Meteorologische Zeitschrift, Vol. **26**, No. 5, 549–566 (published online September 24, 2017)

A limited-area spatio-temporal stochastic pattern generator for simulation of uncertainties in ensemble applications

Michael Tsyrulnikov* and Dmitry Gayfulin

HydroMetCenter of Russia

(Manuscript received July 13, 2016; in revised form November 10, 2016; accepted November 17, 2016)

Abstract

A generator of spatio-temporal pseudo-random Gaussian fields that satisfy the "proportionality of scales" property (Tsyroulnikov, 2001) is presented. The generator is based on a third-order in time stochastic differential equation with a pseudo-differential spatial operator defined on a limited area 2D or 3D domain in the Cartesian coordinate system. The generated pseudo-random fields are homogeneous and isotropic in space-time (with the scaled vertical and temporal coordinates). The correlation functions in any spatio-temporal direction belong to the Matérn class. The spatio-temporal correlations are non-separable. A spectral in space and finite-difference in time numerical solver is implemented and accelerated by exploiting properties of real-world geophysical fields, in particular, the smoothness of their spatial spectra. The generator is designed to create additive or multiplicative, or other spatio-temporal perturbations that represent uncertainties in numerical prediction models in geophysics. The generator is tested with the meteorological COSMO model as a source of additive spatio-temporal perturbations to the forecast-model fields. The program code of the generator is publicly available.

Keywords: Model error, Random field, Stochastic differential equation, Covariance function, Non-separable correlations, Isotropy

1 Introduction

1.1 Stochastic dynamic prediction

Since the works of Epstein (1969) and Tatarsky (1969), we know that accounting for the uncertainty in the initial forecast fields can improve weather (and other geophysical) predictions. Assigning a probability distribution for the truth at the start of the forecast (instead of using deterministic initial data) and attempting to advance this distribution in time according to the dynamic (forecast) model is called stochastic dynamic prediction.

The advantage of the stochastic dynamic prediction paradigm is twofold. First, the resulting forecast probability distribution provides a valuable measure of the *uncertainty* in the prediction, leading to probabilistic forecasting and flow-dependent background-error statistics in data assimilation. Second, for a nonlinear physical model, switching from the deterministic forecast to the mean of the forecast probability distribution improves the mean-square accuracy of the prediction, i.e. it can improve the deterministic forecasting.

1.2 Model errors

Since Pitcher (1977), we have realized that not only uncertainties in the initial data (analysis errors) matter, forecast model (including boundary conditions) imperfections also play an important role. As the simulation of model errors is the subject of this study, we define them now. Let the forecast model be of the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(\mathbf{x}), \tag{1.1}$$

where t is time, $\mathbf{x}$ is the vector that represents the (discretized) state of the system, and $\mathbf{F}$ is the model (forecast) operator. The imperfection of the model Eq. (1.1) means that the (appropriately discretized) truth does *not* exactly satisfy this equation. The discrepancy is called the model error (e.g. Orrell et al., 2001):

$$\boldsymbol{\xi}^t = \mathbf{F}(\mathbf{x}^t) - \frac{\mathrm{d}\mathbf{x}^t}{\mathrm{d}t}. \tag{1.2}$$

The true model error $\boldsymbol{\xi}^t$ is normally unknown. In order to include model errors in the stochastic dynamic prediction paradigm, one *models* $\boldsymbol{\xi}^t(t)$ as a *random process*, $\boldsymbol{\xi}(t)$, or, in other words, as a spatio-temporal random field $\xi(t,\mathbf{s})$ (where $\mathbf{s}$ is the spatial vector). The probability distribution of $\boldsymbol{\xi}(t)$ (in most cases, dependent on the flow) is assumed to be known.

Rearranging the terms in Eq. (1.2), and replacing the unknown $\boldsymbol{\xi}^t$ with its stochastic counterpart $\boldsymbol{\xi}$, we realize that the resulting model of truth is the stochastic dynamic equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(\mathbf{x}) - \boldsymbol{\xi}. \tag{1.3}$$

*Corresponding author: Michael Tsyrulnikov, HydroMetCenter of Russia, 11–13 B. Predtechensky Lane, Moscow, Russia, e-mail: michael.tsyrulnikov@gmail.com

Thus, the extended stochastic dynamic prediction (or modeling) paradigm requires two input probability distributions (that of initial errors and that of model errors) and aims to transform them to the output (forecast) probability distribution.

1.3 Ensemble prediction

Stochastic dynamic modeling of complex geophysical systems is hampered by their high dimensionality and non-linearity. For realistic models, the output probability distribution is analytically intractable. An affordable approximate solution is provided by the Monte-Carlo method, which is called in geosciences the *ensemble prediction*.

In ensemble prediction, the input uncertainties (i.e. initial and model errors) are represented by simulated *pseudo-random draws* from the respective probability distributions. A relatively small affordable number of these draws are fed to the forecast model giving rise to an *ensemble* of predictions (forecasts). *If initial and model errors are sampled from the correct respective distributions, then the resulting forecast ensemble members are draws from the correct probability distribution of the truth given all available external data (initial and boundary conditions).* This mathematically justifies the ensemble prediction principle. From the practical perspective, members of the forecast ensemble can be interpreted as "potential truths" consistent with all available information.

In what follows, we concentrate on the model error field $\xi(t, \mathbf{s})$. We briefly review existing models for $\xi(t, \mathbf{s})$ and then present our stochastic pattern generator, whose goal is to simulate pseudo-random draws of $\xi(t, \mathbf{s})$ from a meaningful and flexible distribution.

1.4 Practical model error modeling

In meteorology, our knowledge of the actual model error probability distribution is scarce. Justified stochastic model-error models are still to be devised and verified. In the authors' opinion, the best way to stochastically represent spatio-temporal forecast-model-error fields is to treat each error source separately, so that, say, each physical parametrization is accompanied by a spatio-temporal stochastic model of its uncertainty. Or, even better, to completely switch from deterministic physical parameterizations to stochastic ones. There is a growing number of such developments (see Berner et al., 2017, for a review), but the problem is so complex that we cannot expect it to be solved in the near future. Its solution is further hampered by the fact that the existing meteorological observations are too scarce and too inaccurate for model errors to be objectively identified by comparison with measurement data with satisfactory accuracy (Tsyrulnikov and Gorin, 2013).

As a result, in meteorology, *ad-hoc* model-error models are widely used. The existing approaches can be classified as either non-stochastic or stochastic. Non-stochastic schemes can be multi-model (different ensemble members are generated using different forecast models) or multi-parameterization (each ensemble member is generated using the forecast model with a unique combination of different physical parameterization schemes or their parameters). These techniques are capable of introducing significant diversity in the ensemble (Berner et al., 2011), but the resulting ensemble members cannot be considered as independent and drawn from the same probability distribution (an assumption normally made in using the ensembles). Besides, there are not enough different models and not enough substantially different physical parameterizations to generate large ensembles. Finally, running many forecast models is a technologically very demanding task.

Stochastic approaches, on the contrary, offer the opportunity to generate as many ensemble members taken from the same probability distribution as needed, while working with just one forecast model and one set of physical parameterizations. In atmospheric ensemble prediction and ensemble data assimilation, the most widely used stochastic techniques are SPPT (Stochastic Perturbations of Physical Tendencies, Buizza et al., 1999), SKEB (Stochastic Kinetic Energy Backscatter scheme, Shutts, 2005), and SPP (Stochastically Perturbed Parameterizations, Christensen et al., 2015; Ollinaho et al., 2017). In the SPPT, multiplicative perturbations to the *tendencies* produced by the model's physical parameterizations are introduced. The multiplier is a spatio-temporal random field centered at 1. In the SKEB, *additive* perturbations are computed by modulating a spatio-temporal random field by the local kinetic energy dissipation rate. In the SPP, selected *parameters* of the physical parameterization schemes are perturbed again using a spatio-temporal field, which thus is seen to be needed in all of the above stochastic model error representation schemes. Stochastic parameterization schemes can also demand such fields (e.g. Bengtsson et al., 2013).

1.5 Generation of spatio-temporal random fields

The simplest non-constant pseudo-random field is the white noise, i.e. the uncorrelated in space and time random field. The white noise is the default forcing in stochastic differential equations, e.g. Jazwinski (1970) or Arnold (1974). Its advantage is the complete absence of any spatio-temporal structure, it is a pristine source of stochasticity. But in model-error modeling, this lack of structure precludes its direct use as an additive or multiplicative perturbation field because model errors are related to the weather pattern and so should be

correlated (dependent) both in space and time. Tsyrulnikov (2005) showed in a simulation study that model errors can exhibit complicated spatio-temporal behavior.

A correlated pseudo-random spatio-temporal field can be easily computed by generating independent random numbers at points of a *coarse* spatio-temporal grid and then assigning each of them to all model grid points within the respective coarse-grid cell (Buizza et al., 1999). As a result, the model-grid field becomes correlated in space and time. The decorrelation space and time scales are, obviously, defined by the respective coarse grid spacings (e.g. in Buizza et al. (1999) these were about 1000 km in space and 6 h in time). This technique is extremely simple, but it suffers from two flaws.

First, the resulting model-grid field appears to be discontinuous and inhomogeneous. Second, it has no *space-time interactions*, that is, the temporal length scales do not depend on the respective spatial scales. In reality, longer spatial scales "live longer" than shorter spatial scales, which "die out" quicker. This 'proportionality of scales' is widespread in geophysical fields (see Tsyroulnikov, 2001, and references therein) and other media, (e.g. Meunier and Zhao, 2009, p. 129), so we believe this property should be represented by model-error models. Note also that the "proportionality of scales" is a special case of the *non-separability* of spatio-temporal covariances. For a critique of simplistic separable space-time covariance models, see Cressie and Huang (1999), Stein (2005), Gneiting et al. (2006), and Section 2 below.

Another popular space-time pseudo-random field generation technique employs a spectral transform in space and then imposes independent temporal auto-regressions for the coefficients of the spectral expansion (Berner et al., 2009; Palmer et al., 2009; Charron et al., 2010; Bouttier et al., 2012). This technique is more general and produces homogeneous fields, but the above implementations use *the same time scale* for all spatial wavenumbers so that there are still no space-time interactions in the generated spatio-temporal fields (though Charron et al. (2010) noted that the decorrelation time scales can be made dependent on the spatial scales and Palmer et al. (2009) allowed for this dependence in their SKEB pattern generator equations).

In this article, we propose and test a spatio-temporal Stochastic (pseudo-random) Pattern Generator (SPG) that accounts for the above "proportionality of scales" and imposes meaningful space-time interactions. The SPG operates on a limited-area domain. It is based on a (spectral-space) solution to a stochastic partial differential equation, more precisely, to a stochastic differential equation in time with a pseudo-differential spatial operator. In what follows, we present the technique, examine properties of the resulting spatio-temporal fields on 2D and 3D spatial domains, describe the numerical scheme, and give an example of an application of the SPG in a forecast model used for operational weather prediction. The technique is implemented as a Fortran program freely available from https://github.com/gayfulin/SPG.

2 Model error fields: separability vs. "proportionality of scales"

In this motivational section, we show for a simple 1D (in space) example that space-time interactions in the model error random field do play a significant role. Specifically, we demonstrate that these interactions determine whether the spatial length scale of the resulting forecast error field grows, in a first approximation, in time or remains constant.

We note that for small enough model error perturbations and small enough lead times, the forecast error due to the accumulated model errors can be approximated by the time integrated model error, $\bar{\xi}(t,s) = \int_0^t \xi(t,s)\,\mathrm{d}t$, also known as the model-error drift (Orrell et al., 2001). Therefore, the methodology in this section is to take two fields, one with separable spatio-temporal correlations and the other with "proportional scales", integrate them in time, and look at the spatial length scales of the two time-integrated random fields.

Theoretically, the time integration reduces (filters out) small-scale-in-time components of the field. As a separable field has no space-time interactions, its time integral should have exactly the same spatial length scale as $\xi(t,s)$. For a proportional-scales field, smaller scales in time are associated with smaller scales in space, so the amount of small spatial scales in the time integrated field should decrease in time leading to an increase in the spatial length scale.

To verify these theoretical conclusions, we set up the following numerical experiment. We considered a 1D domain of size 100 km and the time integration period of 3 h. In this 2D spatio-temporal domain, we introduced a grid with 100 points in space and 100 points in time. On this grid, we simulated two random fields, both with unit variance and exactly the same spatial and temporal exponential correlations. The first field had separable correlations $C_1(\Delta t, \Delta s) = \exp(-|\Delta s|/L) \cdot \exp(-|\Delta t|/T)$, whereas the second field had non-separable correlations $C_2(\Delta t, \Delta s) = \exp(-\sqrt{(\Delta s/L)^2 + (\Delta t/T)^2})$, which can be shown to satisfy the "proportionality of scales" property. The spatial length scale L was selected in such a way that the spatial correlation function intersects the 0.7 level at the distance of 50 km. The temporal length scale was selected to be equal to L/U, where $U = 20$ m/s was taken as the characteristic flow velocity. Note that both the separability and the exponential temporal correlation function are what the scale-independent first-order auto-regressions used in Berner et al. (2009), Palmer et al. (2009), Charron et al. (2010), and Bouttier et al. (2012) imply.

Knowing the two correlation functions, we simulated pseudo-random realizations of the two fields (by building the two covariance matrices, computing their square roots, and applying the latter to vectors of independent $N(0,1)$ random variables), see Fig. 1. In comparing the two panels of Fig. 1, one can see that the two fields look quite different. Visually, the most striking difference is the lack of isotropy in the separable case. The

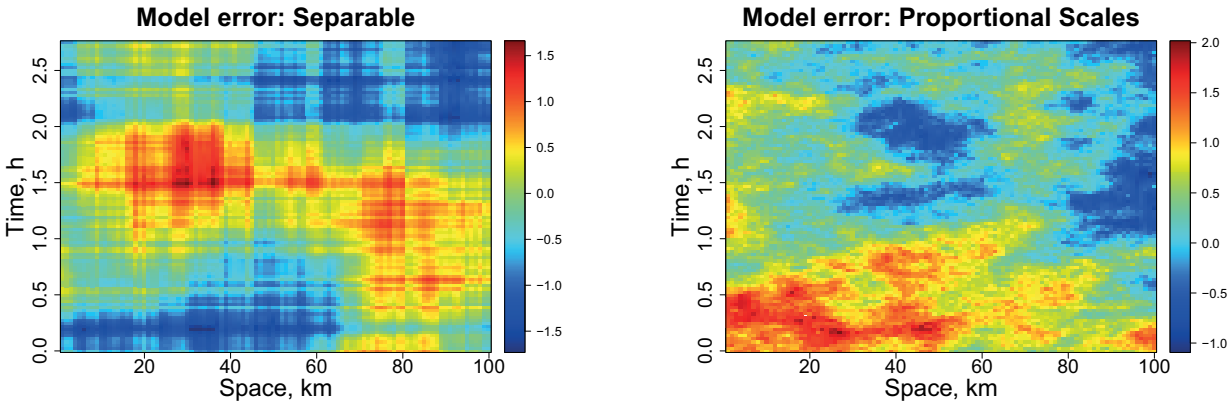

Figure 1: Simulated spatio-temporal fields. *Left*: With separable space-time correlations. *Right*: With non-separable proportional-scales correlations.

proportional-scales field looks much more realistic than the separable one.

To get a more objective criterion, we computed the time integrated model error field $\bar{\xi}(t, s)$ (the model error drift, a proxy to the model-error induced forecast error, see above in this section). Fig. 2 shows the spatial cross-sections of the arbitrarily chosen realizations of the model error fields (left) and the drift fields (right). The realizations generated by the separable random field model are given in black and the realizations of the proportional-scales field are represented by the red curves. One can see that, indeed, the time integration did not change the spatial structure of the *separable* field (compare the two black curves in Fig. 2, left and right). In contrast, the time integrated *proportional-scales* field becomes much smoother in space (compare the two red curves in Fig. 2, left and right).

Even more objectively, we estimated the *spatial micro-scale* of the drift $\bar{\xi}(t, s)$. The estimator was $(\mathsf{Var}\,\bar{\xi}/\mathsf{Var}\,\delta\bar{\xi})^{1/2} \cdot h$, where $\delta\xi$ is the forward finite difference in space, the variance Var was estimated by averaging over the space coordinate and over an ensemble of 100 realizations), and h is the spatial mesh size. The resulting spatial micro-scales for the two fields in question are displayed in Fig. 3 as functions of time. As expected, in the separable case, the spatial micro-scale did not change as a result of the time integration (the flat black line), whereas in the non-separable proportional-scale case, the spatial length scale of the model error drift rapidly grew in time. It is worth emphasizing that in data assimilation, the spatial length scale is a very important attribute of the forecast error field and thus needs to be correctly represented by a forecast (background) ensemble.

Thus, we have shown that the specific type of the spatio-temporal interactions in a model (tendency) error field has important consequences for the spatial structure of the resulting practically relevant forecast error field. We have no evidence on the actual model error spatio-temporal structure, but we know that non-separability and, more specifically, proportionality of scales is ubiquitous in geophysics (Tsyroulnikov, 2001). Therefore, we postulate that the SPG should produce proportional-scales fields.

3 SPG: Requirements and approach

The general requirements are:

1. The SPG should produce univariate stationary in time and homogeneous (stationary) and isotropic in space Gaussian pseudo-random fields $\xi(t, \mathbf{s})$ in 3D and 2D spatial domains.

2. The SPG should be fast enough so that it does not significantly slow down the forecast model computations.

3. Variance as well as the spatial and temporal length scales of $\xi(t, \mathbf{s})$ are to be tunable.

   We also impose more specific requirements:

4. The random field $\xi(t, \mathbf{s})$ should have finite variance and continuous realizations (sample paths).

5. The spatio-temporal covariances should obey the "proportionality of scales" principle: larger (shorter) spatial scales should be associated with larger (shorter) temporal scales (Tsyroulnikov, 2001).

6. The SPG *ansatz* should be flexible enough to allow for practicable solutions in both physical space and spectral space.

Two comments are in order. Firstly, stationarity, homogeneity, isotropy, and Gaussianity imposed by requirement 3 are just the simplest natural properties of a spatio-temporal random field. The SPG is intended to be used as a building block in practical schemes like the above SPPT, SKEB, SPP, or others. Its role

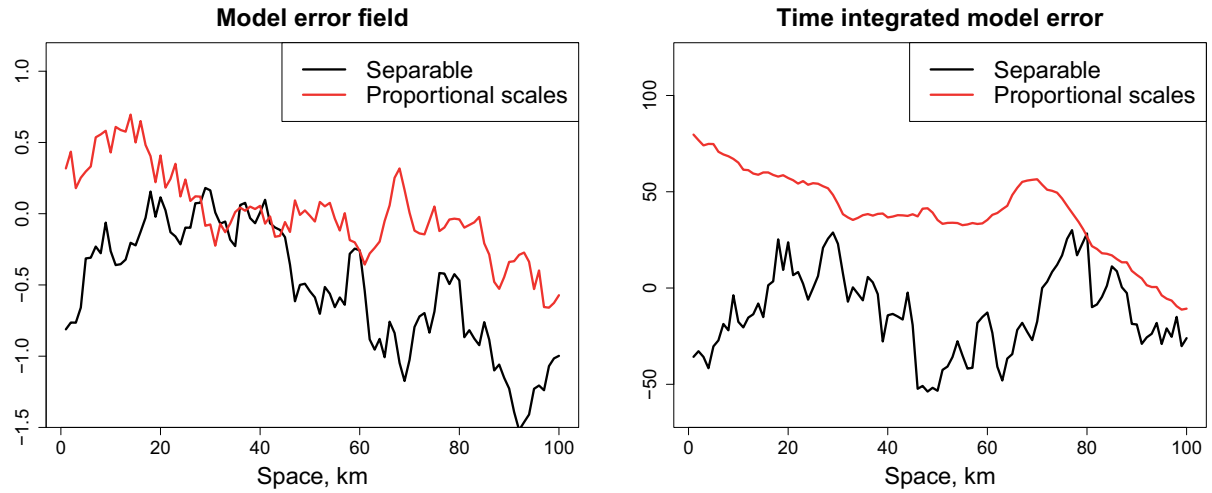

Figure 2: Spatial cross-sections of the simulated fields. *Left*: Model error fields. *Right*: Time integrated model error fields.

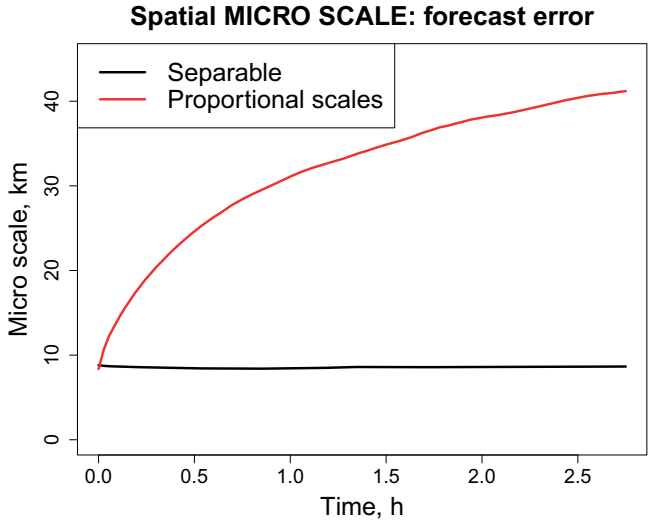

Figure 3: Spatial micro-scale as a function of integration time for the two time integrated random fields (separable and proportional-scales).

is to be the source of meaningful and easily tunable spatio-temporal stochasticity, whereas physical model error features (flow dependence, non-Gaussianity, etc.) are to be provided by the specific model error modeling scheme on a point-by-point basis.

Secondly, requirement 3 demands an SPG equation to be solvable in physical space as well as in spectral space for the following reasons. All the above mentioned existing pattern generators are spectral space based because this is the simplest way to get a homogeneous and isotropic field in physical space. So, following this path, we would like to have a spectral-space solver. But we envision that a combination of a homogeneous and isotropic spatial structure (provided by the SPG) and point-by-point flow dependent and/or non-Gaussian features (provided by the specific model error modeling scheme) can appear too restrictive in the near future. Specifically, this combination approach cannot produce variable local spatial and temporal length scales if used in schemes like the SKEB or SPP (for example, we may wish to reduce the local length scales in meteorologically active areas like cyclones or convective systems). Therefore, we wish the SPG equation to allow for a physical-space solver that would be capable of imposing variable in space and time structures.

As a starting point in the development of the SPG, we select the general class of *linear evolutionary stochastic partial differential equations* (SPDE). This choice is motivated by the flexibility of this class of spatio-temporal models (e.g. Lindgren et al., 2011). In particular, for an SPDE, it is relatively easy to introduce inhomogeneity in space and time as well as local anisotropy – either by changing coefficients of the spatial operator or by changing local properties of the driving noise. One can also produce non-Gaussian fields by making the random forcing non-Gaussian (e.g. Åberg and Podgórski, 2011; Wallin and Bolin, 2015). Physical-space discretizations of SPDEs lead to *sparse* matrices, which give rise to fast numerical algorithms. If an SPDE has constant coefficients, then it can be efficiently solved using spatial spectral-space expansions.

In this study, we develop the SPG that relies on a spatio-temporal stochastic model with *constant coefficients* so that both physical-space and spectral-space solvers can be employed. To facilitate the spectral-space solution, the general strategy is to define the SPG model on a standardized spatial domain. The operational pseudo-random fields are then produced by mapping of the generated fields from the standardized domain to the forecast-model domain. In 3D, the standardized spatial domain is chosen to be the unit cube with the periodic boundary conditions in all three dimensions, in other words, the three-dimensional (3D) unit torus. In 2D, the standardized domain is the 2D unit torus. The 3D and 2D cases are distinguished by the dimensionality $d = 2$ or $d = 3$ in what follows. To simplify the presentation, the default dimensionality will be $d = 3$.

4 Tentative first-order SPG model

4.1 Physical-space model

The random field in question $\xi(t,\mathbf{s})$ is a function of the time coordinate t and the space vector $\mathbf{s}=(x,y,z)$, where (x,y,z) are the three spatial coordinates. Each of the spatial coordinates belongs to the the unit circle $\mathbb{S}^1$, so that $\mathbf{s}$ is on the unit torus $\mathbb{T}^3 \equiv \mathbb{S}^1\times\mathbb{S}^1\times\mathbb{S}^1$ ($\mathbb{T}^2$ in the 2D case).

We start with the simplest general form of the first-order Markov model:

$$\frac{\partial\xi(t,\mathbf{s})}{\partial t} + A\,\xi(t,\mathbf{s}) = \alpha(t,\mathbf{s}), \tag{4.1}$$

where A is the spatial linear operator to be specified and α is the white in space and time driving noise. We postulate α to be the white noise in order for a physical-space numerical solver (demanded by requirement 3 in Section 3) to be fast. This is because generation of the white noise is computationally inexpensive (its values on a grid in space and time are just independent Gaussian random variables).

The SPG is required to be fast, so we choose A to be a *differential* operator (because, as we noted, in this case a physical-space discretization of A gives rise to a very sparse matrix).

Further, since we wish $\xi(t,\mathbf{s})$ to be homogeneous and isotropic in space, we define A to be a polynomial of the negated spatial Laplacian:

$$A = P(-\Delta) = \sum_{j=0}^{q} c_j(-\Delta)^j, \tag{4.2}$$

where $P(x)$ is the polynomial and q its degree (a positive integer). We will refer to q as the spatial order of the SPG model. Note that the negation of the Laplacian is convenient because $(-\Delta)$ is a non-negative definite operator.

The model Eq. (4.2) appears to be too rich for the purposes of the SPG at the moment, so in what follows we employ an even more reduced (but still quite flexible) form

$$A = P(-\Delta) = \mu(1-\lambda^2\Delta)^q, \tag{4.3}$$

where μ and λ are positive real parameters. With the spatial operator A defined by Eq. (4.3), the tentative first-order SPG model becomes

$$\frac{\partial\xi(t,\mathbf{s})}{\partial t} + \mu(1-\lambda^2\Delta)^q\,\xi(t,\mathbf{s}) = \alpha(t,\mathbf{s}). \tag{4.4}$$

4.2 Spectral-space model

On the torus $\mathbb{T}^d$, a Fourier series is an expansion in the basis functions $\mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})} \equiv \mathrm{e}^{\mathrm{i}(mx+ny+lz)}$, where the wavevector $\mathbf{k}$ is, for $d=3$, the triple of integer wavenumbers, $\mathbf{k}=(m,n,l)$. We perform the Fourier decomposition for both $\alpha(t,\mathbf{s})$ and $\xi(t,\mathbf{s})$,

$$\alpha(t,\mathbf{s}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \tilde\alpha_{\mathbf{k}}(t)\mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})} \tag{4.5}$$

and

$$\xi(t,\mathbf{s}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \tilde\xi_{\mathbf{k}}(t)\mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})} \tag{4.6}$$

(where $\mathbb{Z}$ denotes the set of integer numbers) and substitute these expansions into Eq. (4.4). From the orthogonality of the basis functions, we obtain that Eq. (4.4) decouples into the set of ordinary stochastic differential equations (OSDE, e.g. Jazwinski, 1970; Arnold, 1974) in time:

$$\frac{\mathrm{d}\tilde\xi_{\mathbf{k}}}{\mathrm{d}t} + \mu(1+\lambda^2k^2)^q\,\tilde\xi_{\mathbf{k}}(t) = \tilde\alpha_{\mathbf{k}}(t), \tag{4.7}$$

where $k=|\mathbf{k}|=\sqrt{m^2+n^2+l^2}$. The white driving noise α is homogeneous, hence the spectral-space coefficients $\tilde\alpha_{\mathbf{k}}(t)$ are probabilistically independent random processes. This is well known for random fields on the d-dimensional real space $\mathbb{R}^d$ (where spectra are continuous), see, e.g., Chapter 2 in Adler (1981) or Section 8 in Yaglom (1987), and can be directly verified in our case of the fields on the torus (where spectra are discrete). Therefore, for different wavevectors $\mathbf{k}$, the resulting spectral-space equations are probabilistically completely *independent* from each other. This greatly simplifies the solution of the SPG equations, because instead of handling the complicated SPDE Eq. (4.4), we solve a number of independent simple OSDEs, Eq. (4.7).

Further, from the postulated whiteness of the spatio-temporal random field $\alpha(t,\mathbf{s})$, all $\tilde\alpha_{\mathbf{k}}(t)$ are white-in-time random processes with the same intensity σ, see Appendix A in Tsyrulnikov and Gayfulin (2016):

$$\tilde\alpha_{\mathbf{k}}(t) = \sigma\,\Omega_{\mathbf{k}}(t), \tag{4.8}$$

where $\Omega_{\mathbf{k}}(t)$ are the independent *standard* white noises, i.e. the derivatives of the independent standard Wiener processes $W_{\mathbf{k}}(t)$ such that

$$\Omega_{\mathbf{k}}(t)\mathrm{d}t = \mathrm{d}W_{\mathbf{k}}(t). \tag{4.9}$$

Thus, the first-order SPG model reduces to a series of OSDEs

$$\mathrm{d}\tilde\xi_{\mathbf{k}} + \mu(1+\lambda^2k^2)^q\,\tilde\xi_{\mathbf{k}}\,\mathrm{d}t = \sigma\,\mathrm{d}W_{\mathbf{k}}. \tag{4.10}$$

For practical purposes, the series is truncated, so that $\mathbf{k}\equiv(m,n,l)$ is limited: $|m|<m_{\max}$, $|n|<n_{\max}$, and $|l|<l_{\max}$, where $m_{\max}$, $n_{\max}$, and $l_{\max}$ are the truncation limits. If not otherwise stated, all the truncation limits are the same and denoted by $n_{\max}$.

4.3 Stationary spectral-space statistics

Equation (4.10) is a first-order OSDE with constant coefficients sometimes called the Langevin equation (e.g.

Arnold (1974) or Jazwinski (1970), Example 4.12). Its generic form is

$$\mathrm{d}\eta + a\eta\,\mathrm{d}t = \sigma \mathrm{d}W, \tag{4.11}$$

where $\eta(t)$ is the random process in question, a and σ are constants, and $W(t)$ is the standard Wiener process. The solution to Eq. (4.11) is known as the Ornstein-Uhlenbeck random process, whose stationary (steady-state) temporal covariance function is

$$B_\eta(t) = \frac{\sigma^2}{2a}\,\mathrm{e}^{-a|t|} \tag{4.12}$$

(e.g. Jazwinski, 1970, Example 4.12). From Eq. (4.12), it is clear that a has the meaning of the inverse temporal length scale $\tau = 1/a$.

Now, consider the stationary covariance function of the elementary random process $\tilde{\xi}_{\mathbf{k}}(t)$,

$$\mathbb{E}\,\tilde{\xi}_{\mathbf{k}}(t_0)\cdot\tilde{\xi}_{\mathbf{k}}(t_0+t) = b_{\mathbf{k}}\cdot C_{\mathbf{k}}(t), \tag{4.13}$$

where $b_{\mathbf{k}}$ is the variance and $C_{\mathbf{k}}(t)$ the correlation function. According to Eq. (4.6), $\tilde{\xi}_{\mathbf{k}}$ is the spatial spectral component of the random field in question $\xi(t,\mathbf{s})$. Therefore, $b_{\mathbf{k}} = \mathsf{Var}\,\tilde{\xi}_{\mathbf{k}}$ is called the *spatial spectrum* of $\xi(t,\mathbf{s})$. From Eqs. (4.10) and (4.12), we have

$$b_{\mathbf{k}} = \frac{\sigma^2}{2\mu(1+\lambda^2k^2)^q} \tag{4.14}$$

and $C_{\mathbf{k}}(t) = \exp(-|t|/\tau_{\mathbf{k}})$, where

$$\tau_{\mathbf{k}} = \frac{1}{\mu(1+\lambda^2k^2)^q} \tag{4.15}$$

is the temporal length scale associated with the spatial wavevector $\mathbf{k}$.

Note that by the spectrum (e.g. $b_{\mathbf{k}}$) we always mean the *modal* spectrum, i.e. the variance associated with a single basis function (a single wavevector $\mathbf{k}$); the modal spectrum is not to be confused with the variance (or energy) spectrum.

4.4 Physical-space statistics

In the stationary regime (i.e. after an initial transient period has passed), the above independence of the spectral random processes $\tilde{\xi}_{\mathbf{k}}(t)$ (see Section 4.2) implies that the random field $\xi(t,\mathbf{s})$ is spatio-temporally *homogeneous* (stationary), i.e. invariant under shifts in space and time:

$$\mathbb{E}\,\xi(t,\mathbf{s})\cdot\overline{\xi(t+\Delta t,\mathbf{s}+\Delta\mathbf{s})} = B(\Delta t,\Delta\mathbf{s}), \tag{4.16}$$

where $\mathbb{E}$ is the expectation operator and

$$B(t,\mathbf{s}) = \sum_{\mathbf{k}} b_{\mathbf{k}}\,C_{\mathbf{k}}(t)\,\mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})}. \tag{4.17}$$

In particular, the spatial covariance function is

$$B(\mathbf{s}) = B(t=0,\mathbf{s}) = \sum_{\mathbf{k}} b_{\mathbf{k}}\,\mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})}, \tag{4.18}$$

where it is seen that the spatial spectrum $b_{\mathbf{k}}$ is the Fourier transform of the spatial covariance function $B(\mathbf{s})$.

The temporal covariance function is

$$B(t) = B(t,\mathbf{s}=\mathbf{0}) = \sum_{\mathbf{k}} b_{\mathbf{k}}\,C_{\mathbf{k}}(t). \tag{4.19}$$

Finally, the variance is

$$\mathsf{Var}\,\xi = B(t=0,\mathbf{s}=\mathbf{0}) = \sum_{\mathbf{k}} b_{\mathbf{k}}. \tag{4.20}$$

4.5 "Proportionality of scales" requires that $q = \frac{1}{2}$

The more precise formulation of the "proportionality of scales" requirement 3 states that for large k, the temporal length scale $\tau_{\mathbf{k}}$ should be inversely proportional to k:

$$\tau_{\mathbf{k}} \sim \frac{1}{k} \quad \text{as} \quad k\to\infty. \tag{4.21}$$

From Eq. (4.15), this condition entails, importantly, that

$$q = \frac{1}{2}. \tag{4.22}$$

Below, we show that the choice $q = 1/2$ causes the generated spatio-temporal random fields to possess, besides the "proportionality of scales", many other nice properties (sections 5.4 and 5.5).

4.6 The spatial operator of order $q = \frac{1}{2}$

The model's spatial operator A becomes (see Eq. (4.3))

$$A = \mu(1-\lambda^2\Delta)^{\frac{1}{2}} \equiv \mu\sqrt{1-\lambda^2\Delta}. \tag{4.23}$$

This is a pseudo-differential operator (e.g. Shubin, 1987) with the *symbol*

$$a(k) = \mu\sqrt{1+\lambda^2k^2}, \tag{4.24}$$

so that the action of A on the test function $\varphi(\mathbf{s})$ is defined as follows. First, we Fourier transform $\varphi(\mathbf{s})$ getting $\{\tilde{\varphi}_{\mathbf{k}}\}$. Then, $\forall\mathbf{k}\in\mathbb{Z}^d$, we multiply $\tilde{\varphi}_{\mathbf{k}}$ by the symbol $a(k)$. Finally, we perform the backward Fourier transform of $\{a(k)\tilde{\varphi}_k\}$, which gives us the function $(A\varphi)(\mathbf{s})$.

Thus, the action of the above fractional negated and shifted Laplacian on test functions in spectral space is well defined. Importantly, in physical space, the pseudo-differential operator A can be approximated by a discrete-in-space linear operator which is represented by a *very sparse matrix*, see Tsyrulnikov and Gayfulin (2016, Appendix B). So, in both spectral space and physical space, the resulting operator A with the fractional degree $q = 1/2$ is numerically tractable.

4.7 First-order model cannot satisfy the SPG requirements

Let us compute $\mathsf{Var}\,\xi$ using Eqs. (4.14) and (4.20). Since $b_{\mathbf{k}}$ is a smooth function of the wavevector $\mathbf{k}$, we may approximate the sum in Eq. (4.20) with the integral (where $b(\mathbf{k}) = b_{\mathbf{k}}$ for integer wavenumbers), getting

$$\mathsf{Var}\,\xi \propto \int_{\mathbb{R}^d} \frac{1}{\sqrt{1+\lambda^2 k^2}}\,\mathrm{d}\mathbf{k} \propto \int_{\mathbb{R}} \frac{k^{d-1}}{\sqrt{1+\lambda^2 k^2}}\,\mathrm{d}k. \tag{4.25}$$

To check the convergence of the latter integral in Eq. (4.25), we examine the $k \to \infty$ limit. For large k, the integrand is, obviously, proportional to k^{d-2}. As we know, the integral of this kind converges if the integrand decays faster than $k^{-1-\epsilon}$ with some $\epsilon > 0$. This implies that the integral in Eq. (4.25) diverges for all $d \geq 1$. In other words, the spectrum Eq. (4.14) decays too slowly for $\mathsf{Var}\,\xi$ to be finite.

Thus, the SPG model Eq. (4.4) cannot simultaneously satisfy the proportional-scales requirement 3 (which leads to $q = 1/2$) and the finite-variance requirement 3. Consequently, the SPG model is to be somehow changed. The solution is to increase the temporal order of the model.

5 Higher-order in time model

5.1 Formulation

The SPG model of higher temporal order is

$$\left(\frac{\partial}{\partial t} + \mu\sqrt{1-\lambda^2\Delta}\right)^p \xi(t,\mathbf{s}) = \alpha(t,\mathbf{s}), \tag{5.1}$$

where p is the temporal order of the modified SPG model (a positive integer). In spectral space, the model reads (cf. Section 4.2)

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \mu\sqrt{1+\lambda^2 k^2}\right)^p \tilde{\xi}_{\mathbf{k}}(t) = \sigma\,\Omega_{\mathbf{k}}(t). \tag{5.2}$$

In this section, we explore the steady-state statistics of $\xi(t,\mathbf{s})$ and find out which values of the temporal order p solve the above infinite variance problem.

5.2 Stationary spectral-space statistics

For each $\mathbf{k}$, Eq. (5.2) is a pth-order in time OSDE. Using Table 3 in Appendix A, we can write down the stationary variance $b_{\mathbf{k}}$ and the temporal correlation function $C_{\mathbf{k}}(t)$ of the solution to Eq. (5.2), the process $\tilde{\xi}_{\mathbf{k}}(t)$:

$$b_{\mathbf{k}} \propto \frac{\sigma^2}{\mu^{2p-1}(1+\lambda^2 k^2)^{p-\frac{1}{2}}} \tag{5.3}$$

(where the sign $\propto$ means proportional to) and

$$C_{\mathbf{k}}(t) = \left(1 + \frac{|t|}{\tau_{\mathbf{k}}} + r_2\frac{|t|^2}{\tau_{\mathbf{k}}^2} + \ldots + r_{p-1}\frac{|t|^{p-1}}{\tau_{\mathbf{k}}^{p-1}}\right)\mathrm{e}^{-\frac{|t|}{\tau_{\mathbf{k}}}}. \tag{5.4}$$

Here $r_2, \ldots, r_{p-1}$ are real numbers (given for $p = 1, 2, 3$ in Table 3) and $\tau_{\mathbf{k}}$ are still defined by Eq. (4.15). Specifically, for the temporal order $p = 3$, we have

$$b_{\mathbf{k}}|_{p=3} = \frac{3\sigma^2}{16\mu^5(1+\lambda^2 k^2)^{\frac{5}{2}}} \tag{5.5}$$

and

$$C_{\mathbf{k}}(t)|_{p=3} = \left(1 + \frac{|t|}{\tau_{\mathbf{k}}} + \frac{1}{3}\frac{|t|^2}{\tau_{\mathbf{k}}^2}\right)\mathrm{e}^{-\frac{|t|}{\tau_{\mathbf{k}}}}. \tag{5.6}$$

As Eq. (4.15) is unchanged in the higher-order model, the "proportionality of scales" condition Eq. (4.21) is still satisfied. In order to achieve the desired dependency of $\tau_{\mathbf{k}}$ not only on k (which we already have from Eq. (4.15)), but also on λ ($\tau_{\mathbf{k}}$ should increase with λ), we parameterize μ as

$$\mu = \frac{U}{\lambda}, \tag{5.7}$$

where $U > 0$ is the velocity-dimensioned tuning parameter. Note that λ affects both the spatial length scale of ξ (due to Eq. (5.3)) and the temporal length scale (thanks to Eq. (4.15)). In contrast, U affects only the temporal length scale.

5.3 Finite-variance criterion

Substituting b_k from Eq. (5.3) into Eq. (4.20), approximating the sum over the wavevectors by the integral, and exploiting the isotropy of the integrand yields

$$\mathsf{Var}\,\xi \approx \mathsf{const}\cdot\int_0^\infty \frac{\sigma^2}{(1+\lambda^2 k^2)^{p-\frac{1}{2}}}\,k^{d-1}\,\mathrm{d}k, \tag{5.8}$$

so that we have $\mathsf{Var}\,\xi < \infty$ (requirement 3) whenever

$$p > \frac{d+1}{2}. \tag{5.9}$$

5.4 Isotropy

In this section, we show that, remarkably, $q = 1/2$ is the unique spatial order for which the field $\xi(t,\mathbf{s})$ (with the appropriately scaled time coordinate) appears to be isotropic in space-time. In particular, the *shape* of the correlation function is the same in any direction in space-time.

5.4.1 Spatial isotropy

We note that the spatial isotropy of the random field ξ is the invariance of its covariance function $B(\mathbf{s})$ under rotations. If we were in $\mathbb{R}^d$ rather than on $\mathbb{T}^d$, isotropy of $B(\mathbf{s}) = B(s)$, where $s = |\mathbf{s}|$ is the spatial distance, would be equivalent to isotropy of its Fourier transform (spectrum) $b(\mathbf{k})$, so that the latter would be a function of the modulus of the wavevector, $k = |\mathbf{k}|$. On the torus, spectra are discrete, i.e. m, n, l take only *integer*

values, so, strictly speaking, $b_{\mathbf{k}}$ cannot be isotropic there. To avoid this technical difficulty, we employ (for the theoretical analysis only) the approximation of a sum over the wavevectors by an integral (see also sections 4.7 and 5.3).

Specifically, we assume that $b_{\mathbf{k}}$ is smooth enough (which is tantamount to the assumption that $B(\mathbf{s})$ decays on length scales much smaller than the domain's extent) for the validity of the approximation

$$B(\mathbf{s}) = \sum_{\mathbf{k}\in\mathbb{Z}^{\mathbf{d}}} b_{\mathbf{k}}\, \mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})} \approx \int_{\mathbb{R}^d} b(\mathbf{k})\, \mathrm{e}^{\mathrm{i}(\mathbf{k},\mathbf{s})} \mathrm{d}\mathbf{k}, \qquad (5.10)$$

where $b(\mathbf{k})$ is a smooth function of the real vector argument $\mathbf{k} \in \mathbb{R}^d$ such that $\forall \mathbf{k} \in \mathbb{Z}^d, b(\mathbf{k}) = b_{\mathbf{k}}$. The integral in Eq. (5.10) with the isotropic $b(\mathbf{k})$, see Eq. (5.3), can be easily shown to be invariant under rotations of $\mathbf{s}$. This implies that $B(\mathbf{s})$ and so the random field ξ are indeed approximately spatially isotropic.

In the theoretical analysis in this section, we will rely on the approximation Eq. (5.10) and thus assume that the "spectral grid" is dense enough for the spatial spectra to be treated as continuous.

5.4.2 Isotropy in space-time

Consider the OSDE Eq. (5.2) in the stationary regime. Following Yaglom (1987, Section 8), the stationary random process can be spectrally represented as the stochastic integral

$$\tilde{\xi}_{\mathbf{k}}(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega t}\, Z_{\mathbf{k}}(\mathrm{d}\omega), \qquad (5.11)$$

where ω is the angular frequency (temporal wavenumber) and Z is the orthogonal stochastic measure such that

$$\mathbb{E}\,|Z_{\mathbf{k}}(\mathrm{d}\omega)|^2 = b_{\mathbf{k}}(\omega)\,\mathrm{d}\omega, \qquad (5.12)$$

where $b_{\mathbf{k}}(\omega)$ is the spectral density of the process $\tilde{\xi}_{\mathbf{k}}(t)$ (i.e. the Fourier transform of its covariance function $b_{\mathbf{k}}C_{\mathbf{k}}(t)$, see Eq. (4.13)) and, at the same time, the spatio-temporal spectrum of the field ξ. In the spectral expansion of the driving white noise $\Omega_{\mathbf{k}}(t)$ (see Eq. (5.2)),

$$\Omega_{\mathbf{k}}(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega t}\, Z_{\Omega_{\mathbf{k}}}(\mathrm{d}\omega), \qquad (5.13)$$

we have $\mathbb{E}\,|Z_{\Omega_{\mathbf{k}}}(\mathrm{d}\omega)|^2 = \mathsf{const}\cdot \mathrm{d}\omega$ because the white noise has constant spectral density. Next, we substitute Eqs. (5.11) and (5.13) into Eq. (5.2), getting

$$(\mathrm{i}\omega + \mu\sqrt{1+\lambda^2 k^2})^p Z_{\mathbf{k}}(\mathrm{d}\omega) = Z_{\Omega_{\mathbf{k}}}(\mathrm{d}\omega). \qquad (5.14)$$

In this equation, taking expectation of the squared modulus of both sides, recalling that $\mu = U/\lambda$, and introducing the scaled angular frequency $\omega' = \omega/U$, we finally obtain

$$b_{\mathbf{k}}(\omega') \equiv b_{\mathbf{K}} \propto \frac{1}{(\lambda^{-2} + (\omega')^2 + k^2)^p} = \frac{1}{(\lambda^{-2} + \mathbf{K}^2)^p}, \qquad (5.15)$$

where

$$\mathbf{K} = \left(\frac{\omega}{U}, \mathbf{k}\right) \equiv \left(\frac{\omega}{U}, m, n, l\right) \qquad (5.16)$$

is the spatio-temporal wavevector.

From Eq. (5.15), one can see that with the scaled frequency (note that the change $\omega \to \omega/U$ corresponds to the change of the time coordinate $t \to t{\cdot}U$), the spatio-temporal spectrum $b_{\mathbf{k}}(\omega') \equiv b_{\mathbf{K}}$ becomes *isotropic in space-time*. This implies that the correlation function of ξ is isotropic in space-time as well (with the scaled time coordinate). Note that this remarkable property can be achieved only with the spatial order $q = 1/2$.

It is worth noting that the functional form of the spatio-temporal spectrum Eq. (5.15) together with the constraint Eq. (5.9) imply that the conditions of Theorem 3.4.3 in Adler (1981) are satisfied, so that spatio-temporal sample paths of the random field ξ are almost surely continuous, as we demanded in Section 3 by requirement 3.

5.5 Spatio-temporal covariances: the Matérn class

The spatio-temporal field satisfying the p-th order SPG model Eq. (5.1) has the spatio-temporal correlation function belonging to the so-called *Matérn* class of covariance functions (e.g. Stein, 1999; Guttorp and Gneiting, 2006). To see this, we denote

$$\nu = p - \frac{d+1}{2} > 0, \qquad (5.17)$$

where the positivity follows from Eq. (5.9). Then, Eq. (5.15) can be rewritten as

$$b_{\mathbf{K}} \propto \frac{1}{(\lambda^{-2} + \mathbf{K}^2)^{\nu + \frac{d+1}{2}}}. \qquad (5.18)$$

Note that here $d+1$ is the dimensionality of space-time. Equation (5.18) indeed presents the spectrum of the Matérn family of correlation functions, see, e.g., Eq. (32) in Stein (1999). The respective isotropic correlation function is given by the equation that precedes Eq. (32) in Stein (1999) or by Eq. (1) in Guttorp and Gneiting (2006):

$$B(r) \propto (r/\lambda)^{\nu} K_{\nu}(r/\lambda), \qquad (5.19)$$

where $r = \sqrt{s^2 + (Ut)^2}$ is the distance (in our case, the Euclidean distance in space-time with the coordinates (x, y, z, Ut)) and K_ν is the MacDonald function (the modified Bessel function of the second kind).

The Matérn family is often recommended for use in spatial analysis due to its notable flexibility with only two free parameters: ν and λ, see, e.g., Stein (1999) and Guttorp and Gneiting (2006). Specifically, λ controls the length scale, whereas $\nu > 0$ determines the degree of smoothness: the greater ν, the smoother the field. Note that the smoothness is understood as the number of the

Table 1: Spatio-temporal correlation functions $B(r)$ for some combinations of dimensionality d and temporal order p.

d	p	$\nu = p - \frac{d+1}{2}$	$B(r)$
2	2	1/2	$\exp(-\frac{r}{\lambda})$
2	3	3/2	$(1+\frac{r}{\lambda})\exp(-\frac{r}{\lambda})$
2	4	5/2	$\left(1+\frac{r}{\lambda}+\frac{1}{3}\left(\frac{r}{\lambda}\right)^2\right)\exp(-\frac{r}{\lambda})$
3	3	1	$\frac{r}{\lambda}K_1(\frac{r}{\lambda})$

mean-square derivatives of the random field in question. The degree of smoothness depends on the behavior of the correlation function at small distances and manifests itself in the field's realizations as the amount of small-scale noise (for illustration see, e.g., Tsyrulnikov and Gayfulin, 2016, Appendix D).

Table 1 lists the resulting correlation functions (in any direction in space-time) for several combinations of d and p (see Guttorp and Gneiting, 2006, for details).

With the fixed d, the larger p corresponds, according to Eq. (5.17), to the larger ν and so to the smoother in space and time field ξ. This can be used to change the degree of smoothness of the generated field by changing the temporal order of the SPG model.

From the constraint Eq. (5.9), the minimal temporal order p that can be used in both 2D and 3D is equal to 3. *This value $p = 3$ will be used by default in what follows and in the current SPG computer program.*

5.6 Spatio-temporal correlation functions: illustrations

Here, we show spatial, temporal, and spatio-temporal correlation functions computed using Eq. (5.19). To make the plots more accessible, it is arbitrarily assumed that the extent of the standardized spatial domain (the torus) in each dimension is 3000 km, so that the distance is measured in kilometers. The default SPG setup parameters are $\lambda = 125$ km and $U = 20$ m/s.

5.6.1 Spatial correlation functions

Fig. 4 presents the spatial correlation functions for different length scales in 2D and 3D. One can notice, first, that the actual length scale is well controlled by the parameter λ. Second, it is seen that in 2D (the left panel), where, according to Eq. (5.17), $\nu = 3/2$, the correlation functions are somewhat *smoother* at the origin than in 3D (the right panel), where $\nu = 1$. This is consistent with the above statement that the greater ν, the smoother the field. But in general, the 2D and 3D spatial correlation functions are quite similar.

5.6.2 Temporal correlation functions

Equation (5.19) implies that the spatial and temporal correlations have the same *shapes*. The latter feature is very appealing because atmospheric spectra are known to be similar in the spatial and in the temporal domain, e.g., the well-known "−5/3" spectral slope law is observed both in space and time, see, e.g., Monin and Yaglom (2013, Section 23). Thus, the SPG reproduces the observed in the atmosphere similarity of spatial and temporal spectra.

Fig. 5 shows the temporal correlation functions for different values of U. In comparing Fig. 5 with Fig. 4 (right), one can observe that the spatial and temporal correlations indeed have the same shape.

5.6.3 Spatio-temporal correlations

Fig. 6 presents the spatial correlation functions for different time lags. Fig. 7 displays the spatio-temporal correlation function. In both Fig. 6 and Fig. 7, another manifestation of the spatio-temporal "proportionality of scales" is seen: the larger the time lag, the broader the spatial correlations. Note that this is consistent with the behavior of the spatio-temporal covariances found by Cressie and Huang (1999, Figure8) in real-world wind speed data.

5.7 Introducing anisotropy in the vertical plane

We have formulated the SPG model under the 3D isotropy assumption. This implies that the ratio of the horizontal length scale to the horizontal domain size is the same as the ratio of the vertical length scale to the vertical domain size. This may be reasonable but, obviously, the independent specification of the horizontal and vertical length scales would be much more flexible. To get this capability, we can employ two equivalent modifications to the SPG model. One approach is to change the radius of the "vertical circle" in the torus from 1 to the δ^{-1}, where δ is a positive parameter. Another approach is to replace the Laplacian $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ by its anisotropic version $\Delta' = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \delta^2\frac{\partial^2}{\partial z^2}$. With both approaches, the vertical length scale increases by the factor of δ.

5.8 Preserving isotropy in the horizontal plane for non-square domains

If the size of the domain in physical space in the x direction, D_x, differs from the domain size in the y direction, D_y, then mapping from a square SPG domain to the rectangular physical domain would result in an elliptic (also called geometric) anisotropy in the horizontal plane. This undesirable feature can be avoided by replacing Δ' defined in Section 5.7 with

$$\Delta_* = \frac{\partial^2}{\partial x^2} + \gamma^2\frac{\partial^2}{\partial y^2} + \delta^2\frac{\partial^2}{\partial z^2}, \qquad (5.20)$$

where $\gamma = D_x/D_y$. The only change in all the above spectral equations is that the wavenumbers n and l are to

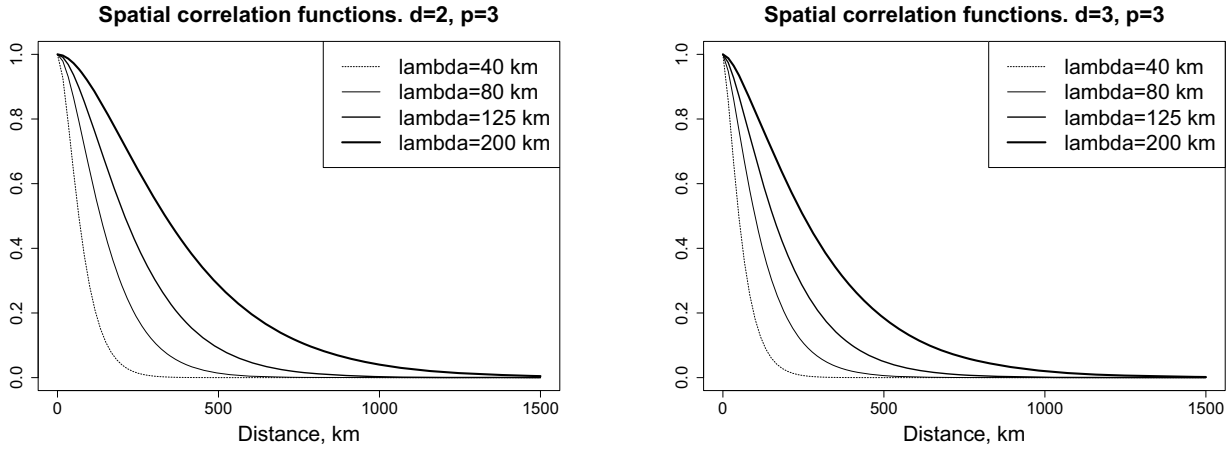

Figure 4: Spatial correlation functions for $p = 3$ in 2D (the left panel) and 3D (the right panel)—for the four spatial length scales indicated in the legend.

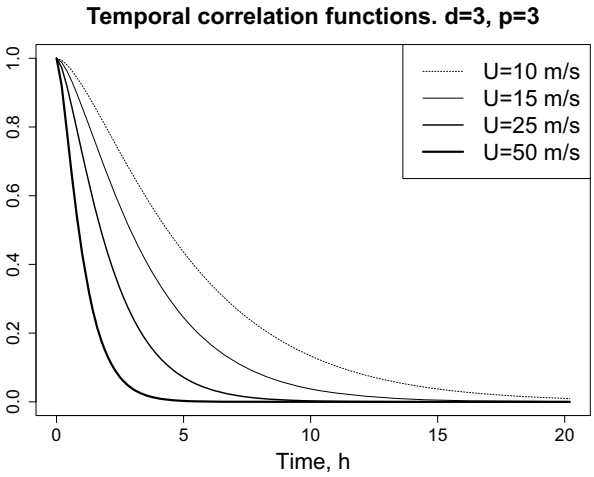

Figure 5: Temporal correlation functions in 3D for the four values of U indicated in the legend.

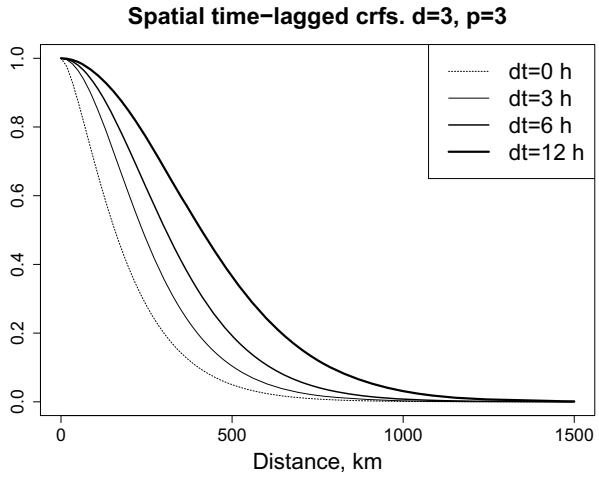

Figure 6: Spatial correlation functions in 3D for the four time lags indicated in the legend.

be multiplied by γ and δ, respectively. The total spatial wavenumber squared k^2 is to be replaced everywhere by its scaled version

$$k_*^2 = m^2 + (\gamma n)^2 + (\delta l)^2. \qquad (5.21)$$

For more technical details, see Tsyrulnikov and Gayfulin (2016). This device indeed allows the preservation of the horizontal isotropy and to change the vertical length scale over a broad range (not shown). We have refrained from introducing this feature to our basic SPG equations for the sake of simplicity of presentation.

6 A discrete-time solver for the third-order in time SPG model

In *physical space*, our final evolutionary model Eq. (5.1) with $p = 3$ can be discretized using the approximation of the operator $\sqrt{1 - \lambda^2\Delta}$ proposed in Tsyrulnikov and Gayfulin (2016, Appendix B). The respective physical-space solver looks feasible but we do *not* examine it in this study. Below, we present our basic *spectral-space* technique. From this point on, we will consider only the *spectral* SPG.

6.1 The spectral solver

To numerically integrate the SPG equations in spectral space, we discretize Eq. (5.2) (with $p = 3$) using an implicit scheme. The model operator $(\frac{\mathrm{d}}{\mathrm{d}t} + a_{\mathbf{k}})^3$, where $a_{\mathbf{k}} = \frac{U}{\lambda}\sqrt{1 + \lambda^2 k_*^2}$ and k_*^2 is defined in Eq. (5.21), is discretized by replacing the time derivative $\frac{\mathrm{d}}{\mathrm{d}t}$ with the backward finite difference $\frac{\mathcal{I}-\mathcal{B}}{\Delta t}$, where Δt is the time step, $\mathcal{I}$ is the identity operator, and $\mathcal{B}$ is the backshift operator. The white noise in the r.h.s. of Eq. (5.2) is discretized using Eq. (4.9), where $\mathrm{d}W_{\mathbf{k}}(t)$ is replaced with $\Delta W_{\mathbf{k}}(t) = W_{\mathbf{k}}(t + \Delta t) - W_{\mathbf{k}}(t)$, and simulated as a zero-mean Gaussian random variable with the variance Δt. As

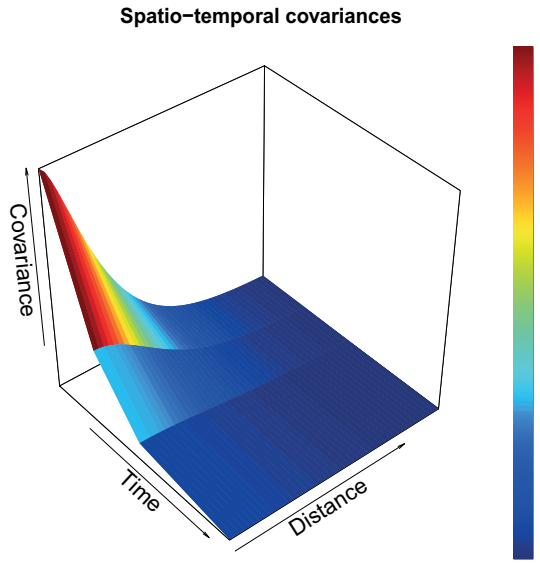

Figure 7: SPG spatio-temporal covariances.

a result, we obtain the discrete-time evolution equation

$$\hat{\xi}_{\mathbf{k}}(i) = \frac{1}{\varkappa^3}\cdot \left[3\varkappa^2\hat{\xi}_{\mathbf{k}}(i-1) - 3\varkappa\hat{\xi}_{\mathbf{k}}(i-2) + \hat{\xi}_{\mathbf{k}}(i-3) + \sigma\Delta t^{\frac{5}{2}}\zeta_{\mathbf{k}t}\right], \tag{6.1}$$

where $i = 0, 1, 2, \ldots$ denotes the discrete time instance, $\varkappa = 1 + a_{\mathbf{k}}\Delta t$, and $\zeta_{\mathbf{k}t} \sim CN(0, 1)$ are independent complex standard Gaussian pseudo-random variables (for their definition, see, e.g., Tsyrulnikov and Gayfulin, 2016, Appendix A.6). Note that the solution of the time-discrete Eq. (6.1) is denoted by the hat, $\hat{\xi}_{\mathbf{k}}(i)$, in order to distinguish it from the solution of the time-continuous Eq. (5.2), which is denoted by the tilde, $\tilde{\xi}_{\mathbf{k}}(t)$.

It can be shown that the numerical stability of the scheme Eq. (6.1) is guaranteed whenever $\varkappa > 1$, which is always the case because $a_{\mathbf{k}} > 0$ (see Eq. (4.24)).

Note that the derivation of the numerical scheme for a higher-order (i.e. with $p > 3$) SPG model is straightforward: one should just raise the difference operator $\frac{\mathcal{I}-\mathcal{B}}{\Delta t}$ to a power higher than 3.

6.2 Correction of spectral variances

Because of discretization errors, the discrete-time scheme Eq. (6.1) gives rise to the steady-state spectral variances $\hat{b}_{\mathbf{k}} = \mathsf{Var}\,\hat{\xi}_{\mathbf{k}}(i)$ that are different from the "theoretical" ones, $b_{\mathbf{k}}$ (given in Eq. (5.5)). The idea is to correct (multiply by a number) the solution $\hat{\xi}_{\mathbf{k}}(i)$ to Eq. (6.1) so that the steady-state variance of the corrected $\hat{\xi}_{\mathbf{k}}(i)$ is made equal to $b_{\mathbf{k}}$. To this end, we derive $\hat{b}_{\mathbf{k}}$ from Eq. (6.1) (using Eq. (B.2) in Appendix B) and then, knowing the "theoretical" $b_{\mathbf{k}}$, we introduce the correction coefficients, $\sqrt{b_{\mathbf{k}}/\hat{b}_{\mathbf{k}}}$, to be applied to $\hat{\xi}_{\mathbf{k}}(i)$. This simple device ensures that for any time step, the spatial spectrum and thus the *spatial* covariances are perfect. But the *temporal* correlations do depend on the time step; this aspect is discussed below in Section 6.4.1.

6.3 "Warm start": ensuring stationarity from the beginning of the time integration

To start the numerical integration of the third-order scheme Eq. (6.1), we obviously need three initial conditions. If the integration is the continuation of a previous run, then we just take values of $\hat{\xi}_{\mathbf{k}}(i)$ at the last three time instances i from that previous run; this ensures the continuity of the resulting trajectory. If we start a new integration, we have to somehow generate values of $\hat{\xi}_{\mathbf{k}}(i)$ at $i = 1, 2, 3$. Let us denote them here as the vector $\boldsymbol{\xi}^{\mathrm{ini}} = (\hat{\xi}_{\mathbf{k}}(1), \hat{\xi}_{\mathbf{k}}(2), \hat{\xi}_{\mathbf{k}}(3))^{\top}$. Simplistic choices like specifying zero initial conditions give rise to a substantial initial transient period, which distorts the statistics of the generated field in the short time range.

In order to have the steady-state regime right from the beginning of the time integration and thus avoid the initial transient period completely, we simulate $\boldsymbol{\xi}^{\mathrm{ini}}$ as a pseudo-random draw from the multivariate Gaussian distribution with zero mean and the steady-state covariance matrix of $\hat{\xi}_{\mathbf{k}}(i)$. In Appendix B, we derive the components of this 3×3 matrix, namely, its diagonal elements (all equal to the steady-state variance), see Eq. (B.2), and the lag-1 and lag-2 covariances, see Eq. (B.3).

6.4 Computational efficiency

In this subsection, we describe two techniques that allow us to significantly decrease the computational cost of running the spectral SPG.

6.4.1 Making the time step Δt dependent on the spatial wavevector k

For an ordinary differential equation, the accuracy of a finite-difference scheme depends on the time step. More precisely, it depends on the ratio of the time step Δt to the temporal length scale τ of the process in question. For high accuracy, $\Delta t \ll \tau$ is needed.

In our problem, $\tau_{\mathbf{k}}$ decays with the total scaled wavenumber k_*, see Eqs. (4.15), (4.21), and (5.21). This implies that for higher k_*, smaller time steps are needed. To maintain the accuracy across the wavenumber spectrum, we choose the time step to be a fraction of the time scale:

$$(\Delta t_{\mathbf{k}}) = \beta\tau_{\mathbf{k}}. \tag{6.2}$$

For decreasing β, the numerical integration scheme becomes more accurate and, at the same time, more time consuming.

We note that in atmospheric spectra, small scales have, normally, much less variance (energy) than large

scales. But with the constant β, the computational time would be, on the contrary, spent predominantly on high wavenumbers (because the latter require a smaller time step and are much more abundant in 3D or 2D). Thus, to save computer time whilst ensuring reasonable overall (i.e. for the whole range of wavenumbers) accuracy, we specify β to be wavenumber dependent (growing with the wavenumber) in the following ad-hoc way:

$$\beta_{\mathbf{k}} = \beta_{\min} + (\beta_{\max} - \beta_{\min}) \left(\frac{k_*}{\max k_*} \right)^2 , \qquad (6.3)$$

where $\beta_{\min}$ and $\beta_{\max}$ are the tunable parameters. The choice of the "optimal" $\beta_{\min}$ and $\beta_{\max}$ is discussed below in Section 6.4.2.

6.4.2 Introduction of a coarse grid in spectral space

Here we propose another technique to reduce the computational cost of the spectral solver. The technique exploits the *smoothness* of the SPG spectrum $b_{\mathbf{k}}$ Eq. (5.5). This smoothness allows us to introduce a *coarse grid in spectral space* and save a lot of computer time by performing the integration of the discrete-time spectral OSDEs Eq. (6.1) only for those wavevectors that belong to the coarse grid. The spectral coefficients $\hat{\xi}_{\mathbf{k}}(i)$ are then interpolated from the coarse grid to the dense (full) grid in spectral space.

The latter interpolation would introduce correlations between different spectral coefficients $\hat{\xi}_{\mathbf{k}}(i)$, which would destroy the spatial homogeneity. In order to avoid this, we employ a device used to generate so-called surrogate time series (Theiler et al., 1992, Section 2.4.1). At each t, we multiply the interpolated (i.e. dense-grid) $\hat{\xi}_{\mathbf{k}}(i)$ by $e^{i\theta_{\mathbf{k}}}$, where $\theta_{\mathbf{k}}$ are independent *random phases*, i.e. independent for different $\mathbf{k}$ random variables uniformly distributed on the segment $[0, 2\pi]$. It can be easily seen that this multiplication removes any correlation between the spectral coefficients.

Note also that the random phase rotation does not destroy the Gaussianity because $\hat{\xi}_{\mathbf{k}}(i)$ are complex circularly-symmetric random variables with uniformly distributed and independent of $|\hat{\xi}_{\mathbf{k}}(i)|$ arguments (phases) (e.g. Tse and Viswanath, 2005, Section A.1.3).

In order to preserve the temporal correlations, we keep the set of $\theta_{\mathbf{k}}$ constant during the SPG-model time integration.

The exact spectrum $b_{\mathbf{k}}$ after the trilinear (bilinear in 2D) interpolation of $\hat{\xi}_{\mathbf{k}}(i)$ from the coarse to the full spectral grid is imposed in a way similar to that described in Section 6.2 (for details, see Tsyrulnikov and Gayfulin, 2016).

The coarse grid is specified as the direct product of 1D grids. Each of the (non-uniform) 1D coarse grids is specified as follows. Its jth point is located at the fine-grid wavenumber n_j, which equals j for $|j| \le n_0$ (where n_0 is an integer) and equals the closest integer to $n_0(1+\varepsilon)^{|j|-n_0}$ for $|j| > n_0$. Here, ε is a tunable small positive number. In the following numerical experiments, the coarse-grid parameters were $n_0 = 20$ and $\varepsilon = 0.2$, which resulted in the following positive 1D coarse-grid points: 0 1 2 3 … 19 20 24 29 35 42 50 60 72 86 103 124 150 (the 1D grid extent was 300 points and, correspondingly, the maximal wavenumber was 150). At the time of writing, the coarse spectral grid was introduced only in the horizontal.

Table 2: CPU times (seconds) of the 2D SPG computations per 1 h of SPG model time and the relative error in the temporal length scale $T_{0.5}$. Spectral stands for spectral-space, interpol. means interpolation from the sparse spectral grid, and FFT is the fast Fourier transform.

Accelerators	CPU spectral	CPU interpol.	CPU FFT	Speedup	Rel. err. $T_{0.5}$
NO	0.66	0	0.027	1	3 %
YES	0.010	0.012	0.027	14	4 %

6.4.3 Numerical acceleration: results

As the two above acceleration techniques guarantee that the *spatial* spectrum is always precise, we tested how these techniques impacted the *temporal* correlations and what was the speedup. We performed a numerical experiment with the 2D SPG on the grid with 300×300 points, the mesh size $h = 7$ km, and the setup parameters $\lambda = 80$ km, $U = 10$ m/s, and $\delta = \gamma = 1$. The time interval Δt_{FFT} between the successive backward Fourier transforms determines the effective resolution of the generated field in time. To make the temporal resolution consistent with the spatial resolution, we selected Δt_{FFT} close to h/U, namely, $\Delta t_{\mathrm{FFT}} = 15$ min. The computations were performed on a single CPU.

The results are presented in Table 2. We compared the non-accelerated scheme with the constant $\beta = 0.1$ and without the sparse spectral grid (the second row) and the accelerated scheme with $\beta_{\min} = 0.15$, $\beta_{\max} = 3$, and with the sparse spectral grid (the third row). From column 2, it is seen that the combined effect of the two numerical acceleration techniques on the cost of the *spectral-space* computations (see column 2) was dramatic: the speedup was as large as 66 times. The contributions of the two acceleration techniques to the spectral-space speedup were comparable in magnitude (not shown). Most importantly, this spectral-space speedup was achieved at very little cost: the temporal length scale $T_{0.5}$ (defined as the time shift at which the correlation function first intersects the 0.5 level) was distorted by only 4 % with respect to the theoretical model (column 6). Note, however, that the cost of the interpolation from the sparse spectral grid (column 3) and of the discrete backward Fourier transform (column 4) reduced the total speedup of the 2D SPG to 14 times (see column 5).

In 3D, the SPG operating on the spatial grid with $300 \times 300 \times 64$ points, took 40–70 times more CPU time as compared with the above 2D case, with the accuracy being similar to that indicated in Table 2 (not shown).

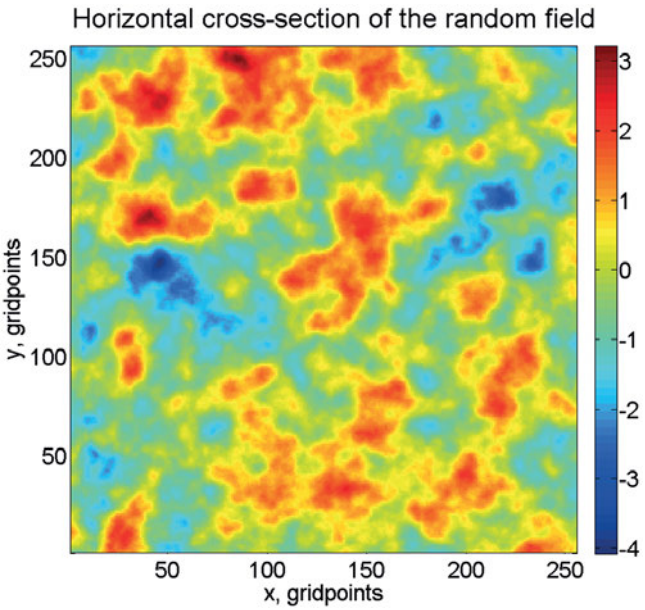

Figure 8: Horizontal (x-y) cross-section of an SPG field.

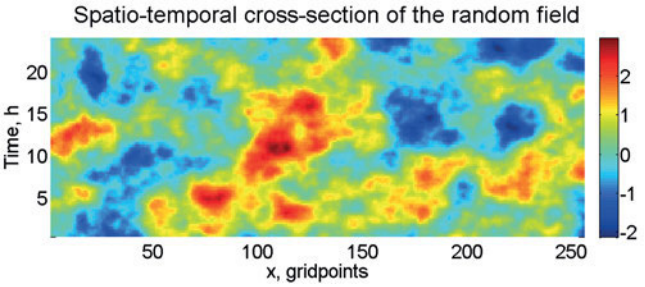

Figure 9: Spatio-temporal (x-t) cross-section of an SPG field.

The total speedup was only 8 times due to an increased share of the Fourier transform.

6.5 Examples of the SPG fields

Fig. 8 shows a horizontal x-y cross-section and Fig. 9 a spatio-temporal x-t cross-section of the pseudo-random field $\xi(t, x, y)$ simulated by the SPG with the setup parameters indicated in Section 6.4.3. Note that with 300 grid points in each spatial direction, only 256 contiguous grid points are shown in the Figs. 8 and 9 and are intended to be used in a mapping to a physical-space domain. This is done in the SPG for practical purposes in order to avoid correlations between the opposite sides of the spatial domain, which would be spurious in real-world applications.

7 Application to the COSMO model

COSMO is a limited-area meteorological forecast model developed by the European COnsortium for Small Scale MOdelling (Baldauf et al., 2011). The model solves the non-hydrostatic primitive equations for the moist compressible atmosphere. The model equations are formulated in rotated spherical coordinates in the horizontal and a generalized terrain-following geometric-height based vertical coordinate. Currently, the COSMO model is used operationally in many meteorological services across Europe and beyond with horizontal grid spacings down to 1 km, see www.cosmo-model.org. In our experiments, the COSMO domain was an area covering most of Europe with the grid having $n_x \times n_y$ grid points ($n_x = 700$ and $n_y = 620$) in the horizontal, $n_z = 40$ vertical levels, the horizontal grid spacing 7 km, and the time step 66 s.

The SPG was embedded into the Fortran code of the COSMO model and parallelized using MPI. In our experiments, the SPG's 3D grid was $270 \times 270 \times 32$ points. The Fourier transforms were performed every `stride`-th COSMO time step, where `stride` was selected to be equal to 14 for the same reason as described in Section 6.4.3. The resulting temporal resolution was, correspondingly, $\Delta t_{\mathrm{FFT}} = 14 * 66 = 924$ s (i.e. about 15 min).

The SPG was used to generate additive flow independent (that is, purely stochastic) model-error perturbations to the 3D model fields u, v, T, and p every model time step. The u, v, and T field perturbations were mutually probabilistically independent. Pressure perturbations were computed from the temperature perturbation using the hydrostatic equation with the zero pressure perturbation at the model's top level. Technically, every `stride`-th model time step, the SPG physical space fields ΔT, Δu, Δv, and Δp were computed. Then, they were divided by `stride` and the resulting fields ΔT/`stride`, Δu/`stride` etc. were added to the model fields at each of the next `stride` model time steps.

The model's initial and boundary conditions were unperturbed in order to isolate the effect of the imposed model error perturbations. The magnitudes and the length scales of the SPG perturbations were specified using the following heuristics.

Results reported by Raynaud et al. (2012, Figure 3) imply that the 6-h rms forecast errors in their global data assimilation system (evaluated by comparison with observations) were, in the mid-troposphere, about 0.5 K for temperature and 1.5 m/s for each of the two mutually orthogonal horizontal wind components. Model errors were represented in that system by inflating the forecast ensemble spread. The tuned inflation coefficient was 1.1, i.e. the the model error contribution was 0.1 of the forecast ensemble spread. Consequently, the contribution of model errors over the period of 6 h can be assessed as $0.5 \cdot 0.1 = 0.05$ K and $1.5 \cdot 0.1 = 0.15$ m/s, respectively. For the period of 66 s (one COSMO model time step), this amounts to $\frac{0.05 \cdot 66}{3600 \cdot 6} = 1.5 \cdot 10^{-4}$ K for temperature and $5 \cdot 10^{-4}$ m/s for each wind component. These were the standard deviations of the SPG model error perturbation fields we imposed.

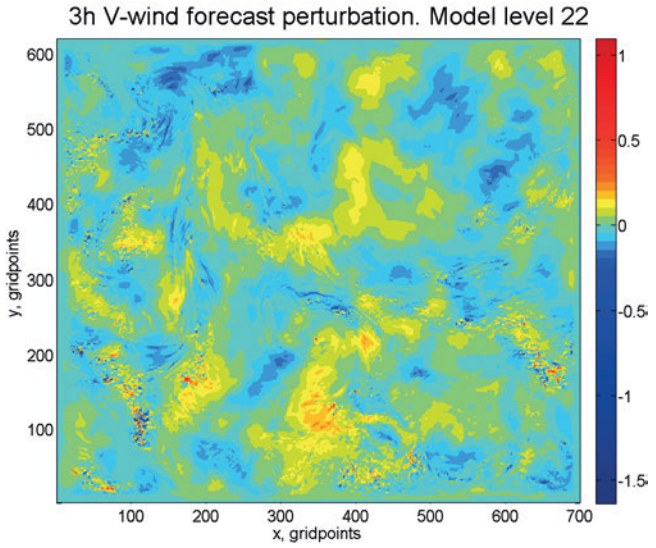

Figure 10: COSMO 3-h forecast V-wind perturbation field in response to additive model-error SPG perturbations of temperature, pressure, and both wind components.

The spatial SPG length scale was chosen to be representative of meso-scale structures, $\lambda = 100$ km. The characteristic flow velocity was $U = 10$ m/s and $\delta = \gamma = 1$. As a remark, we did not try to geostrophic balance the model error wind and mass perturbations because of their very small magnitude.

As an example, Fig. 10 displays the forecast meridional wind perturbation field (computed as the difference between the perturbed and unperturbed forecasts starting 0000 UTC 28 April 2016) at the model level 22 (about 3 km above the ground) after 3 h of model integration. Comparing the forecast perturbation field in Fig. 10 with a model error perturbation field (which looks very much like that in Fig. 8, not shown) suggests that even flow independent additive model error perturbations can give rise to a flow dependent forecast perturbation (as desired in ensemble prediction and ensemble data assimilation). In particular, in the forecast perturbation field, there are many very localized features (seen mostly in the southern part of the domain as almost point-wise disturbances). Examining their structure in the vertical revealed that these were convection plumes (not shown), so the model error perturbations were capable of triggering convection.

The rms v-wind forecast perturbation magnitude was 0.06 m/s. The median of the absolute deviation of the forecast perturbation from zero was 0.04 m/s, which means that the outliers at the grid points with the convective plumes were not too numerous. The actual magnitude of the 3-h forecast perturbation 0.04–0.06 m/s can be compared with the target magnitude 0.15 m/s for the 6-h forecast perturbation (see above in this section), i.e. roughly, 0.075 m/s for the 3-h perturbation. The actual magnitude is seen to be somewhat less than the target magnitude; this looks reasonable because the growth of the model error *drift* (see Section 2) due to a *stochastic* forcing should be slower than linear in time (e.g. with the white-noise forcing, the growth should follow the $\sqrt{t}$ law).

Finally, we note that at the time of writing, the numerical acceleration techniques described in Section 6.4 were implemented only in the stand-alone version of the SPG. Without those accelerators, the cost of running the SPG within COSMO was about 0.8 % of the total COSMO model wall-clock time per generated SPG field.

8 Conclusions

- The proposed Stochastic Pattern Generator (SPG) produces pseudo-random spatio-temporal Gaussian fields on 2D and 3D limited area spatial domains with the tunable variance, horizontal, vertical, and temporal length scales.
- The SPG model is defined on a standardized domain in space, specifically, on the unit 2D or 3D cube with periodic boundary conditions. Fields on a limited-area geophysical domain in question are obtained by mapping from the standardized domain.
- The SPG is based on a linear third-order in time stochastic model driven by the white in space and time Gaussian noise.
- The spatial operator of the stochastic model is built to ensure that solutions to the SPG model, i.e. the generated pseudo-random fields, satisfy the "proportionality of scales" property: large-scale (small-scale) in space field components have large (small) temporal length scales.
- Beyond the "proportionality of scales", the generated fields possess a number of other nice properties:
  - The spatio-temporal realizations are (almost surely) continuous.
  - With the appropriately scaled temporal and vertical coordinates, the fields are isotropic in space-time (i.e. having the same shape of the correlation function in any direction in space-time).
  - The correlation functions belong to the Matérn class.
- The SPG numerical solver is spectral-space based.
- Two techniques to accelerate the spectral-space computations are proposed and implemented. The first technique selects the time step of the spectral-space numerical integration scheme to be dependent on the wavenumber, so that the discretization error is smaller for more energetic larger spatial scales and is allowed to be larger for less energetic smaller scales. The second technique introduces a coarse grid in spectral space. The resulting speedup of the spectral-space computations from both techniques combined is as large as 40–70. The total speedup (that includes the fast Fourier transform) is 8–14.
- The SPG is embedded in the meteorological limited area forecast model COSMO and tested as a source of additive model-error perturbations.

- Potential applications of the SPG include ensemble prediction and ensemble data assimilation in meteorology, oceanography, hydrology, and other areas. The SPG can be used to generate spatio-temporal perturbations of the model fields (in the additive or multiplicative or other mode) and of the boundary conditions.
- A more detailed exposition of some aspects of the SPG theory and design can be found in Tsyrulnikov and Gayfulin (2016).

Acknowledgements

The SPG has been developed as part of the Priority Project KENDA (Kilometre scale Ensemble Data Assimilation) of COSMO. We have used the discrete fast Fourier package `fft991` developed by C. Temperton at ECMWF in 1978. Christoph Schraff and Michael Baldauf kindly helped us to interpret the appearance of the small-scale noise in the COSMO forecast perturbation fields. We would like to thank Denis Blinov for his help with the COSMO model technology. The authors are grateful to the two anonymous reviewers for their valuable comments on the manuscript.

Appendices

A Stationary statistics of a higher-order OSDE

We examine the OSDE Eq. (5.2) in its generic form:

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)^p \eta(t) = \sigma\Omega(t), \tag{A.1}$$

where $\eta(t)$ is the random process in question, σ and a are the positive numbers, p is the positive integer, and $\Omega(t)$ is the standard white noise, see, e.g., Rozanov (1982, Section 1.1.3) or Kuo (2001, Section 3.1.4) and also Tsyrulnikov and Gayfulin (2016, Appendix A).

The goal here is to find the variance and the correlation function of $\eta(t)$ in the stationary regime. The technique is to reduce the p-th order OSDE to a system of first-order OSDEs.

To simplify the exposition, we consider the third-order OSDE ($p = 3$) and rewrite Eq. (A.1) as

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)\left\{\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)\left[\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)\eta(t)\right]\right\} = \sigma\Omega(t). \tag{A.2}$$

Here, by η_1 we denote the term in brackets,

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)\eta = \eta_1 \tag{A.3}$$

and by η_2 the term in braces,

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)\eta_1 = \eta_2, \tag{A.4}$$

so that Eq. (A.2) implies that

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)\eta_2 = \sigma\Omega. \tag{A.5}$$

In Eqs. (A.3)–(A.5), the last equation is the familiar first-order OSDE forced by the white noise, whereas the other equations are not forced by the white noise. Generalizing the above construction, Eqs. (A.2)–(A.5), to the arbitrary $p > 0$, we form the following first-order vector-matrix OSDE (a system of first-order OSDEs):

$$\mathrm{d}\boldsymbol{\eta} + \mathbf{A}\boldsymbol{\eta}\mathrm{d}t = \boldsymbol{\Sigma}\boldsymbol{\Omega}\mathrm{d}t, \tag{A.6}$$

where $\boldsymbol{\eta} = (\eta, \eta_1, \ldots, \eta_{p-2}, \eta_{p-1})$, $\boldsymbol{\Omega} = (0, 0, \ldots, 0, \Omega)$, and the design of the matrices $\mathbf{A}$ and $\boldsymbol{\Sigma}$ is obvious (not shown).

With Eq. (A.6) in hand, we derive a differential equation for the covariance matrix $\mathbf{P} = \mathbb{E}\,\boldsymbol{\eta}\boldsymbol{\eta}^*$, where * denotes transpose complex conjugate (e.g. Jazwinski, 1970, example 4.16). First, we compute the increment of $\mathbf{P}$:

$$\Delta\mathbf{P} = \mathbb{E}(\boldsymbol{\eta}+\mathrm{d}\boldsymbol{\eta})(\boldsymbol{\eta}+\mathrm{d}\boldsymbol{\eta})^* - \mathbb{E}\,\boldsymbol{\eta}\boldsymbol{\eta}^* = \mathbb{E}\,\boldsymbol{\eta}\mathrm{d}\boldsymbol{\eta}^* + \mathbb{E}\,\mathrm{d}\boldsymbol{\eta}\boldsymbol{\eta}^* + \mathbb{E}\,\mathrm{d}\boldsymbol{\eta}\mathrm{d}\boldsymbol{\eta}^*. \tag{A.7}$$

Then, using Eq. (A.6) and the fact that $\mathbb{E}\,|\Omega\mathrm{d}t|^2 = \mathbb{E}\,|\mathrm{d}W|^2 = \mathrm{d}t$, we obtain the differential of $\mathbf{P}$ from Eq. (A.7):

$$\mathrm{d}\mathbf{P} = -\mathbf{A}\mathbf{P}\mathrm{d}t - \mathbf{P}\mathbf{A}^*\mathrm{d}t + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^*\mathrm{d}t. \tag{A.8}$$

In the stationary regime $\mathrm{d}\mathbf{P} = 0$, so the equation for the stationary covariance matrix is

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^* = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^*. \tag{A.9}$$

This a system of linear algebraic equations for the unknown entries of the matrix $\mathbf{P}$. Because both $\mathbf{P}$ and $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^*$ are self-adjoint matrices, the number of unknowns, $p(p+1)/2$, is equal to the number of independent equations. We analytically solve this system of equations and look at the *first* diagonal entry of the solution $\mathbf{P}$, which represents the required $\mathsf{Var}\,\eta$ (because the random field in question η is defined above to be the first entry of the vector $\boldsymbol{\eta}$). Dropping tedious derivations, we present in Table 3 (the second row) the formulas for the temporal orders $p = 1$, $p = 2$, $p = 3$, and for the general p.

Finally, we derive the temporal correlation function for the pth-order OSDE. To this end, we multiply Eq. (A.1) by $\eta(s)$ with $s < t$ and take the expectation. Since a is non-stochastic, we may interchange the expectation and the differential operator $\left(\frac{\mathrm{d}}{\mathrm{d}t}+a\right)^p$, getting the pth-order ordinary differential equation for the temporal covariance function, whose solutions for different p are presented in row 3 of Table 3.

Table 3: Variances $\mathsf{Var}\,\eta$ and correlation functions $C_{\eta(t)}$ of the stationary solution to Eq. (A.1) for different temporal orders p.

p	1	2	3	Arbitrary p
$\mathsf{Var}\,\eta$	$\sigma^2/(2a)$	$\sigma^2/(4a^3)$	$3\sigma^2/(16a^5)$	$\sigma^2/(a^{2p-1})$
$C_{\eta(t)}$	$\mathrm{e}^{-a\lvert t\rvert}$	$(1+a\lvert t\rvert)\,\mathrm{e}^{-a\lvert t\rvert}$	$(1+a\lvert t\rvert+a^2t^2/3)\,\mathrm{e}^{-a\lvert t\rvert}$	$R_{p-1}(a\lvert t\rvert)\cdot\mathrm{e}^{-a\lvert t\rvert}$

Note that $R_{p-1}(.)$ in the last row stands for the polynomial of order $p-1$.

B Stationary statistics of a discrete-time higher-order OSDE

Consider the continuous-time OSDE, Eq. (A.1), with $p=3$. The implicit scheme Eq. (6.1) we use to numerically solve it is reproduced here as

$$\eta_i = \frac{1}{\varkappa^3}\left[3\varkappa^2\eta_{i-1} - 3\varkappa\eta_{i-2} + \eta_{i-3} + \sigma(\Delta t)^2\,\Delta W_i\right], \tag{B.1}$$

where $\varkappa = 1 + a\Delta t$. Here, the goal is to find the stationary variance $V = \lim_{i\to\infty}\mathsf{Var}\,\eta_i$ along with lag-1 and lag-2 stationary covariances, $c_1 = \lim_{i\to\infty}\mathbb{E}\,\eta_i\eta_{i-1}$ and $c_2 = \lim_{i\to\infty}\mathbb{E}\,\eta_i\eta_{i-2}$, respectively. To reach this goal, we build three linear algebraic equations for the three unknowns, V, c_1, and c_2. The first equation is obtained by applying the variance operator to both sides of Eq. (B.1). The second and third equations are obtained by multiplying Eq. (B.1) by η_{i-1} and η_{i-2}, respectively, and applying the expectation operator to both sides of the resulting equations. Omitting the derivations, we write down the results:

$$V = \frac{\varkappa^4 + 4\varkappa^2 + 1}{(\varkappa^2-1)^5}\,(\Delta t)^5\sigma^2. \tag{B.2}$$

$$c_1 = \frac{3\varkappa(\varkappa^2+1)}{(\varkappa^2-1)^5}\,(\Delta t)^5\sigma^2, \quad c_2 = \frac{6\varkappa^2}{(\varkappa^2-1)^5}\,(\Delta t)^5\sigma^2. \tag{B.3}$$

One can verify that as $\Delta t \to 0$, V tends to the continuous-time variance $\frac{3}{16}\frac{\sigma^2}{a^5}$, see Table 3.

References

Åberg, S., K. Podgórski, 2011: A class of non-Gaussian second order random fields. – Extremes **14**, 187–222.

Adler, R.J., 1981: The geometry of random fields – Wiley.

Arnold, L., 1974: Stochastic differential equations – Wiley.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. – Mon. Wea. Rev. **139**, 3887–3905.

Bengtsson, L., M. Steinheimer, P. Bechtold, J.-F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. – Quart. J. Roy. Meteor. Soc. **139**, 1533–1543.

Berner, J., G. Shutts, M. Leutbecher, T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. – J. Atmos. Sci. **66**, 603–626.

Berner, J., S.-Y. Ha, J. Hacker, A. Fournier, C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. – Mon. Wea. Rev. **139**, 1972–1995.

Berner, J., U. Achatz, L. Batté, L. Bengtsson, A. de la Cámara, H.M. Christensen, M. Colangeli, D.R.B. Coleman, D.C. Stamen, I. Dolaptchiev, C.L.E. Franzke, P. Friederichs, P. Imkeller, H. Järvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T.N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Weniger, P.D. Williams, J.-I. Yano, 2017: Stochastic parameterization: Towards a new view of weather and climate models. – Bull. Amer. Meteor. Soc. **98**, 565–587.

Bouttier, F., B. Vié, O. Nuissier, L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. – Mon. Wea. Rev. **140**, 3706–3721.

Buizza, R., M. Miller, T. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. – Quart. J. Roy. Meteor. Soc. **125**, 2887–2908.

Charron, M., G. Pellerin, L. Spacek, P. Houtekamer, N. Gagnon, H.L. Mitchell, L. Michelin, 2010: Toward random sampling of model error in the Canadian ensemble prediction system. – Mon. Wea. Rev. **138**, 1877–1901.

Christensen, H., I. Moroz, T. Palmer, 2015: Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. – J. Atmos. Sci. **72**, 2525–2544.

Cressie, N., H.-C. Huang, 1999: Classes of nonseparable, spatio-temporal stationary covariance functions. – J. Amer. Statist. Assoc. **94**, 1330–1339.

Epstein, E.S., 1969: Stochastic dynamic prediction. – Tellus **21**, 739–759.

Gneiting, T., M.G. Genton, P. Guttorp, 2006: Geostatistical space-time models, stationarity, separability, and full symmetry. – Monographs Statist. Appl. Probab. **107**, 151 pp.

Guttorp, P., T. Gneiting, 2006: Studies in the history of probability and statistics XLIX. On the Matern correlation family. – Biometrika **93**, 989–995.

Jazwinski, A., 1970: Stochastic processes and filtering theory – Academic Press, 376 pp.

Kuo, H.-H., 2001: White noise theory. – In: Kannan D., V. Lakshmikantham (Eds.): Handbook of stochastic analysis and applications, 107–158.

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

Meunier, N., J. Zhao, 2009: Observations of photospheric dynamics and magnetic fields: from large-scale to small-scale flows. – Space Sci. Rev. **144**, 127–149.

Monin, A., A.M. Yaglom, 2013: Statistical fluid mechanics, Volume II: Mechanics of turbulence – Courier Corp.

Ollinaho, P., S.-J. Lock, M. Leutbecher, P. Bechtold, A. Beljaars, A. Bozzo, R.M. Forbes, T. Haiden, R.J. Hogan, I. Sandu, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble. – Quart. J. Roy. Meteor. Soc. **143**, 408–422.

Orrell, D., L. Smith, J. Barkmeijer, T. Palmer, 2001: Model error in weather forecasting. – Nonlin. Proc. Geophys. **8**, 357–371.

Palmer, T., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. – ECMWF Tech. Memo. **598**, ECMWF, Shinfield Park, 42 pp.

Pitcher, E.J., 1977: Application of stochastic dynamic prediction to real data. – J. Atmos. Sci. **34**, 3–21.

Raynaud, L., L. Berre, G. Desroziers, 2012: Accounting for model error in the Météo-France ensemble data assimilation system. – Quart. J. Roy. Meteor. Soc. **138**, 249–262.

Rozanov, Y.A., 1982: Markov random fields – Springer.

Shubin, M.A., 1987: Pseudodifferential operators and spectral theory – Springer.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. – Quart. J. Roy. Meteor. Soc. **131**, 3079–3102.

Stein, M.L., 1999: Interpolation of spatial data: some theory for kriging – Springer, New York.

Stein, M.L., 2005: Space–time covariance functions. – J. Amer. Statist. Assoc. **100**(469), 310–321.

Tatarsky, B., 1969: The use of dynamic equations for a probabilistic forecasting of the pressure fields. – Izv. Akad. Nauk SSSR, FAO **5**, 293–297.

Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, 1992: Testing for nonlinearity in time series: the method of surrogate data. – Physica D **58**(1-4), 77–94.

Tse, D., P. Viswanath, 2005: Fundamentals of wireless communication – Cambridge University Press.

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

Tsyrulnikov, M.D., 2005: Stochastic modelling of model errors: A simulation study. – Quart. J. Roy. Meteor. Soc. **131**(613), 3345–3371.

Tsyrulnikov, M., D. Gayfulin, 2016: A spatio-temporal stochastic pattern generator for ensemble prediction and ensemble data assimilation in geophysical applications. – published online, arXiv preprint arXiv:1605.02018.

Tsyrulnikov, M., V. Gorin, 2013: Are atmospheric-model tendency errors perceivable from routine observations? – COSMO Newsletter No. **13**, 3–18.

Wallin, J., D. Bolin, 2015: Geostatistical modelling using non-Gaussian Matérn fields. – Scand. J. Statist. **42**, 872–890.

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