6th International Conference on Earthquake Geotechnical Engineering
1-4 November 2015
Christchurch, New Zealand

# Efficient Parallel Generation of Random Field of Mechanical Properties for Geophysical Application

L. Paludo[1], V. Bouvier[2], L. Corrêa[3], R. Cottereau[4], D. Clouteau[5]

**ABSTRACT**

We study the generation of random fields of mechanical properties for problems where the domain is much larger than the characteristic distance over which the properties fluctuates. This statistical description is sufficient if we are interested in higher frequency signals or the seismic coda. We expose three generation methods: Spectral Method, Randomization and a variant of the Spectral Method for isotropic media. Preliminary numerical results indicate that the random generation step with these methods becomes a numerical bottleneck for geophysical problems at today' state-of-the-art size. We address this scalability issue by dividing the domain in independent overlapping subsets. The proposed approach has the potential to remove that bottleneck.

## 1 Introduction

The Earth crust presents heterogeneities at several scales. Their modeling is necessary when we are interested in studying the seismic coda or higher frequency signals (Aki and Chouet, 1975). Unfortunately, the complete description of the medium requires an enormous amount of parameters. The stochastic description of those parameters may provide an interesting alternative. In particular, this approach is appealing when asymptotic regimes are considered, such as homogenization (Capdeville et al., 2010) or weak scattering regime (Ryzhik et al., 1996). In these regimes, the full description of the parameters is not mandatory since the solution of the mechanical problem depends only on some statistics of those parameters.

The equation describing elastic wave propagation in elastic (non-dissipative) media can be expressed as:

$$\rho(\boldsymbol{x})\frac{\partial^2 v}{\partial t^2}(\boldsymbol{x},t) - \nabla_{\boldsymbol{x}}\left\{C(\boldsymbol{x}) : \nabla_{\boldsymbol{x}} \otimes v(\boldsymbol{x},t)\right\} = 0, \qquad (\boldsymbol{x},t) \in \Omega \times \mathbb{R} \tag{1}$$

where $\rho(\boldsymbol{x})$ is the medium density, $C(\boldsymbol{x})$ is the fourth-order elastic tensor, and $v(\boldsymbol{x},t)$ is the displacement field. The material behaviour is often considered isotropic and parameterized by the P-wave velocity, S-wave velocity and density. However, in the Earth's crust, the composite mineralogy and the presence of fractures of various sizes induce large fluctuations of these

[1]Laboratoire MSSMat - CentraleSupélec, France, luciano.de-carvalho@ecp.fr
[2]Laboratoire MSSMat - CentraleSupélec, France, victor.bouvier@student.ecp.fr
[3]Laboratoire MSSMat - CentraleSupélec, France, lucio.de-abreu-correa@ecp.fr
[4]Laboratoire MSSMat - CentraleSupélec, France, regis.cottereau@ecp.fr
[5]Laboratoire MSSMat - CentraleSupélec, France, didier.clouteau@ecp.fr

parameters. To take into account media heterogeneity, we assume $C = \{C(\boldsymbol{x}) : \boldsymbol{x} \in \Omega\}$ is a stochastic random field. This description is able to generate 3D elasticity tensor fields even of a random anisotropic material (Ta et al., 2010).

There are several methods to compute random fields with a given correlation function, basically falling into two categories: (i) direct approaches and (ii) spectral approaches. In the former case, the generation of a realization of a random field is performed in the space domain usually using the Cholesky factorization to determine the square root of the covariance matrix $R$ (Rue, 2001). The computational cost scales as $O(N^3)$ in the general case, but can be improved to sub-$O(N^2)$ using a polynomial approximation of $R^{1/2}$ (Chow and Saad, 2014). Nevertheless, the covariance matrix is often sparse and circulant and factorization algorithms can be optimized for this particular case (Dietrich and Newsam, 1997). Alternatively, the random field can be simulated using a spectral approach. The factorization is then performed on the Power Spectral Density (Fourier transform of the covariance). One such method was introduced by M. Shinozuka and G. Deodatis and is called Spectral Representation Method (Shinozuka and Deodatis, 1991). Another spectral method relies on the Monte Carlo method and is called Randomization Method (Cameron, 2003; Kramer et al., 2007; Kurbanmuradov et al., 2013). A third possibility is to consider a mix of the previous two methods in discretizing the spectrum amplitude as in the spectral representation method, and the angle as in the Monte Carlo method. It is available only for isotropic media and is therefore called the Isotropic Spectral Method.

Seismic wave propagation problems are now routinely performed over hundreds or thousands of cores (Komatitsch et al., 2002). It is then necessary to generate samples of the random fields of parameters on very large scales, in particular when the correlation length is small compared to the wave length or the propagation length. An important requirement is that the sample generation cost (CPU time and memory) should remain small compared to the simulation time. Preliminary numerical results will show that this is not the case with the methods described above. To mitigate scalability issues we propose to treat the problem as a set of smaller independent problems, gluing them together through transition volumes. To generate a sample the statistical inputs are: first order marginal density, correlation model, correlation length, average and standard deviation. The only information communicated between the processors is the seed for random number generation. It ensures the $\mathcal{C}^\infty$ regularity continuity of the generated fields while minimizing the communications.

## 2 Stochastic field generation

In this paper we are interested in the case of large domains. We mean by large that the dimension of the domain $L$ is much larger than both the correlation length $\ell_c$ (or some characteristic size over which the fluctuations of the random field are significant) and the discretization step $h$. If the size of the domain is small compared to the correlation length, 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. If the discretization step is much larger than the correlation length, the sampling becomes simple and numerically inexpensive. Indeed, for the mesh considered, the random field is essentially a white noise with Gaussian first-order marginal density. We therefore restrict our attention in this paper to the case where $h < \ell_c \ll L$.

We only consider here the sampling of standard Gaussian fields because they are the basic building block of a large number of numerical schemes. The first-order marginal density can be

modified locally by combining a direct and inverse Rosenblatt transforms (Rosenblatt, 1952), although one has to pay attention to the influence on the correlation function of the resulting random field (Grigoriu, 1998; Puig and Akian, 2004). Therefore we want to generate a random field $u$ that follows three assertions : $(i)$ $u$ is a standard gaussian field with a given correlation function $R$, $(ii)$ $u$ is ergodic and $(iii)$ $u \in \mathcal{C}^p(\Omega)$ almost surely for a given $p \in \mathbb{N}$. In the particular case when $R$ is only a function of $y-x$, $u$ it is called a stationnary process and when it depends only of $|| y-x ||$, $u$ is called an isotropic process. The ergodicity hypothesis is compulsory when each generated sample should represent well the required statistics.

A common approach to sample a random field $\{u(\boldsymbol{x}) : \boldsymbol{x} \in \Omega \subset \mathbb{R}^d\}$ with a given correlation function $R$ is to search it as a linear combination of independent and identically distributed random variables, where $\Omega \subset \mathbb{R}^d$ is the domain and $d$ is the number of dimensions of space. The spectral representation is a classic way to sample gaussian random field (Shinozuka and Deodatis, 1991) :

$$u(\boldsymbol{x}) = \int_{\boldsymbol{k}\in\Omega} \hat{R}^{1/2}(\boldsymbol{k}) \exp(i\boldsymbol{k}\cdot\boldsymbol{x}) \mathrm{d}W(\boldsymbol{k}) \quad (\boldsymbol{x} \in \Omega) \tag{2}$$

where $\{W(\boldsymbol{k}) : \boldsymbol{k} \in \Omega\}$ is a Brownian motion, $\hat{R}$ is the Fourier transform of $R$ and $\boldsymbol{k}\cdot\boldsymbol{x}$ the inner product between $\boldsymbol{k}$ and $\boldsymbol{x}$. There are several methods in the literature to compute the stochastic integral (2). In the next sections we will introduce the Spectral Method, Randomization and a variant of the Spectral Method for isotropic media.

### *2.1 Spectral Method*

The spectral method by M. Shinozuka and G. Deodatis (Shinozuka and Deodatis, 1991), proposes the quadrature (3) :

$$u^{S.M.}(\boldsymbol{x}) = \sum_{n=0}^{N} \hat{R}^{1/2}(\boldsymbol{k}_n) \exp(i\boldsymbol{k}_n\cdot\boldsymbol{x}) \sqrt{\Delta_n}\xi(n) \quad (\boldsymbol{x} \in \Omega) \tag{3}$$

where $\xi = \{\xi(n) : n \leq N\}$ is a white noise, $\boldsymbol{k}_n \in \Omega_n$ for all $n \leq N$, $(\Omega_n)_{0\leq n\leq N}$ is a partition of $\Omega$ and $\Delta_n$ is the Lebesgue measure of $\Omega_n$. This representation ensures the $\mathcal{C}^\infty$ regularity on $\Omega$ of the random field $u$ and decouples $u(\boldsymbol{x}_1)$ and $u(\boldsymbol{x}_2)$ for $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$. Nevertheless, some conditions must be respected when using the Fourier transform in discrete spaces. Assuming $\Omega = [0, L]^d$ where $L$ is the domain size, we define $\Delta x = \frac{L}{M}$ and $\Delta k$ respectively the discretization steps in space and wave number space. To avoid the field periodicity we must compute $\Delta k \leq \frac{2\pi}{L}$. It generates a dependence between the number of points in the spectral space and the domain size $L$. As a result, when generating $u$ over a large domain or a refined mesh, computational cost grows rapidly.

### *2.2 Randomization*

Another classic way to compute (2) is to consider it as the expectancy of a random variable $\exp(i\boldsymbol{K}.\boldsymbol{x})$ : $\int_{k\in\Omega} \hat{R}(\boldsymbol{k}) \exp(i\boldsymbol{k}.\boldsymbol{x}) \mathrm{d}\boldsymbol{k} = \mathbb{E}[\exp(i\boldsymbol{K}.\boldsymbol{x})]$ where $K$ follows the probability density $\hat{R}(\boldsymbol{k})$. It is called the Randomization Method (Kramer et al., 2007; Kurbanmuradov et al., 2013) :

$$u^{R\cdot}(\boldsymbol{x}) = \frac{1}{\sqrt{N_r}} \sum_{n=0}^{N_r} \xi(n) \exp(i\boldsymbol{k}_n \cdot \boldsymbol{x}) \quad (\boldsymbol{x} \in \Omega) \tag{4}$$

where $(\boldsymbol{k}_n)_{n \leq N_r}$ is a set of $N_r$ realisations of $\boldsymbol{K}$. The Randomization Method does not introduce aliasing or periodicity ; there is no condition on $N_r$ involving $\Delta x$ or $L$. On the other hand sampling randomly the Fourier space doesn't guarantee that we represent accurately the spectrum we want. It is a major drawback of this method given that we would like to rely on one single realization to represent the properties field.

### *2.3 Isotropic Spectral Method*

When considering isotropic fields one can reduce the complexity of Spectral Representation method from $O(N^2)$ to $O(N^{1+1/d})$ using spherical coordinates to describe the vector $\boldsymbol{k}$. We choose randomly the two angles $\theta_n$ and $\phi_n$ that define the direction of $\boldsymbol{k}_n$ and his norm $r_n$ deterministically. The deterministic radius $(r_n)_{n \leq N_r}$ assures that we explore all the spectrum and the random direction reduces the integral from a volume to a line with no further drawback. $(\boldsymbol{k}_n)_{n \leq N_r}$ is defined as :

$$\boldsymbol{k}_n = r_n \{\cos(\theta_n)\sin(\phi_n), \sin(\theta_n)\sin(\phi_n), \cos(\phi_n)\}^T \quad (r_n \in \mathbb{R}^+) \tag{5}$$

where $\{\theta_n : n \leq N_r\}$ and $\{\phi_n : n \leq N_r\}$ are respectively white noises in $[0, 2\pi]$ and $[0, \pi]$.

$$u^{I.S.\cdot}(\boldsymbol{x}) = \sum_{n=0}^{N_r} \sqrt{\hat{R}(r_n) r_n \sin(\phi_n) \Delta_n} \exp(i\boldsymbol{k}_n \cdot \boldsymbol{x}) \xi(n) \quad (\boldsymbol{x} \in \Omega) \tag{6}$$

## 3 Numerical observation of scaling

Dealing with problems with many degrees of freedom is very demanding and we need to use many cores to perform calculations. As we are looking for a sampling method suited to this context, we perform a numerical scaling test on the presented methods. We started calculations with a cube of volume $2^{10}$ m$^3$ on 1 processor. At each iteration we double size the volume and use twice as much processors (weak scaling). The correlation length is $l_c = 1$ m and the mesh is structured with a step of $\Delta x = \frac{l_c}{10}$ in all three directions. The results are in Figure 1.

For each method the measured time was normalized with respect to the time taken by one single processor. On the line graph we see that the wall time grows almost exponentially at each iteration. It reveals that none of the presented method has a good scalability. In other words, as the domain grows bigger the cost of generation per volume increases. It is caused by the link between the size of the domain and the number of elements needed in the wave number domain. For the Randomization method this link is made when we require a given accuracy and, as we can see in Figure 1 it makes its scaling very similar to the Spectral Method.

## 4 Localization of the sampling

As the domain becomes larger the computational cost of generating a sample grows rapidly and without threshold. To bound this overflowing instead of performing the whole domain at once,

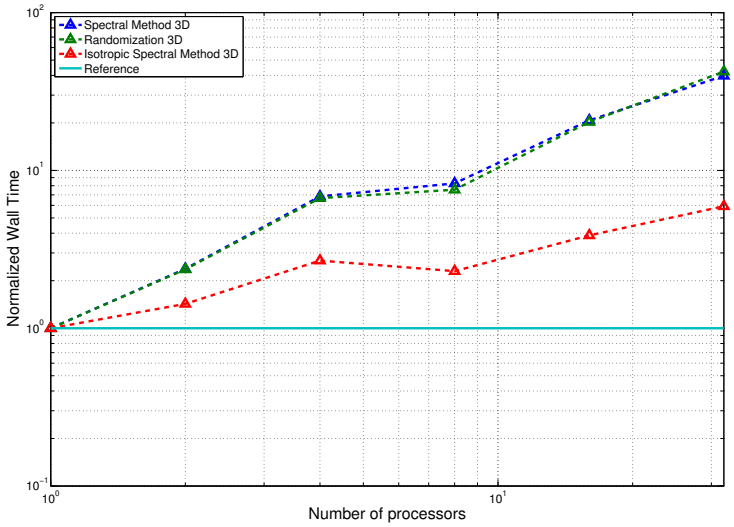


Figure 1: Weak scaling behavior for each generation method

it could be interesting to sample over several smaller subdomains. The issue here is how to ensure regularity between the fields generated on different subdomains. We address this problem by making a transition overlapping volume between subdomains.

Points separated by a distance larger than the correlation length are, by definition, uncorrelated. In the algorithms presented so far the mutual contribution of every point on the grid was considered. The idea now is to bound the number of operations needed to generate the sample based only on the size of the field over each processor and not the global size. It allows us to keep the number of operations per processor constant, even when $\frac{L}{l_c} \gg 1$. We subdivide the domain in smaller independent parts $\Omega_i$ with a partition of unity $\psi = (\psi_i)_{i\in I}$ of $\Omega : \sum_{i\in I} \psi_i(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \Omega$. We write the random field as:

$$u_{Loc}(\boldsymbol{x}) = \sum_{i\in I} \sqrt{\psi_i(\boldsymbol{x})} u_{\Omega_i}(\boldsymbol{x}) \quad (\boldsymbol{x} \in \Omega) \tag{7}$$

where $u_{\Omega_i}$ is a localized sample of $u_{Loc}$ over the subdomain $\Omega_i$ and for $i \neq j, \psi_i = 0$ over $\Omega_j$. With this decomposition the mean, variance an correlation function are now:

$$\mathbb{E}[u_{Loc}(\boldsymbol{x})] = \mathbb{E}\left[\sum_{i\in I} \sqrt{\psi_i(\boldsymbol{x})} u_{\Omega_i}(\boldsymbol{x})\right] = \sum_{i\in I} \sqrt{\psi_i(\boldsymbol{x})}\, \mathbb{E}[u_{\Omega_i}(\boldsymbol{x})] = 0 \quad (\boldsymbol{x} \in \Omega) \tag{8}$$

$$\mathbb{E}[u^2_{Loc}(\boldsymbol{x})] = \mathbb{E}\left[\left(\sum_{i\in I} \sqrt{\psi_i(\boldsymbol{x})} u_{\Omega_i}(\boldsymbol{x})\right)^2\right] = \sum_{i\in I} \psi_i(\boldsymbol{x})\, \mathbb{E}[u^2_{\Omega_i}(\boldsymbol{x})] = 1 \quad (\boldsymbol{x} \in \Omega) \tag{9}$$

$$R_{Loc}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i\in I}\sum_{j\in I} \sqrt{\psi_i(\boldsymbol{x})}\sqrt{\psi_j(\boldsymbol{y})_j}\, \mathbb{E}\left[u_{\Omega_i}(\boldsymbol{x}) u_{\Omega_j}(\boldsymbol{y})\right] = \sum_{i\in I} \sqrt{\psi_i(\boldsymbol{x})\psi_i(\boldsymbol{y})}\, R(\boldsymbol{x},\boldsymbol{y}) \quad (\boldsymbol{x},\boldsymbol{y} \in \Omega) \tag{10}$$

We can see that field variance and average remain unchanged but the correlation function has a multiplying factor of $\sqrt{\psi_i(\boldsymbol{x})\psi_i(\boldsymbol{y})}$ when compared to the original function. The sum of all multiplying factors goes to one if $\psi_i(\boldsymbol{x}) = \psi_i(\boldsymbol{y})$. It means that the approximation is good as long as the partitions of unity $\psi$ vary slowly (in a larger scale) compared to $R$. This have to be taken into account when chosing the size of the overlap volume. An example is shown in Figure 2.

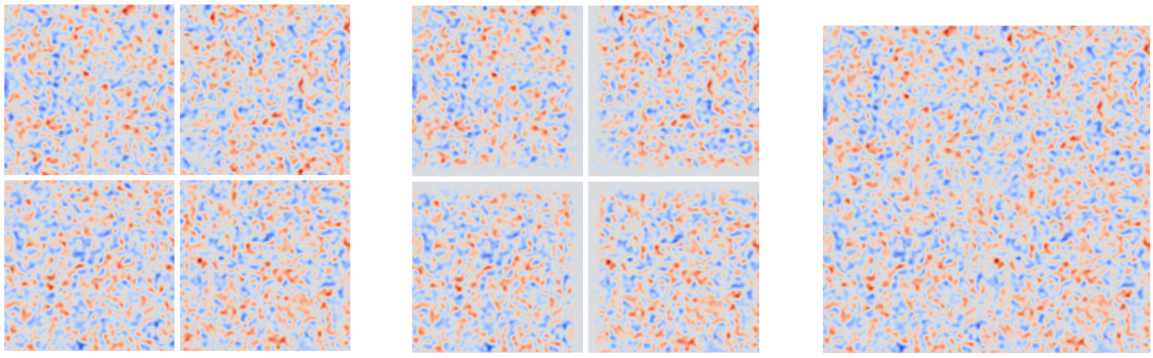

Figure 2: Generation of four independent gaussian random fields (Left), action of the decomposition functions (Center) and resultant field after merge (Right).

An analysis of Figure 3 reveals the power of this method to mitigate scalability issues. The parameters are the same used to make the Figure 1 (initial volume $= (2^{10})$ m$^3$, $l_c = 1$m, $\Delta x = \frac{l_c}{10}$). The transition volume is equal to $5l_c$.

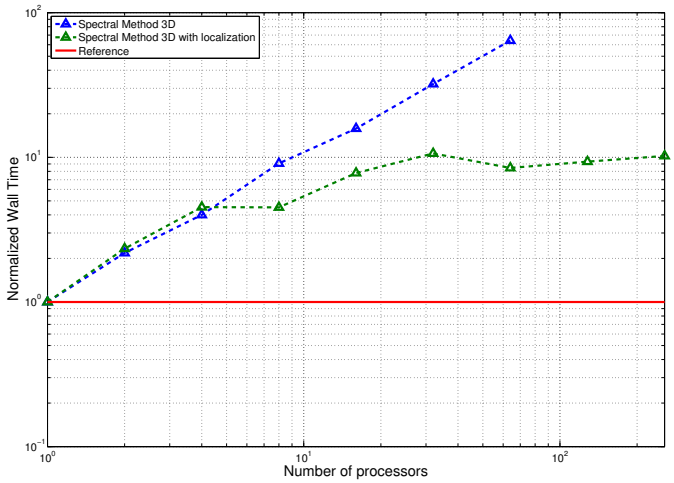


Figure 3: Weak scaling behavior comparison for the Spectral Representation Method

We observe that for few processors the extra cost of calculating the overlapping volumes makes computation cost with the localization approach slightly bigger . As we go further in the number of processors we see that the cost of calculation becomes much smaller and then stabilizes. It suggests that this method allows to calculate large domains efficiently as we can keep the time taken by each processor constant. It should be noted that the wall time where we reach stabilization can be diminished if we improve the algorithm. For instance Fast Fourier Trans-

form can be used to calculate the Spectral Method if we limit our interest to structured meshes. Anyway, finding this stabilization area is a direct contribution of the localization method.

## 5 Simulation

First we compare the displacement-log in a simulation using first homogeneous media and then randomly heterogeneous media. For each property the statistical inputs are: first order marginal density, correlation model, correlation length, average and standard deviation. We used the Spectral Method to generate samples in a cube of side 500m and correlation length 40m. The source excitation is along the Y-axis. We take the displacement field along a line that passes through the source aligned with the X-axis. The statistical properties are in Table 1 and displacement log in both cases is shown in Figure 4.

Table 1: Statistical parameters used in simulation.

| | Density | Lambda | Mu |
|---|---|---|---|
| Average | $2800\frac{kg}{m^3}$ | $1092.7\ 10^5$ | $1120.0\ 10^5$ |
| Variance | $6\ 10^5$ | $1\ 10^{16}$ | $1\ 10^{16}$ |
| First Order Marginal | Lognormal | | |
| Correlation Size | 40m | | |

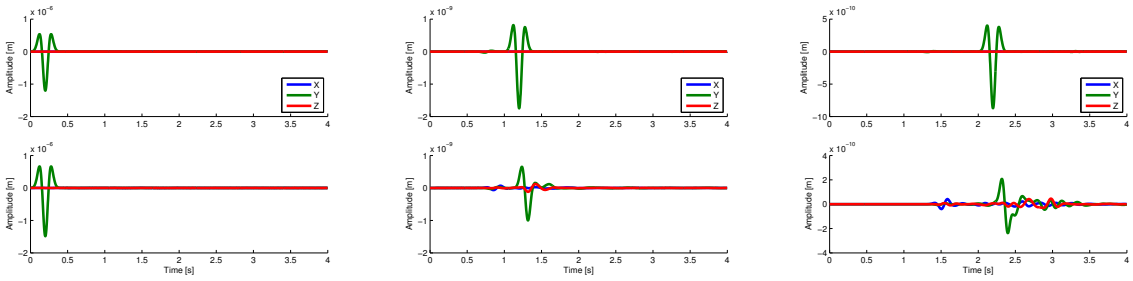


Figure 4: Displacement log in three points: in the source (left), 200m from the source (center), 400m from the source (right). In each log the homogeneous case is on top and random heterogeneous at the bottom

We can see that in the homogeneous case the displacement field remains unchanged. On the other hand in the heterogeneous case the displacement field changes a lot with the distance from the source. As expected wave scattering generates displacement in axis other than Y.

In the second simulation our domain is the Greek island of Argostoli. We generated with the Isotropic Spectral Method samples to create a 3D elasticity tensor fields of a random isotropic material. We can see one realization of a density field on Figure 5. The problem was performed in a 8,2 millions points unstructured mesh. The correlation length, $l_c$, is 1km and the domain size is 100km $\times$80km $\times$15km.

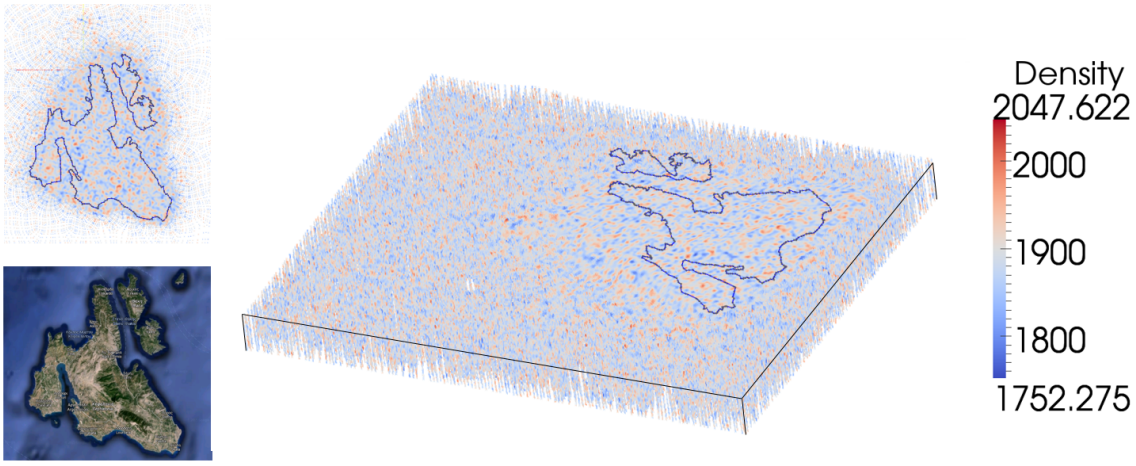


Figure 5: Random density field generated with the isotropic spectral method.

Using the density tensor of Figure 5 and similarly fluctuating properties of the P-wave and S-wave velocities, we performed a wave propagation simulation in a spectral element code. Results are shown in Figure 6 The generation of the properties took about 2% of the total calculation time.

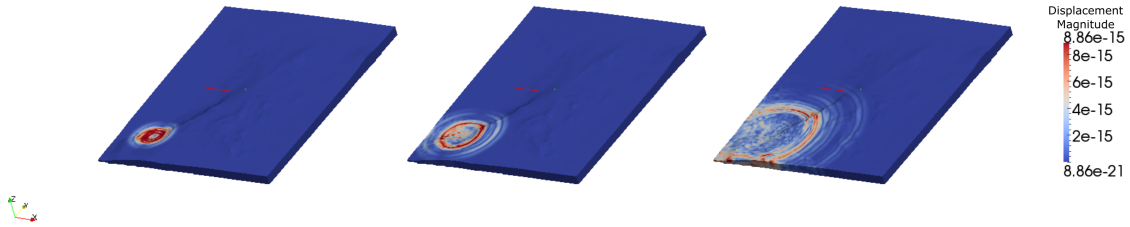


Figure 6: Three snapshots of wave propagation simulation in the Greek island Argostoli.

## 6 Conclusion

When generating random fields in domains where the correlation length is small compared to the wave length or the propagation length the sampling step can become a bottleneck. We showed three sampling approaches in the spectral domain and conclude that the existing methods do not present a good scalability. We addressed this scalability issue by localizing the sample. The localization of the sample allows to generate several independent random fields and combine them in a continuous field. Results and theory have shown that a transition volume of 5 to 10 $l_c$ is enough to make statistics homogeneous over the whole domain. Note that we are interested in cases where $L$ is hundreds to thousands times $l_c$, thus the cost of calculating the overlapping volumes is very little compared to the whole generation. Although, the numerical tests have to be pushed further, there seems to be a possibility to remove the scientific issue of large scale simulation of random fields.

## 7 Acknowledgements

This work, within the SINAPS@ project, benefited from French state funding managed by the National Research Agency under program RNSR Future Investments bearing reference No. ANR-11-RSNR-0022-04.

## References


K. Aki and B. Chouet. Origin of coda waves: source, attenuation, and scattering effects. *J. Geophys. Res.*, 1975, 80(23):3322–3342.

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

Y. Capdeville, L. Guillot, and J.-J. Marigo. 2D non-periodic homogenization to upscale elastic media for P-SV waves. *Geophys. J. Int.*, 2010, 182(2):903–922.

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

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

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

Dimitri Komatitsch, Jeroen Ritsema, and Jeroen Tromp. The spectral-element method, Beowulf computing, and global seismology. *Science (New York, N.Y.)*, 2002, 298(5599):1737–42.

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

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

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

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

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

L. Ryzhik, G. Papanicolaou, and J. B. Keller. Transport equations for elastic and other waves in random media. *Wave Motion*, 1996, 24:327–370.

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

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