large space–time data sets. *J. Roy. Statist. Soc. B*, 77(1):3–33, 2015.

D. Tjøstheim. Some doubly stochastic time series models. *J. Time Ser. Anal.*, 7(1):51–72, 1986.

M. D. Tsyroulnikov. Proportionality of scales: An isotropy-like property of geophysical fields. *Quart. J. Roy. Meteor. Soc.*, 127(578):2741–2760, 2001.

M. Tsyrulnikov and D. Gayfulin. Spatio-temporal stochastic pattern generator for ensemble prediction and ensemble data assimilation in geophysical applications. *arXiv:1605.02018*, 2016.

M. Tsyrulnikov and A. Rakitko. A hierarchical Bayes ensemble Kalman filter. *Physica D*, 338:1–16, 2017.

P. Whittle. *Systems in stochastic equilibrium*. John Wiley & Sons, Inc., 1986.

C. K. Wikle, L. M. Berliner, and N. Cressie. Hierarchical Bayesian space-time models. *Environ. Ecol. Stat.*, 5(2):117–154, 1998.

A. M. Yaglom. *Correlation theory of stationary and related random functions, Volume 1: Basic results*. Springer Verlag, 1987.