Scalable parallel scheme for sampling of Gaussian random fields over very large domains

L. de Carvalho Paludo, V. Bouvier, R. Cottereau

Received 26 April 2016; | Revised 6 June 2016; | Accepted 6 June 2016

DOI: xxx/xxxx

RESEARCH ARTICLE

Scalable parallel scheme for sampling of Gaussian random fields over very large domains

L. de Carvalho Paludo | V. Bouvier | R. Cottereau*

[1]Laboratoire MSSMat, CNRS, CentraleSupélec, Université Paris-Saclay, France

Correspondence
*Rs Cottereau. Email: regis.cottereau@centralesupelec.fr

Summary

This paper proposes a new scheme for the generation of Gaussian random fields over large domains (domain size much larger than the correlation length). The scheme decomposes the simulation domain into overlapping subdomains, and essentially generates independent random fields over each of them before merging them on the overlaps. It is naturally suited for simulation over clusters of computers. With this approach the number of operations for each processor depends only on the number of local degrees of freedom and not on the total number over all processors. Hence weak scalability is perfectly met. The paper describes the general scheme and introduces two error estimates for comparison with classical sampling schemes. Improvements in terms of scalability are demonstrated both theoretically and through numerical examples. The behavior in the overlap is studied in detail. Simulations using the localized approach were performed in up to 512 processors and allowed to generate in 41 s a realization of a random field over a cube of side 300 correlation lengths (close to 2 billions sampling points), much more efficiently and rapidly than with classical methods.

KEYWORDS:
Random field, Gaussian Process, Spectral Representation, Parallel Computing, Domain Decomposition

1 | INTRODUCTION

In a large number of scientific domains, random fields are routinely used to quantify uncertainties and parameterize fluctuating properties. These domains include micromechanics [1,2,3,4,5], geomechanics and seismic engineering [6,7,8,9], modeling of concrete and quasi-brittle materials [10,11], geophysics [12,13,14,15,16,17], structural mechanics [18,19], computational fluid dynamics [20,21], dynamics of chemical reactions [22], and many others. With the increase of computational power, larger and larger domains are being simulated. While solvers have been improved to make efficient use of large memory-distributed clusters, random field generation is still performed mostly sequentially. As the size of the domain increases, the random field generation step becomes a computational bottleneck for the simulation as a whole.

Random field generation schemes [23,24] are essentially separated in two groups, depending whether they perform in space (direct methods) or wave-number (spectral methods). Direct methods generate realizations by filtering a white noise through the square root of the covariance matrix. The simulation cost is essentially related to the computation of that square root. The Cholesky factorization [25] can be used but scales as $\mathcal{O}(N^3)$, where N is the number of points of the sample. As the covariance matrix is often sparse and circulant, optimized factorization algorithms can be introduced, reaching as low as $\mathcal{O}(N \log N)$

complexity[26,27]. Other direct methods compute polynomial approximations of the square root, and obtain preconditioned iterative schemes that are interesting for sampling large dimensional random fields[28,29]. A recent method proposes to dispatch the generation over smaller subdomains and to introduce statistical dependence between the random variables of the different subdomains[30]. However, none of these methods allows to decrease the complexity to the level of $\mathcal{O}(N)$, which is necessary for weak scaling over large clusters of processors.

Alternatively, the generation can be performed in the spectral domain[31,32,33,34,35], where the convolution is a simple product. The numerical cost is essentially that of computing the inverse Fourier transform, so the complexity can be lowered to $\mathcal{O}(N \log N)$ with the Fast Fourier Transform (FFT). However, the parallelization of the FFT algorithm over memory-distributed clusters is not trivial, and involves a large amount of communication between processors. This may cause performance loss, and eventually lead to a complexity above the theoretical $\mathcal{O}(N \log N)$ when a large number of processors are involved.

The objective of this paper is to construct a sampling scheme that overcomes these limitations. This paper describes an approach that requires less inter-processors communication and that has exactly $\mathcal{O}(N)$ complexity for large numbers of processors. The scheme decomposes the simulation domain into overlapping subdomains, each of which is assigned to a single processor. Each processor generates its own statistically-independent random field sample over its subdomain, and then merges it in the overlap areas with the neighboring random fields. The local realizations can be based indifferently on direct or spectral methods. With the merging formula that is introduced in this paper (Section 4), it will be shown that the first and second order statistics of the random field over the entire domain are conserved.

This paper is organized in four parts. The theoretical framework and classical numerical schemes are introduced in Section 2. Two error norms are then introduced in Section 3 in order to compare the different methods. These errors are based on two different estimations of the correlation of the simulated fields, based either on ensemble average or spatial average. The behavior of the errors and complexity of the classical generation techniques are discussed at the end of the same section. Section 4 is the core of the paper. It introduces the new parallel simulation scheme, shows that overall statistics of the random field are conserved, and discusses error and complexity of the scheme. Finally the scalability of a numerical implementation of our scheme (based on the FFT for the generation of the local random fields) is illustrated in Section 5.

2 | SPECTRAL METHODS TO SAMPLE GAUSSIAN RANDOM FIELDS

2.1 | Objectives and limitations

In this paper, we consider a probability space $(\Theta, \mathcal{T}, \mathbb{P})$, and we aim at generating realizations of a random field u with the following properties:

- u is indexed over $\mathbb{R}^d$, with $d \in 1, 2, 3$. In practice, however, realizations will be generated only inside a subdomain $[-L, L]^d$, typically over a grid with step h;
- u is second-order and homogeneous (or stationary). This implies in particular that its mean $\mu = \mathbb{E}[u(x)]$ and variance $\sigma^2 = \mathbb{E}[(u(x) - \mu)^2]$ are finite and do not depend on position x, and that its autocovariance $\mathcal{R}(|y - x|) = \mathbb{E}[(u(x) - \mu)(u(y) - \mu)]$ only depends on the distance between x and y;
- u is a centered Gaussian random field, with $\mu = 0$ and given autocovariance $\mathcal{R}(z)$ verifying unit variance $\mathcal{R}(0) = \sigma^2 = 1$. The power spectral density $\hat{\mathcal{R}}(k)$, which is the Fourier transform of $\mathcal{R}(z)$, is assumed to decay rapidly enough at high frequencies, in the sense that there exists $A_0 > 0$ and $\delta > 1$ such that $|\hat{\mathcal{R}}(k)| \leq A_0 |k|^{-\delta}$. The autocovariance is characterized by a correlation length $\ell_c = 2 \int_{\mathbb{R}^+} \mathcal{R}(z) dz$.

Note that we restrict ourselves to the sampling of Gaussian fields because they are the basic building block of most sampling schemes (although more refined schemes also exist[36]). The first-order marginal density can be modified locally by combining a direct and inverse Rosenblatt transforms[37], although one must consider the influence on the transformation of the correlation function[38,39]. We are also restricting ourselves to homogeneous random fields but they can be used as a building block to obtain non-homogeneous random fields by multiplication for example by a function of space (see for instance an example in time in[40]). Finally, we consider isotropic correlation structures $\mathcal{R}(|y - x|)$ instead of the more general case $\mathcal{R}(y - x)$ for simplicity. The results and analyses presented in the paper extend straightforwardly to anisotropic correlation structures.

Note finally that the scheme we propose is only interesting in the particular situation when $h < \ell_c \ll L$, where h is the discretization step (distance between two nodes of the simulation grid), ℓ_c is the correlation length and L is the size of the

domain. However, this should not be seen as a restriction but rather as the addressing of the difficult case. Indeed, when $L < \ell_c$ or $L \approx \ell_c$, the field can be effectively sampled over a coarse grid (with a step size relevant for the correlation length) and then interpolated onto the mesh of interest. When $h > \ell_c$, the sampling becomes essentially that of a white noise, which is numerically inexpensive. It is only natural then that our scheme would aim at addressing the remaining situation where $h < \ell_c \ll L$.

2.2 | Spectral Representation – Structured approaches

The generation of a second-order stationary random field u that follows a correlation function $\mathcal{R}$ can be obtained by linear combination of independent and identically distributed (i.i.d.) random variables in the following manner:

$$u_1^{\text{STR}}(x) = \sum_{\ell=-N}^{N} \sqrt{2\hat{\mathcal{R}}(\ell\Delta k)\Delta k}\ \cos(\ell\Delta k \cdot x + \Phi_\ell), \forall x \in \mathbb{R}, \tag{1}$$

where the $\{\Phi_\ell\}_{-N\leq\ell\leq N}$ are a set of independent identically distributed random variables, with uniform distribution over $[0, 2\pi]$. This so-called spectral representation was proposed in[41,42,43]. A similar formulation consists in using

$$u_2^{\text{STR}}(x) = \sum_{\ell=-N}^{N} \sqrt{2\hat{\mathcal{R}}(\ell\Delta k)\Delta k}\ \zeta_\ell \cos(\ell\Delta k \cdot x + \Phi_\ell), \forall x \in \mathbb{R}, \tag{2}$$

where the $\{\Phi_\ell\}_{-N\leq\ell\leq N}$ and $\{\zeta_\ell\}_{-N\leq\ell\leq N}$ are sets of independent random variables, the former distributed uniformly over $[0, 2\pi]$ the other two following a unit centered Gaussian distribution. One difference between the two formulas[44] is that the latter is a Gaussian random field for any discretization N, while the former is only asymptotically Gaussian, for $N \to \infty$. On the other hand, u_1^{STR} has stronger ergodicity properties than u_2^{STR}, since the spatial estimation of the autocovariance for each realization of $u_1^{\text{STR}}(x)$ verifies the objective $\mathcal{R}$ while ensemble average is necessary for the same to happen with $u_2^{\text{STR}}(x)$ (see Section 3 for details). Although the derivation will be presented here in 1D for simplicity, these formulations are extended straightforwardly to 3D, as detailed in[45].

These constructions imply that the averages vanish $\mathbb{E}[u_1^{\text{STR}}(x)] = \mathbb{E}[u_2^{\text{STR}}(x)] = 0$, thanks to the linearity of the averaging operator, independence of the set of random variables and the property that $\mathbb{E}[\cos(\alpha + \Phi_\ell)] = 0$ for any $\alpha \in \mathbb{R}$ and any $-N \leq \ell \leq N$. Concerning the autocovariance of the random field, we observe

$$\mathbb{E}[u_1^{\text{STR}}(x)u_1^{\text{STR}}(y)] = \mathbb{E}[u_2^{\text{STR}}(x)u_2^{\text{STR}}(y)] = \sum_{\ell=-N}^{N} \hat{\mathcal{R}}(\ell\Delta k)\cos(\ell\Delta k \cdot (y-x))\Delta k \tag{3}$$

thanks to the property that $\mathbb{E}[\cos(\alpha + \Phi_i)\cos(\beta + \Phi_j)] = \delta_{ij}\cos(\beta - \alpha)$ for any $\alpha, \beta \in \mathbb{R}$ and where δ_{ij} is such that $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$. Since the power spectral density $\hat{\mathcal{R}}(k)$ is real and positive, by property of the Fourier transform, Eq. (3) is an approximation of the power spectral density, for sufficiently small Δk (with respect to the inverse of the correlation length $1/\ell_c$) and large N (depending on the rate of decrease of the power spectral density). Finally, by virtue of the central limit theorem, the generated random fields are Gaussian (for N asymptotically large).

This approach will be denoted as "structured approach" to insist on the fact that it is based on a uniform discretization in the spectral space. The numerical complexity of computing a realization of the random fields in Eq. (1) or (2) is $\mathcal{O}(N^2)$. When using uniform discretization in wavenumber as above, the generated field is periodical, so that care should be taken to make the periodicity larger than the domain on which the field is generated. An alternative consists in using non-uniform discretization, as proposed for instance in[46]. This approach will not be discussed in this paper because it is not numerically competitive with the Fast Fourier Transform, which can efficiently be used to compute realizations of a random field with Eq. (1) or (2). In that case, the random field is obtained on a uniform grid in physical space, with step Δx and N points. For instance, for the example of Eq. (1):

$$u_1^{\text{FFT}}(m\Delta x) = 2\sum_{\ell=1}^{N} \sqrt{\hat{\mathcal{R}}(\ell\Delta k)\Delta k}\ \cos(\ell m\Delta k \cdot \Delta x + \Phi_\ell), \forall 1 \leq m \leq N, \tag{4}$$

The numerical complexity to compute a realization of the random field is then decreased to $\mathcal{O}(N \log_2 N)$.

2.3 | Spectral Representation – Randomization

In the previous section, the objective autocovariance is retrieved by approximation of the Fourier Transform with a uniform discretization over the spectral domain. An alternative consists in using a randomized discretization in space. More specifically,

it is observed that the properties of the power spectral density $\hat{\mathcal{R}}(k)$ make it a candidate probability distribution function for a random variable K with values in $\mathbb{R}$. Indeed, the power spectral density is positive, real, and normalized in the sense that $\int_{\mathbb{R}} \hat{\mathcal{R}}(k)dk = \sigma^2 = 1$. Generation of a random field with the properties of Section 2.1 can therefore be done as:

$$u^{\text{RND}}(x) = \sqrt{\frac{2}{N}} \sum_{\ell=-N}^{N} \cos(K_\ell \cdot x + \Phi_\ell), \forall x \in \mathbb{R}, \tag{5}$$

where the $\{\Phi_\ell\}_{-N\leq\ell\leq N}$ and $\{K_\ell\}_{-N\leq\ell\leq N}$ are sets of independent random variables, following uniform distributions over $[0, 2\pi]$ and probability density function $\hat{\mathcal{R}}(k)$ over $\mathbb{R}$, respectively. This formulation was apparently initially proposed in [47,48] and more recent contributions include [49,50,51]. An intermediate formula between the randomization and the structured approach is discussed in [31], where random, almost-evenly distributed frequencies are chosen.

For a given $-N \leq \ell \leq N$, by independence of K_ℓ and Φ_ℓ, we have $\mathbb{E}[\cos(K_\ell \cdot x + \Phi_\ell)] = \mathbb{E}[\cos(K_\ell \cdot x)]\mathbb{E}[\cos(\Phi_\ell)] - \mathbb{E}[\sin(K_\ell \cdot x)]\mathbb{E}[\sin(\Phi_\ell)]$. Since $\mathbb{E}[\cos(\Phi_\ell)] = \mathbb{E}[\sin(\Phi_\ell)] = 0$, and using the linearity of the expectation operator, we see that the average of the generated field is as expected $\mathbb{E}[u^{\text{RND}}(x)] = 0$. Concerning the covariance, similar arguments and using $\mathbb{E}[\cos^2(\Phi_\ell)] = \mathbb{E}[\sin^2(\Phi_\ell)] = 1/2$ and $\mathbb{E}[\cos(\Phi_\ell)\sin(\Phi_\ell)] = 0$, yield

$$\mathbb{E}[u^{\text{RND}}(x)u^{\text{RND}}(y)] = \frac{1}{N} \sum_{\ell=-N}^{N} \mathbb{E}[\cos(K_\ell \cdot (x-y))] = \frac{1}{N} \sum_{\ell=-N}^{N} \mathcal{R}(x-y) = \mathcal{R}(x-y). \tag{6}$$

The last equality results from the fact that $\mathbb{E}[\cos(K_\ell(x-y))] = \int_{\mathbb{R}} \cos(k(x-y))\hat{\mathcal{R}}(k)dk$, where the latter can be seen as a Fourier transform.

As before, the complexity of this method is $\mathcal{O}(N^2)$. The main difficulty with this approach lies in the simulation of realizations of the random variables $\{K_\ell\}_{1\leq\ell\leq N}$. While it may be readily available (and numerically cheap) to perform for some very specific distributions, it is not so for general distributions. In the general case, the Rosenblatt Transform [37] can be used after drawing samples from another distribution, but it requires to find the inverse of the cumulated PDF of $2\hat{\mathcal{R}}$, which may be numerically expensive. This numerical cost of the method of course adds up with that of computing the Fourier Transform of the covariance, as with all other spectral methods. This approach can be generalized in subsets of $\mathbb{R}^d$, so-called bins, to improve the distribution of the random wave numbers over the spectrum. A hybrid approach where the lower part of the spectrum with the deterministic approach and the higher part with RM-bins is discussed in the literature by [49]. The interest is to mitigate problems of representing the whole spectrum extent.

3 | COMPARISON OF THE GENERATION METHODS

3.1 | Methodology of the discussion

When using the spectral formula presented in the previous section, ergodic Gaussian random fields are obtained in the limit of large N by the sum of many independent random variables. When the number of germs is finite, the random fields obtained are not exactly Gaussian because the sum is over finite number of samples. Also, the computation of the correlation function as the inverse Fourier transform of the power spectral density is only approximated through discretization: (i) by cutting the integral at wavenumber $k_{\max}$, and (ii) by discretizing with a wavenumber step Δk. The objective of this section is to assess precisely these errors as a function of the number of germs. As the objective of all spectral simulation methods is to reproduce the second order statistics of the field, the correlation, we choose naturally to define the error as a deviation between the correlation of the simulated field and the objective (true) correlation. However two such errors are eventually introduced. The first error estimate, denoted ε_h, accounts for the error in the integral approximation of the correlation. The second error estimate, denoted ε_ω, quantifies the error expected in $\mathcal{R}$ for a single realization $\omega \in \Omega$. The latter therefore quantifies the ergodicity of the random field. We therefore start in the next section by defining two different estimations of the correlation of the sample field, followed by definitions of the corresponding errors. Finally, the rest of the section is devoted to computing the errors for the different methods proposed. For simplicity, the discussion will be performed only in 1D. Results extend to 3D straightforwardly.

3.2 | Two correlation-based errors

A first estimator $\mathbf{R}$ of the correlation function is:

$$\mathbf{R}(z) = \lim_{L\to+\infty} \frac{1}{2L} \int_{-L}^{L} \mathbb{E}\left[u(x+z)u(x)\right] \mathrm{d}x. \tag{7}$$

This estimator involves averages both in space and in the random dimension and is expected to be representative of the discretized autocovariance, slightly different from the reference $\mathcal{R}$. Considering the residual between this estimator and $\mathcal{R}$ yields a first definition of the error, sensitive to the discretization of the autocovariance:

$$\varepsilon_h = \|\mathbf{R} - \mathcal{R}\|_{\mathbb{L}^2(\mathbb{R})} \tag{8}$$

Based on a single given realization, it is also possible to define an alternative estimator $\mathbf{R}_\omega$ of $\mathcal{R}$:

$$\mathbf{R}_\omega(z) = \lim_{L\to+\infty} \frac{1}{2L} \int_{-L}^{L} u(x+z)u(x)\mathrm{d}x. \tag{9}$$

This estimator corresponds to the so-called unbiased correlation estimator in [44]. It is indeed based on an hypothesis of ergodicity of the field, since the spatial average of Eq. (9) replaces the ensemble average of the definition of the covariance $\mathcal{R}$. Note that this estimator is a random variable and that $\mathbf{R}(z) = \mathbb{E}\left[\mathbf{R}_\omega(z)\right]$. Considering the residual of that estimator with the previous yields a second error indicator ε_ω, more sensitive to intrinsic stochastic properties of the generated fields (such as ergodicity). This error indicator is constructed as a L^∞ norm $\varepsilon_\omega = \|e_\omega(z)\|_{L^\infty(\mathbb{R})} = \max_{z\in\mathbb{R}} e_\omega(z)$ where the function $e_\omega(z) > 0$ is defined for each $z \in \mathbb{R}$ by :

$$\mathbb{P}\left[|\mathbf{R}_\omega(z) - \mathbf{R}(z)| \le e_\omega(z)\right] = \alpha, \tag{10}$$

where $\alpha \in [0,1]$ is a confidence level to be chosen. By definition of the cumulative density function, this choice corresponds to $e_\omega(z)$ being the inverse of the cumulative density function (ICDF) of $|\mathbf{R}_\omega(z) - \mathbf{R}(z)|$, evaluated at α. Note that, when $\mathbf{R}_\omega(z) - \mathbf{R}(z)$ is a centered Gaussian random variable (for fixed z), that ICDF of the absolute value $|\mathbf{R}_\omega(z) - \mathbf{R}(z)|$ can be evaluated with the ICDF of the Gaussian random variable as:

$$\mathbb{P}\left[|\mathbf{R}_\omega - \mathbf{R}| \le e_\omega(z)\right] = 2\mathbb{P}\left[\mathbf{R}_\omega - \mathbf{R} \le e_\omega(z)\right] - 1 = 2\Phi_{\mathbf{R}_\omega - \mathbf{R}}(e_\omega(z)) - 1, \tag{11}$$

where $\Phi_{\mathbf{R}_\omega - \mathbf{R}}(e_\omega(z))$ is the cumulative density function of the centered Gaussian random variable $\mathbf{R}_\omega(z) - \mathbf{R}(z)$. Eventually, we define the error estimate as:

$$\varepsilon_\omega = \max_{z\in\mathbb{R}} \Phi^{-1}_{\mathbf{R}_\omega - \mathbf{R}}((\alpha-1)/2) = \max_{\sigma^2_{\mathbf{R}_\omega - \mathbf{R}} \in \mathbb{R}} \Phi^{-1}_{\mathbf{R}_\omega - \mathbf{R}}\left(\frac{\alpha-1}{2}\right) \tag{12}$$

In that case, the properties of the ICDF mean that the error estimate is just an image of the maximum over $z \in \mathbb{R}$ of the variances of the different random variables $\mathbf{R}_\omega - \mathbf{R}$. For a fixed α, ergodic fields are expected to correspond to smaller errors (small maximum variance) than non-ergodic fields (larger maximum variance).

The rest of the section is devoted to the computation of the two errors ε_h and ε_ω for the different generation formulas proposed in Section 2.

3.3 | A priori error estimation - structured case (STR)

Using Eq. (3) with $y = x + z$, we get for both structured generation methods that the correlation between the fields at x and $x + z$ is independent of x. Hence the integral and limit in Eq. (7) play no role and we obtain:

$$\mathbf{R}^{\mathrm{STR}}(z) = 2\sum_{\ell=1}^{N} \hat{\mathcal{R}}(\ell\Delta k)\cos(\ell\Delta k \cdot z)\Delta k, \tag{13}$$

which is the N-term Fourier series approximation of $\mathcal{R}(z)$. Using the Parseval identity, the error $\varepsilon_h^{\mathrm{STR}}$ for both structured generation schemes can then be written as :

$$\|\mathbf{R}^{\mathrm{STR}}(z) - \mathcal{R}(z)\|^2_{\mathbb{L}^2(\mathbb{R})} \le 2N(\Delta k)^3 \max_{k\in\mathbb{R}} |\hat{\mathcal{R}}'(k)|^2 + 2\frac{A_0^2}{2\delta - 1}\frac{1}{k_{\max}^{2\delta-1}}, \tag{14}$$

where the first term comes discretizing the continuous function $\hat{\mathcal{R}}$ with step $\Delta k = k_{\max}/N$ while the second arises from cutting the spectral domain at $k_{\max}$. Note that $\hat{\mathcal{R}}'(k) = \partial\hat{\mathcal{R}}(k)/\partial k$ and that we have used the bound A_0/k^δ on the power spectrum

density proposed in Section 2.1. Finally, we obtain the following bound for the discretization error for both structured generation schemes:

$$\varepsilon_h^{\text{STR}} \leq \sqrt{C_2^{\text{STR}} \frac{k_{\max}^3}{N^2} + \frac{C_1^{\text{STR}}}{k_{\max}^{2\delta-1}}}, \tag{15}$$

where the constants $C_1^{\text{STR}} = 2A_0^2/(2\delta - 1)$ and $C_2^{\text{STR}} = 2\max_{k\in\mathbb{R}} |\hat{\mathcal{R}}'(k)|^2$ depend on the continuous power spectrum density $\hat{\mathcal{R}}(k)$ and are independent from the discretization parameters. For fixed $k_{\max}$, the first part of the error decreases whenever N increases. Since $\delta > 1$, the second part of the error decreases whenever $k_{\max}$ increases.

We now turn to the second type of errors, and first notice that, whenever $i \neq \ell$ $\lim_{L\to+\infty} \int_{-L}^{L} \cos(k_\ell(x+z)+\Phi_\ell)\cos(k_i x + \Phi_i)dx = 0$. We also notice that, for any k and Φ, $\lim_{L\to+\infty} 2\int_{-L}^{L} \cos(k(x+z)+\Phi)\cos(kx+\Phi)dx = \cos(kz)$. Eventually, we obtain for the first type of structured generation method, that:

$$\mathbf{R}_\omega^{\text{STR1}}(z) = 2\sum_{\ell=1}^{N} \hat{\mathcal{R}}(\ell\Delta k)\cos(\ell\Delta k \cdot z)\Delta k = \mathbf{R}^{\text{STR}}(z). \tag{16}$$

so the corresponding error vanishes $\varepsilon_\omega^{\text{STR1}} = 0$ for any confidence level α. For the second type of structured generation method, and using similar arguments, we obtain that

$$\mathbf{R}_\omega^{\text{STR2}}(z) = 2\sum_{\ell=1}^{N} \hat{\mathcal{R}}(\ell\Delta k)\zeta_\ell^2\cos(\ell\Delta k \cdot z)\Delta k. \tag{17}$$

and, using Eq. (16),

$$\mathbf{R}_\omega^{\text{STR2}}(z) - \mathbf{R}^{\text{STR}}(z) = 2\Delta k\sum_{\ell=1}^{N} \hat{\mathcal{R}}(\ell\Delta k)\cos(\ell\Delta k \cdot z)(\zeta_\ell^2 - 1). \tag{18}$$

As expected, this variable is centered since $\mathbb{E}[\zeta_\ell^2] = 1$. For asymptotically large N, the Central Limit Theorem ensures that this random variable converges in distribution to a centered Gaussian law with variance $8\Delta k^2 \sum_{\ell=1}^{N} \hat{\mathcal{R}}^2(\ell\Delta k)\cos^2(\ell\Delta k \cdot z)$, since ζ_ℓ are unit centered Gaussian random variables, such that $\mathbb{E}[\zeta_\ell^4 - 1] = 2$. The maximum variance is then $\sigma_{\text{STR2}}^2 = 8\Delta k^2 \sum_{\ell=1}^{N} \hat{\mathcal{R}}^2(\ell\Delta k) \approx 4\Delta k \int_{\mathbb{R}} \mathcal{R}^2(x)dx$. Introducing the constant $C_3^{\text{STR}} = 4\int_{\mathbb{R}} \mathcal{R}^2(x)dx$, only dependent on the continuous power spectrum density $\hat{\mathcal{R}}(k)$ and independent from the discretization parameters, we then have

$$\varepsilon_\omega^{\text{STR2}} = \Phi^{-1}\left[C_3^{\text{STR}}\frac{k_{\max}}{N}\right]\left(\frac{\alpha-1}{2}\right) \tag{19}$$

where $\Phi[\sigma^2]$ is the cumulative density function of a centered Gaussian random variable with variance σ^2. For a fixed α, this error therefore decreases when Δk decreases.

3.4 | A priori error estimation - randomized case (RND)

Considering the Randomization method, and shown in Eq. (6), $\mathbf{R}^{\text{RND}}(z) = \mathcal{R}(z)$, so that we first obtain $\varepsilon_h^{\text{RND}} = 0$. On the other hand, the random estimator $\mathbf{R}_\omega^{\text{RND}}(z)$ is:

$$\mathbf{R}_\omega^{\text{RND}}(z) = \frac{1}{2N}\sum_{\ell=1}^{N} \cos(K_\ell \cdot z). \tag{20}$$

To obtain this result, we first observe that, for any $k \neq 0$, and any $b \in \mathbb{R}$, we have $\int_{-L}^{L} \cos(kx+b)dx = 2\cos(b)\sin(kL)/k$. This naturally implies that $\lim_{L\to\infty} \int_{-L}^{L} (\cos(kx+b)dx)/(2L) = 0$. Whenever $k = 0$, we obtain $\lim_{L\to\infty} \int_{-L}^{L} (\cos(b)dx)/(2L) = \cos(b)$. The result above is then obtained by separating $2\cos(K_\ell \cdot x + \Phi_\ell)\cos(K_m \cdot (x+z) + \Phi_m) = \cos((K_\ell + K_m)\cdot x + K_m \cdot z + \Phi_\ell + \Phi_m) + \cos((K_\ell - K_m)\cdot x - K_m \cdot z + \Phi_\ell - \Phi_m)$ and separating in the sum the terms for which $\ell = m$ and $\ell \neq m$.

Observing that $\mathbb{E}[\cos^2(K_\ell \cdot z)] = (1 + \mathcal{R}(2z))/2$ and $\mathbb{E}[\cos(K_\ell \cdot z)]^2 = \mathcal{R}(z)^2$, we obtain that for asymptotically large N, the random variable $\mathbf{R}_\omega^{\text{RND}}(z) - \mathbf{R}^{\text{RND}}(z)$ is a centered Gaussian random variable with variance $((1 + \mathcal{R}(2z))/2 - \mathcal{R}(z)^2)/(4N) \leq 1/(4N)$. Introducing the constant $C^{\text{RND}} = \max_{z\in\mathbb{R}}((1+\mathcal{R}(2z))/2 - \mathcal{R}(z)^2)/4$, only dependent on the continuous power spectrum density $\hat{\mathcal{R}}(k)$ and independent from the discretization parameters, we obtain

$$\varepsilon_\omega^{\text{RND}} = \Phi^{-1}\left[\frac{C^{\text{RND}}}{N}\right]\left(\frac{\alpha-1}{2}\right) \tag{21}$$

3.5 | Comparison of the different generation methods

We finalize this section with a short comparison between the different generation methods. We report in Table 1 the numerical complexity of generating samples with the different approaches, along with the order of convergence (in the number of points in wavenumber space) of the discretization and ergodicity errors. These orders of convergence were obtained by assuming in the formulas of Sections 3.3 and 3.4 that the value of $k_{\max}$ was large enough to induce negligible errors. Also, the FFT can be applied to both forms of the structured method, so the ergodicity error depends on which was chosen. If cost is approximated by complexity, this table implies that the errors decrease very slowly with increases of the numerical cost: dividing by 10 the errors corresponds to multiplying by 100 the numerical cost.

TABLE 1 Complexity and correlation errors for the different generation methods considered.

Generation method	Complexity	Discretization error ε_h	Ergodic error ε_ω
Structured 1	$\mathcal{O}(N^2)$	$\mathcal{O}(N^{-1})$	0
Structured 2	$\mathcal{O}(N^2)$	$\mathcal{O}(N^{-1})$	$\mathcal{O}(N^{-1})$
FFT	$\mathcal{O}(N \log N)$	$\mathcal{O}(N^{-1})$	0 or $\mathcal{O}(N^{-1})$
Randomized	$\mathcal{O}(N^2)$	0	$\mathcal{O}(N^{-1})$

4 | LOCALIZATION METHOD FOR THE SAMPLING OF RANDOM FIELDS OVER LARGE DOMAINS

After comparing existing simulation methods in detail, we now turn to the core of this paper, which is to address the issue of generating realizations of random fields on very large domains. We consider a domain whose size L is much bigger than the correlation length ℓ_c. As discussed in the introduction and the previous section, the complexity of generation algorithms implies that this often cannot be done on a single computer. The use of a distributed-memory cluster is then mandatory. Such an approach can be performed rather straightforwardly with parallelized algorithm, such as provided by the FFTW library. However, working with more processors does not decrease the complexity. Since this complexity here is larger than linear (at best $\mathcal{O}(N \log N)$), the generation will still be impossible for a large enough L/ℓ_c. Also, practical complexity might be worse in parallel environnement due to communication or memory allocation issues. There is therefore a need for a generation algorithm that scales linearly with the number of simulation points N. This section introduces such an algorithm.

4.1 | The core idea

It is interesting the observe that we aim at generating random fields with finite correlation length using sums of cosines. On the one hand, finite correlation length means that points far apart (with distance larger than the correlation length, on order of magnitude) are uncorrelated . On the other hand, cosines induce correlation over the entire length of domain: points separated by one period are exactly correlated. It is therefore quite ureasonable that a very large number of cosine functions (with different frequencies and random phases) should be summed in order to retrieve a field that is only correlated on small scales. When interested in the situation $\ell_c \ll L$, which is our case, it may be possible to break the generation into several independent generations on smaller domains, and then to recombine the independent samples to retrieve small-scale correlations at the interfaces. This is the approach followed here.

We therefore introduce a d-dimensional domain $D = [-L, L]^d$ and a partition of unity of P positive functions $\{\phi_\ell(\mathbf{x})\}_{1\leq\ell\leq P}$ over that domain. By definition, a partition of unity is such that $\sum_{\ell=1}^{n} \phi_\ell(\mathbf{x}) = 1$, $\forall \mathbf{x} \in D$. The functions are assumed positive, in the sense that $\phi_\ell(x) \geq 0$ for all $x \in D$ and $1 \leq \ell \leq P$. The core idea of this paper consists in constructing realizations of

random fields as

$$u^{\mathrm{LOC}}(\mathbf{x}) = \sum_{\ell=1}^{P} \sqrt{\phi_\ell(\mathbf{x})}\, u_\ell(\mathbf{x}), \tag{22}$$

where the random fields $\{u_\ell(\mathbf{x})\}_{1\leq\ell\leq P}$ are P independent centered Gaussian random fields with the objective correlation structure, typically generated with one of the methods described in Section 2. We will show further down that, given some hypotheses on the choice of the partition of unity functions, the reconstructed field $u^{\mathrm{LOC}}(\mathbf{x})$ is a centered Gaussian random field with the same correlation structure, and that the random fields $\{u_\ell(\mathbf{x})\}_{1\leq\ell\leq n}$ can be generated over small subdomains, which correspond to the local support of the different functions $\{\phi_\ell(\mathbf{x})\}_{1\leq\ell\leq n}$ of the partition of unity. The structure of Eq. 22 implies that the support of the local random fields is controlled by that of the partition of unity functions. Hence, choosing partition of unity functions with localized support allows to distribute the generation of the global random field $u^{\mathrm{LOC}}(\mathbf{x})$ into the generation of a set of independent random fields localized over (potentially dramatically) smaller supports.

4.2 | A 1D example with two subdomains

Before moving to a more general description of the algorithm, we introduce it on a simple 1D example with only two subdomains. We will show how to merge two independent random fields $u_1(x)$ and $u_2(x)$, each supported on a small portion of the domain, into a single field u statistically homogeneous (weakly) over the entire domain. We consider the domain $[-L, L]$, and a centered overlap area $[-\alpha, \alpha]$, with $\alpha \in [0, 1]$, and the simplest partition of unity:

$$\phi_1(x) = \begin{cases} 0 & \text{for } x \in [-1, -\alpha] \\ (1 + x/\alpha)/2 & \text{for } x \in [-\alpha, \alpha] \\ 1 & \text{for } x \in [\alpha, 1] \end{cases}, \quad \phi_2(x) = \begin{cases} 1 & \text{for } x \in [-1, -\alpha] \\ (1 - x/\alpha)/2 & \text{for } x \in [-\alpha, \alpha] \\ 0 & \text{for } x \in [\alpha, 1] \end{cases}. \tag{23}$$

These functions are constant outside of the band $[-\alpha, \alpha]$ and linear inside. We observe that $\phi_2(x) = 1 - \phi_1(x)$ so this pair of functions clearly forms a partition of unity over $[-L, L]$. The support of ϕ_1 is $[-\alpha, L]$ while that of ϕ_2 is $[-L, \alpha]$. We now consider the random field

$$u^{\mathrm{LOC}}(x) = \sqrt{\phi_1(x)}u_1(x) + \sqrt{\phi_2(x)}u_2(x), \tag{24}$$

where $u_1(x)$ and $u_2(x)$ are two independent realizations of the same Gaussian random field with vanishing average and unit variance. It is essential to observe that since the support of ϕ_1 and ϕ_2 is smaller than the global domain $[-L, L]$, the local fields have to be generated only on small subdomains, defined as the support of their corresponding weight function. The procedure is illustrated in Figure (1).

By linearity of the averaging operator, $\mathbb{E}[u^{\mathrm{LOC}}(x)] = 0$. Also, the independence of the two processes means that $\mathbb{E}[(u^{\mathrm{LOC}}(x))^2] = \phi_1(x)\mathbb{E}[u_1^2(x)] + \phi_2(x)\mathbb{E}[u_2^2(x)] = 1$ since ϕ_1 and ϕ_2 form a partition of unity. Using the independence property and the fact that the fields are centered, we obtain that

$$\mathbb{E}\left[u^{\mathrm{LOC}}(x)u^{\mathrm{LOC}}(y)\right] = \left(\sqrt{\phi_1(x)\phi_1(y)} + \sqrt{\phi_2(x)\phi_2(y)}\right)\mathcal{R}(x - y), \tag{25}$$

where $\mathcal{R}$ is the common correlation function of the random fields u_1 and u_2. In general, the correlation $\mathbb{E}\left[u^{\mathrm{LOC}}(x)u^{\mathrm{LOC}}(y)\right]$ is of course not equal to $\mathcal{R}(x-y)$. However, if the functions $\phi_1(x)$ and $\phi_2(x)$ fluctuate slowly at the scale of $|x-y|$, then $\phi_1(x) \approx \phi_1(y)$ and $\phi_2(x) \approx \phi_2(y)$, and finally $\sqrt{\phi_1(x)\phi_1(y)} + \sqrt{\phi_2(x)\phi_2(y)} \approx 1$ using again the property of partition of unity. We bound the error on the correlation through

$$\left|\mathbb{E}\left[u^{\mathrm{LOC}}(x)u^{\mathrm{LOC}}(y)\right] - \mathcal{R}(x - y)\right| \leq \left|1 - \sqrt{\phi_1(x)\phi_1(y)} - \sqrt{\phi_2(x)\phi_2(y)}\right| |\mathcal{R}(x - y)| \tag{26}$$

We observe that $1 - \sqrt{\phi_1(x)\phi_1(y)} - \sqrt{\phi_2(x)\phi_2(y)} = \sqrt{\phi_1(x)}\left(\sqrt{\phi_1(x)} - \sqrt{\phi_1(y)}\right) + \sqrt{\phi_2(x)}\left(\sqrt{\phi_2(x)} - \sqrt{\phi_2(y)}\right)$, and, since the partition of unity function are positive (and inferior to 1 by extension), we then get the bound

$$\left|\mathbb{E}\left[u^{\mathrm{LOC}}(x)u^{\mathrm{LOC}}(y)\right] - \mathcal{R}(x - y)\right| \leq \left(\left|\sqrt{\phi_1(x)} - \sqrt{\phi_1(y)}\right| + \left|\sqrt{\phi_2(x)} - \sqrt{\phi_2(y)}\right|\right) |\mathcal{R}(x - y)|, \tag{27}$$

For any function $f \in \mathcal{C}^1([-\alpha, \alpha])$, the mean-value theorem states that $|f(x) - f(y)| \leq |x - y| \sup_{z\in[-\alpha,\alpha]} |f'(z)|$ so we finally obtain, using also the symmetry of the two functions $\phi_1(x)$ and $\phi_2(x)$:

$$\left|\mathbb{E}\left[u^{\mathrm{LOC}}(x)u^{\mathrm{LOC}}(y)\right] - \mathcal{R}(x - y)\right| \leq 2\mathcal{M} \sup_{[-\alpha,\alpha]} \left|\left(\sqrt{\phi_1}\right)'\right|, \tag{28}$$

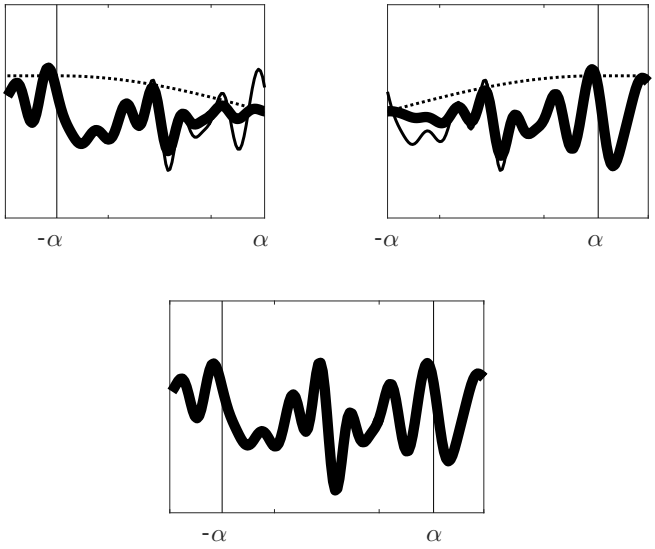


FIGURE 1 Generation of two independent samples (thin solid lines) $u_1(x)$ (top left) and $u_2(x)$ (top right), with the respective partition of unity functions $\phi_1(x)$ and $\phi_2(x)$ (dashed lines), defined respectively over $[-L, \alpha]$ and $[-\alpha, L]$. The thick solid lines represent the products $\phi_1(x)u_1(x)$ (top left) and $\phi_2(x)u_2(x)$ (top left) and the reconstructed global random field $u(x)$ over the entire domain (lower plot)

where we have introduced the constant $\mathcal{M} = \sup_{z\in\mathbb{R}} |z\mathcal{R}(z)|$, which is bounded thanks to the hypotheses on the correlation structure (faster than linear decrease at infinity). It is therefore sufficient to choose adequately the overlap area (the length α in 1D) and the partition of unity to control the error ϵ on the correlation structure

Obviously, the choice of linear partition of unity functions in Eq. (23) was not ideal since the derivative of the square root is then not bounded. Other functions present a better behavior, such as

$$\phi_1(x) = \begin{cases} 0 & \text{for } x \in [-L, -\alpha] \\ \sin^2\left[\frac{\pi}{4}\left(1 + \frac{x}{\alpha}\right)\right] & \text{for } x \in [-\alpha, \alpha] \\ 1 & \text{for } x \in [\alpha, L] \end{cases}, \quad \phi_2(x) = \begin{cases} 1 & \text{for } x \in [-L, -\alpha] \\ \cos^2\left[\frac{\pi}{4}\left(1 + \frac{x}{\alpha}\right)\right] & \text{for } x \in [-\alpha, \alpha] \\ 0 & \text{for } x \in [\alpha, L] \end{cases}. \tag{29}$$

Using such a partition of unity, the choice of overlap length becomes

$$\alpha \geq \frac{\pi}{2}\frac{\mathcal{M}}{\epsilon}, \tag{30}$$

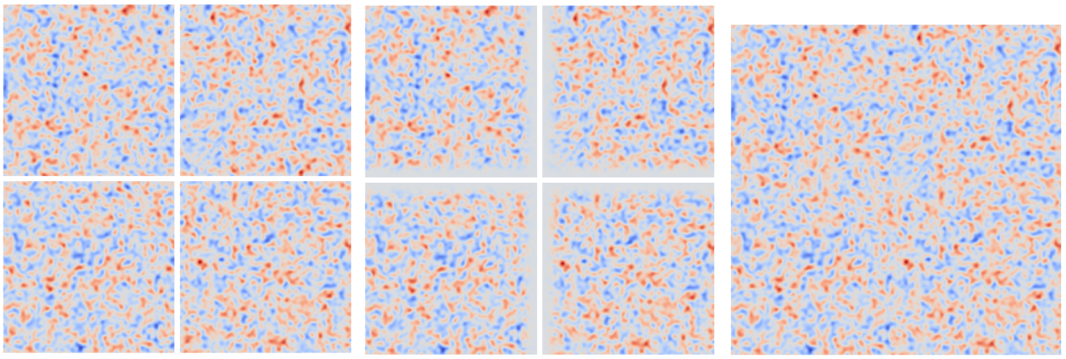
for a chosen level of error $\epsilon = \left|\mathbb{E}\left[u^{\mathrm{LOC}}(x)u^{\mathrm{LOC}}(y)\right] - \mathcal{R}(x-y)\right|$ on the correlation function. Table 2 gives some examples of values for α for different levels of errors and correlation structures. Obviously, there is always the requirement that $\alpha < L$ by definition, so large L might be required when a low error is sought for.

4.3 | Localization method for the generation of d-variate random fields in large domains

We now generalize the method described above to d-variate random fields, and any numbers of subdomains. We therefore come back to the setting of Section 4.1, with a d-dimensional domain $D = [-L, L]^d$ and a partition of unity of P functions $\{\phi_\ell(\mathbf{x})\}_{1\leq\ell\leq P}$ over that domain. As described in Eq. (22), the localized generation method consists in constructing the field $u^{\mathrm{LOC}}(\mathbf{x})$ over D as a merging of P local independent realizations of random fields $u_\ell(\mathbf{x})$ with the same second order statistics as the objective, and localized over the support of each of the partition of unity functions. An illustration in 2D is provided in Fig. 2 . Note that in this illustration, the area where the merging takes place is only a small subset of each of the local subdomains where the independent random fields are generated.

TABLE 2 Needed overlap size depending on the correlation model and error requirements when using the partition of unity in Equation (29) and correlation length $\ell_c = 1$

Correlation Model	Correlation Function, $\mathcal{R}$	$\mathcal{M}$	α minimum required for:		
			$\epsilon = 0.2$	$\epsilon = 0.1$	$\epsilon = 0.01$
Exponential	$\exp(-2z)$	0.1839	0.722	1.444	14.443
Power-law	$\dfrac{1}{\left(1+\frac{\pi^2 z^2}{4}\right)^2}$	0.2067	0.812	1.623	16.234
Gaussian	$\exp(-\pi z^2)$	0.2420	0.950	1.901	19.007
Triangular	$\dfrac{12(2-2\cos(2\pi z)-(2\pi z)\sin(2\pi z))}{(2\pi z)^4}$	0.2579	1.013	2.026	20.255
Low-pass white noise	$\dfrac{3\left(\sin\left(\frac{3\pi}{2}z\right)-\frac{3\pi}{2}z\cos\left(\frac{3\pi}{2}z\right)\right)}{\left(\frac{3\pi}{2}z\right)^3}$	0.2777	1.091	2.181	21.811



(a) $u_\ell(x)$, for $1 \le \ell \le 4$ (b) $\sqrt{\phi_\ell(x)}u_\ell(x)$, for $1 \le \ell \le 4$ (c) $u(x) = \sum_{\ell=1}^{4} u_\ell(x)$

FIGURE 2 Generation of four independent realizations of the local random fields (left figure), multiplication by the square root of the partition of unity functions (center figure), and merging on the overlaps to obtain the global fully correlated random field (right figure).

As before, the average of the global random field vanishes thanks to the linearity of the expectation operator $\mathbb{E}\left[u^{\mathrm{LOC}}(\mathbf{x})\right] = 0$, $\forall \mathbf{x} \in D$. The variance is obtained using the independence of the local random fields and the property of the partition of unity functions: $\mathbb{E}[(u^{\mathrm{LOC}}(\mathbf{x}))^2] = \sum_{\ell=1}^{P} \phi_\ell(\mathbf{x})\ \mathbb{E}[(u_\ell(\mathbf{x}))^2] = 1$, $\forall \mathbf{x} \in D$. Further, using similar arguments as in 1D, we obtain the following error on the correlation structure

$$\left|\mathbb{E}\left[u^{\mathrm{LOC}}(\mathbf{x})u^{\mathrm{LOC}}(\mathbf{y})\right] - \mathcal{R}(\mathbf{x}-\mathbf{y})\right| \le \mathcal{R}(\mathbf{x}-\mathbf{y}) \sum_{\ell=1}^{P} \left|\sqrt{\phi_\ell(\mathbf{x})} - \sqrt{\phi_\ell(\mathbf{y})}\right|, \tag{31}$$

The multi-dimensional version of the mean-value theorem then yields

$$\left|\mathbb{E}\left[u^{\mathrm{LOC}}(\mathbf{x})u^{\mathrm{LOC}}(\mathbf{y})\right] - \mathcal{R}(\mathbf{x}-\mathbf{y})\right| \le \mathcal{R}(\mathbf{x}-\mathbf{y})\,|\mathbf{x}-\mathbf{y}| \sum_{\ell=1}^{P} \sup_{\mathbf{x}\in D_\ell} \left|\nabla\left(\sqrt{\phi_\ell}\right)\right| \le \mathcal{M} \sum_{\ell=1}^{P} \sup_{\mathbf{x}\in D_\ell} \left|\nabla\left(\sqrt{\phi_\ell}\right)\right|, \tag{32}$$

where $\mathcal{M} = \sup_{\mathbf{x}\in D} |\mathbf{x}||\mathcal{R}(\mathbf{x})|$.

For simplicity, we now consider a structured discretization of the domain D into a regular pavement of space (squares in 2D, hexahedra in 3D), where the functions of the partition of unity are defined as tensorizations of 1D functions. These 1D functions are themselves defined as translations and rescaling from a common function $\phi_0(x)$, defined on a reference domain $D_0 = [-L, L]$. In 2D for instance, where the domain is $[-L, L]^2$ and p functions are used in each direction ($p^2 = P$), the length $[-L; L]$ is separated into p intervals of length $2L/p$ in each direction. The centers of these intervals are denoted $\{x_i\}_{1\le i\le p}$ and

$\{y_i\}_{1\leq i\leq p}$. The functions of the partition of unity are then defined (accounting as in the previous section for overlaps over length $\alpha < L$) as

$$\phi_{ij}(\mathbf{x}) = \phi_0\left(p(x-x_i)\right)\phi_0\left(p(y-y_i)\right), \tag{33}$$

which is such that:

$$\sup_{\mathbf{x}\in D_{ij}}\left|\nabla\left(\sqrt{\phi_{ij}(\mathbf{x})}\right)\right| \leq \sqrt{2}p\sup_{x\in D_0}\left|\sqrt{\phi_0(x)}\right|\sup_{x\in D_0}\left|\left(\sqrt{\phi_0(x)}\right)'\right| \leq \sqrt{2}p\frac{\pi}{4\alpha}. \tag{34}$$

The last inequality is obtained assuming the use of the same 1D functions as in the previous section (see Eq. (29)). Finally, acknowledging the fact that only at most 4 functions (2^d in d dimensions) can be non-vanishing in one given point of such 2D structured pavement, the sum over P can be bound with only 4 times the bound above. We obtain

$$\epsilon = \left|\mathbb{E}\left[u^{\mathrm{LOC}}(\mathbf{x})u^{\mathrm{LOC}}(\mathbf{y})\right] - \mathcal{R}(\mathbf{x}-\mathbf{y})\right| \leq \sqrt{2}p\frac{\pi\mathcal{M}}{\alpha}. \tag{35}$$

This equation can be inverted to obtain an estimate of the overlap length required for a given accuracy level. Generalizing the above equation in d dimensions yields the following estimate of the overlap:

$$\frac{\alpha}{p} \leq 2^{d-2}\sqrt{d}\pi\frac{\mathcal{M}}{\epsilon}. \tag{36}$$

Note that, in the derivation above, α is an overlapping length relative to the total length of the sample $2L$. A more relevant quantity is α/p, indicating the size of the overlap with respect to the size of one subdomain.

5 | NUMERICAL RESULTS

In this last section, we turn to the actual implementation of the algorithm presented in the previous section, and to numerical tests. In particular, we consider weak scalability, which tests the scalability of the algorithm for larger numbers of processors and constant numerical cost on each processor. In all tests in this section, the FFT algorithm is used at the local scale (for the generation of the local independent realizations of the random field on each subdomain). Also, 5 grid points are considered per correlation length ℓ_c. The random field generation routines are available freely on GibHub (https://github.com/cottereau/randomField), under a CeCILL-C open-source license. The implementation uses the MPI and HDF5 libraries.

5.1 | A large scale example in 3D

We first prove the efficiency of our algorithm by testing it on a large 3D case. The test is run on 512 processors of the CentraleSupélec–ENS Paris-Saclay supercomputer, with Intel Xeon X7542 processors at 2.66 Ghz. A large domain of size $300\ell_c \times 300\ell_c \times 300\ell_c$ is considered, with close to 2 billion grid points). The pavement of the domain is made with a regular grid of $8\times8\times8$ overlapping subdomains, which means that each subdomain is about $40\ell_c \times 40\ell_c \times 40\ell_c$. The overlap length is $5\ell_c$ of overlap, which corresponds to a negligible error (see Table 2). The random field sample shown in Fig. 3 was produced in only 41 seconds (Figure 3), including I/O which corresponds to an HDF5 file of approximately 16 Gb.

Note that to generate an equivalent random field sample on 1 core of an Intel Xeon CPU E5-4667 v4 at 2.20 GHz, it was necessary to use 256 Gio of RAM and the computation lasted over 45 minutes. Besides these large-shared-memory machines not being so widely spread as distributed-memory machines, the gain in wall time is clearly an interest of the localization technique.

In Fig. 4 , we compare the correlation function numerically obtained from one sample of a Gaussian random field obtained with the classical FFT scheme and the proposal in this paper. In order to present converged results with only one sample, these numerical correlation have been computed from a random field in an elongated 3D box, with largest side equal to 10'000 times the correlation length. These numerical correlations are compared with the theoretical correlation function, which is gaussian. The comparison shows that, as expected, both schemes generate random fields with the target correlation function.

5.2 | Weak Scaling

We now consider the scalability of the algorithm. We therefore consider a sequence of problems of growing size, and matching numbers of processors, such that the numerical cost for one processor remains constant. The ultimate objective is to obtain the same wall time for one problem solved on a given number of processors and a problem twice as large solved on twice as many processors. Although our algorithm is theoretically constructed to do that, the challenge here is that communication between

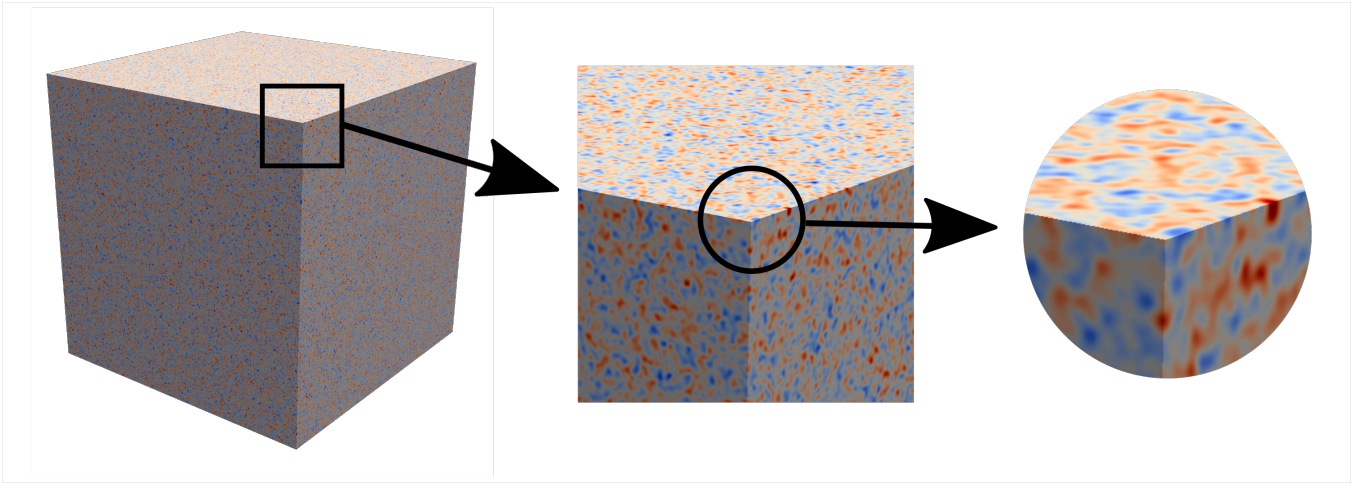


FIGURE 3 Simulation of a random field at 1.7×10^9 DOFs in a $300\ell_c \times 300\ell_c \times 300\ell_c$ domain over 512 Processors.

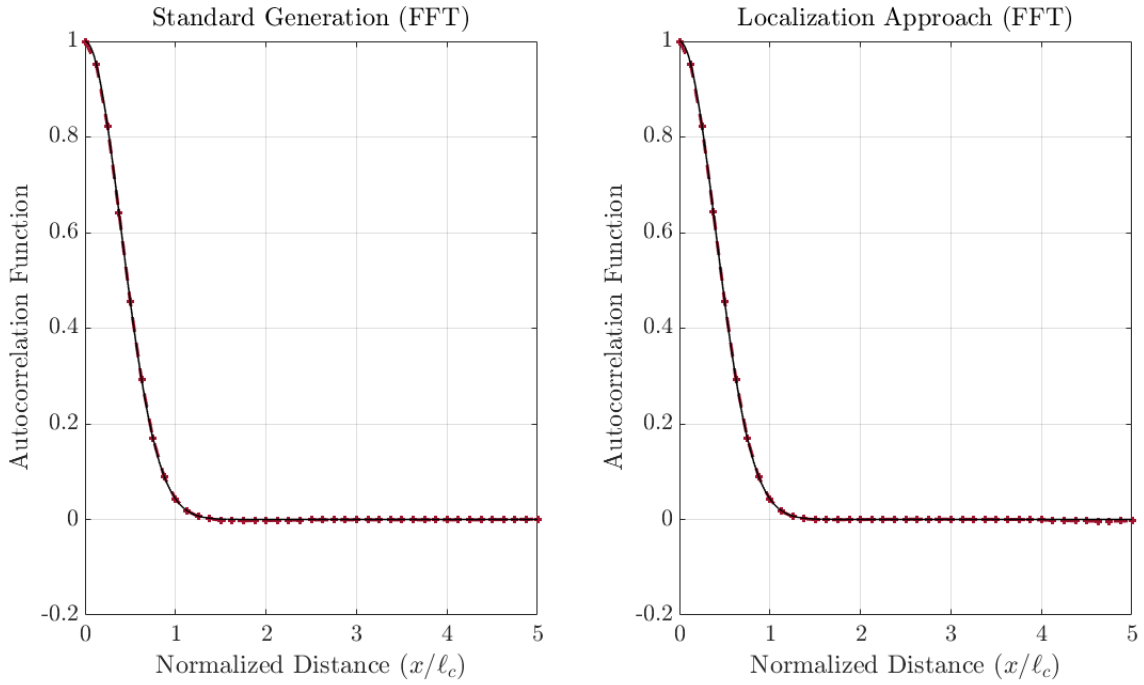


FIGURE 4 Correlation functions estimated from one random field sample generated with the classical FFT method (left figure) and with the proposed method (left figure). The red dashed lines with crosses correspond to the numerical estimation of the correlation function, and the black solid lines represent the theoretical correlation (gaussian).

processors also has to scale linearly. Each simulation was run 10 times to provide an average wall time. Each subdomain is about $30\ell_c \times 30\ell_c \times 30\ell_c$. The overlap length is $5\ell_c$ of overlap. It results in 1.5×10^6 DOFs per subdomain and a total of 5.5×10^8 DOFs for the simulation with 512 CPUs.

Fig. 5 presents the weak scalability results. Results correspond to the generation time, from reading the input to writing the output files. Points indicate the average between 10 simulations while the shaded areas indicate the fluctuation between the slowest and fastest cases. On the same figure are also reported the results using a classical implementation of the FFT algorithm (using FFTW library). The numerical cost for the generation scheme without localization clearly explodes after a while: although the implementation of the FFTW library is rather efficient, at some point, the $\mathcal{O}(N \log N)$ communication and memory issues

become prohibitive for weak scalability. On the other hand, for the localization scheme, we obtain a perfect weak scalability for more than 32 processors, all the way up to 512 processors, which was the maximum number of processors available to us.

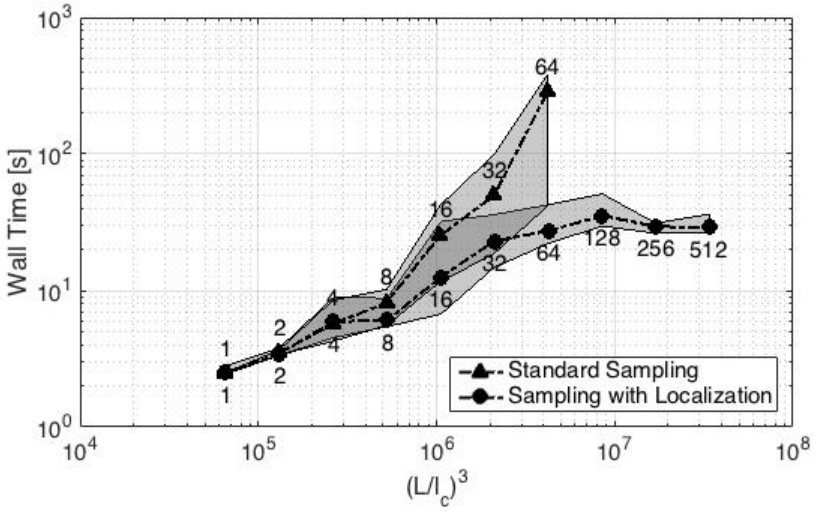


FIGURE 5 3D Weak Scaling without localization and with localization. The number below each point is the number of processors used for that particular simulation. Points indicate the average between 10 simulations while the shaded areas indicate the fluctuation between the slowest and fastest cases.

6 | CONCLUSION

This paper introduced a new parallel technique to generate realizations of Gaussian random fields over very large domains and very large numbers of processors. To our knowledge, no other technique can obtain such excellent scalability or promise to reach sizes compatibles with the largest supercomputers. Scalability is obtained by generating sets of independent realizations over smaller subdomains and merging them in an well chosen manner to retrieve the global statistical properties of the random field. This technique is appropriate in the case where the discretization step h is smaller than the correlation length ℓ_c, which is itself much smaller than the size of the domain L. It has been implemented in a library, freely available on GibHub (https://github.com/cottereau/randomField) under a CeCILL-C open-source license.

ACKNOWLEDGMENTS

This work, within the SINAPS@ and CouESt projects, benefited from French state funding managed by the National Research Agency under program RNSR Future Investments bearing reference No. ANR-11-RSNR-0022-04 and Young Investigators Program No. ANR-14-CE07-0007. This work was performed using HPC resources from the "Mésocentre" computing center of CentraleSupélec and École Normale Supérieure Paris-Saclay, supported by CNRS and the French Minist de l'Enseignement Supeur, de la Recherche, et de l'Innovation (http://mesocentre.centralesupelec.fr/). The first and third authors would also like to thank Alvaro Coutinho and Fernando Rochinha from Universidade Federal do Rio de Janeiro (Brazil), for fruitful interactions on the subject of this paper, and CAPES-COFECUB for funding allowing these interactions through project Ph 822-14 .

References

1. Huyse L., Maes M. A.. Random field modeling of elastic properties using homogenization. *J. Engr. Mech.*. 2001;127(1):27-36.

2. Efendiev Y., Pankov A.. Numerical homogenization of nonlinear random parabolic operators. *SIAM Multiscale Model. Simul.*. 2004;2(2):237-268.

3. Ostoja-Starzewski M.. *Microstructural randomness and scaling in mechanics of materials*. CRC Press; 2007.

4. Le Bris C.. Some numerical approaches for weakly random homogenization. In: *et al.* G. Kreiss, ed. *Numerical Mathematics and Advanced Applications 2009*, :29-45Springer-Verlag; 2010.

5. Cottereau R.. Numerical strategy for unbiased homogenization of random materials. *Int. J. Numer. Meths. Engr.*. 2013;95(1):71-90.

6. Popescu R., Deodatis G., Nobahar A.. Effects of random heterogeneity of soil properties on bearing capacity. *Prob. Engr. Mech.*. 2005;20:324-341.

7. Cottereau R., Clouteau D., Soize C.. Probabilistic impedance of foundation: impact of the seismic design on uncertain soils. *Earthquake Engr. Struct. Dyn.*. 2007;196(17-20):899-918.

8. Phoon K.-K.. *Reliability-based design in geotechnical engineering*. CRC Press; 2008.

9. Alves Fernandes V., Lopez-Caballero F., Costa d'Aguiar S.. Probabilistic analysis of numerical simulated railway track global stiffness. *Comp. & Geotech.*. 2014;55:267-276.

10. Yang Z. J., Su X. T., Chen J. F., Liu G. H.. Monte Carlo simulation of complex cohesive fracture in random heterogeneous quasi-brittle materials. *Int. J. Solids Struct.*. 2009;46:3222-3234.

11. Jehel P., Cottereau R.. On damping created by heterogeneous yielding in the numerical analysis of nonlinear RC frame elements. *Comp. & Struct.*. 2014;. Submitted for publication.

12. Frankel A., Clayton R. W.. A finite-difference simulation of wave propagation in two-dimensional random media. *Bull. Seismological Soc. Amer.*. 1984;74(6):2167-2186.

13. Holliger K.. Upper-crustal seismic velocity heterogeneity as derived from a variety of P-wave sonic logs. *Geophys. J. Int.*. 1996;125(3):813-829.

14. Ta Q.-A., Clouteau D., Cottereau R.. Modeling of random anisotropic elastic media and impact on wave propagation. *Europ. J. Comp. Mech.*. 2010;19(1-3):241-253.

15. Sato H., Fehler M. C., Maeda T.. *Seismic wave propagation and scattering in the heterogeneous earth.* Springer; second ed.2012.

16. Baydoun I., Savin E., Cottereau R., Clouteau D., Guilleminot J.. Kinetic modeling of multiple scattering of elastic waves in heterogeneous anisotropic media. *Wave Motion.* 2014;51(8):1325-1348.

17. Khazaie S., Cottereau R., Clouteau D.. Influence of the spatial correlation structure of an elastic random medium on its scattering properties. *J. Sound Vib.*. 2015;370:132-148.

18. Choi S.-K., Gandhi R. V., Canfield R. A.. *Reliability-based structural design*. Springer; 2006.

19. Clouteau D., Cottereau R., Lombaert G.. Dynamics of structures coupled with elastic media – a review. *J. Sound Vib.*. 2013;332(10):2415-2436.

20. Chorin A.. Gaussian fields and random flow. *J. Fluid Mech.*. 1974;63:21-32.

21. Le Maître O. P., Knio O. M.. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer; 2010.

22. Debusschere B., Najm H. N., Matta A., Knio O., Ghanem R., le Maître O.. Protein labeling reactions in electrochemical microchannel flow: numerical prediction and uncertainty propagation. *Phys. Fluids.* 2003;15(8):2238-2250.

23. Lifshits M. A.. *Gaussian random functions*. Mathematics and its ApplicationsSpringer; 1995.

24. Adler R. J., Taylor J. E.. *Random fields and geometry*. Springer Monographs in MathematicsSpringer; 2007.

25. Rue H.. Fast sampling of Gaussian Markov random fields. *J. Roy. Stat. Soc.*. 2001;B63:325-338.

26. Powell C. E.. Generating realisations of stationary Gaussian random fields by circulant embedding unpublished.

27. Dietrich C. R., Newsam G. N.. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comp.*. 1997;18(4):1088-1107.

28. Aune E., Eidsvik J., Pokern Y.. Iterative numerical methods for sampling from high dimensional Gaussian distributions. *Stat. Comput.*. 2013;23(4):501-521.

29. Chow E., Saad Y.. Preconditioned Krylov subspace methods for sampling multivariate Gaussian distributions. *SIAM J. Sci. Comp.*. 2014;36(2):A588-A608.

30. Panunzio A. M., Cottereau R., Puel G.. Large scale random fields generation using localized Karhunen-Loève expansion. *Adv. Model. Simul. Engr. Sci.*. 2018;. Submitted for publication.

31. Shinozuka M., Jan C.-M.. Digital simulation of random processes and its applications. *J. Sound Vib.*. 1972;25(1):111-128.

32. Shinozuka M., Deodatis G.. Simulation of stochastic processes by spectral representation. *Appl. Mech. Rev.*. 1991;44(4):191-205.

33. Pardo-Igúzquiza E., Chica-Olmo M.. The Fourier integral method: an efficient spectral method for simulation of random fields. *Math. Geol.*. 1993;25(2):177-217.

34. Frenje L., Juhlin C.. Scattering attenuation: 2-D and 3-D finite difference simulations vs. theory. *J. Appl. Geophys.*. 2000;44(1):33-46.

35. Imperatori W., Mai P. M.. Broad-band near-field ground motion simulations in 3-dimensional scattering media. *Geophys. J. Int.*. 2013;192(2):725-744.

36. Shields M. D., Kim H.. Simulation of higher-order stochastic processes by spectral representation. *Prob. Engr. Mech.*. 2017;47:1-15.

37. Rosenblatt M.. Remarks on a multivariate transformation. *Annals Math. Stat.*. 1952;23(3):470-472.

38. Grigoriu M.. Simulation of stationary non-gaussian translation processes. *J. Engr. Mech.*. 1998;124(2):121-126.

39. Puig B., Akian J.-L.. Non-gaussian simulation using Hermite polynomials expansion and maximum entropy principle. *Prob. Engr. Mech.*. 2004;19(4):293-305.

40. Bogdanoff J. L., Goldberg J. E., Bernard M. C.. Response of a simple structure to a random earthquake-type disturbance. *Bull. Seismological Soc. Amer.*. 1961;51(2):293-310.

41. Rice S. O.. Mathematical analysis of random noise. *Bell Labs Tech. J.*. 1945;24(1):46-156.

42. Lévy P.. *Processus stochastiques et mouvement brownien*. Gabay; 1948.

43. Pierson W. J.. Wind generated gravity waves. *Adv. Geophys.*. 1955;2:93-178.

44. Grigoriu M.. On the spectral representation method in simulation. *Prob. Engr. Mech.*. 1993;8(2):75-90.

45. Shinozuka M., Deodatis G.. Simulation of multi-dimensional Gaussian stochastic fields by spectral representation. *Appl. Mech. Rev.*. 1996;49(1):29-53.

46. Borgman L. E.. Ocean Wave Simulation for Engineering Design. *J. Waterways Harbors Div. ASCE.* 1969;95(4):557-586.

47. Goto H., Toki K.. Structural response to nonstationary random excitation. In: :130-144; 1969.

48. Shinozuka M.. Simulation of multivariate and multidimensional random processes. *J. Acoust. Soc. Amer.*. 1971;49(1):357-368.

49. Cameron C.. Relative efficiency of Gaussian stochastic process sampling procedures. *J. Comp. Phys.*. 2003;192(2):546-569.

50. Kramer P. R., Kurbanmuradov O., Sabelfeld K.. Comparative analysis of multiscale gaussian random field simulation algorithms. *J. Comp. Phys.*. 2007;226:897-924.

51. Kurbanmuradov O., Sabelfeld K., Kramer P. R.. Randomized spectral and Fourier-wavelet methods for multidimensional Gaussian random vector fields. *J. Comp. Phys.*. 2013;245:218-234.