Sensitivity analysis of synthetic seismograms in sedimentary basin with respect to uncertain seismological parameters

Florent De Martin, Philippe Thierry, David Keyes, Emmanuel Chaljub, Fabrice Dupros, Faïza Boulahya, Pierre Sochala, Abed Benaichouche, Jérémy Rohmer, Gabriele Piaciucci, et al.

Introduction

Physics-based three-dimensional (3D) numerical simulations are becoming more predictive and have already become essential in improving understanding of natural phenomena, such as earthquakes and tsunamis, flooding, extreme weather, and so on. High performance computing (HPC) now offers the spatial resolution finer than the data gathered in situ to build the models used by the simulations. As a consequence, better understanding the role and impact of epistemic uncertainties (linked to ignorance of the model input parameters) is now a crucial topic in exploiting numerical predictions.

This study focuses on the understanding of the variability and sensitivity of synthetic earthquake ground motions at sedimentary basin scale associated with the epistemic uncertainties of the model of seismic wave propagation. The key question at stake is the following: What is the spatiotemporal variability and sensitivity of seismograms at the surface with respect to seismological parameters? (e.g., shear wave velocity of a specific soil layer, Earth's velocity or attenuation structures, sedimentary basin geometry, etc.)

The model of wave propagation, whereby the variability of earthquakes ground motions will be quantified, is the one designed for the Euroseistest Verification and Validation Project (E2VP). E2VP is an international collaborative project organized jointly by: the Aristotle University of Thessaloniki, Greece; the ITSAK (Institute of Engineering Seismology and Earthquake Engineering of Thessaloniki), Greece; the Cashima research project (supported by CEA – the French Alternative Energies and Atomic Energy Commission – and by ILL – the Laue-Langevin Institute, Grenoble); and ISTerre at Grenoble Alpes University, France. The E2VP target site is the Mygdonian basin near Thessaloniki, Greece, which is the international research and test site of many international seismological and earthquake-engineering projects (Pitilakis *et al*. 2013). To foster the use of linear 3D numerical simulations in practical prediction, E2VP aimed at (a) evaluating the accuracy of the most-advanced numerical methods when applied to realistic 3D models and (b) providing an objective, quantitative comparison between recorded earthquake ground motions and their numerical predictions. So far, these two objectives (a) and (b) have been thoroughly tackled; they are published in Chaljub *et al*. (2015) and Maufroy *et al*. (2015). As a continuation of the E2V project, the source-related variability of the site is presented by Maufroy *et al*. 2017; and this study investigates the variability and sensitivity of earthquake ground motions associated with the epistemic uncertainties of the E2VP model.

Big data for seismic hazard

HPC is actually facing key challenges on both the hardware and software sides. Every vendor is actually working on next generation of machines that will drive the community to the 'Exascale' and beyond. Whatever the computing devices we may consider, the total energy consumption is a challenge that will get down to the pico-Joules scale for any operations or data movement. We all know that those machines will reach several millions of cores and that the scalability of the applications will become a tremendous bottleneck to maintain a decent performance and power efficiency at scale. But instead of targeting pure application scalability up to this huge number of cores, there is also the opportunity to work on a more probabilistic approach to take advantage of the coming computers: in our particular case, where the whole application – including initialization, i/o and communications – can scale up to several thousands of cores, why not starting several simulations at the same time, using a reduced number of cores, but using different input parameters to finally conduct a probabilistic analysis of the results also known as uncertainties quantification (UQ). In other words, why not starting 1,000 simulations using 1,000 cores to reach the millions of cores instead of a single 1,000,000 cores simulation with a far less scalability?

In addition to the challenging geophysical and mathematical problems behind this simple description, we can easily see the forthcoming HPC problems as the programming model that has to include several parallelization levels (MPI + OMP), vectorization for SIMD, intensive i/o (standard parallel file systems vs. Hadoop/Spark). In figure 1, we show the description of the whole concept including pre- and post-processing around the main earthquake simulation engine. The different stages can be viewed as follow: 1° definition of the initial model perturbation to generate a given set of input parameters, 2° simulation including a real time (i.e., runtime) filtering and a post-simulation filtering and decimation to reduce the amount of output data, 3° UQ analysis on parallel file system including Hadoop evaluation, and multi-level MPI communicators to obtain the final results of the global big data application.

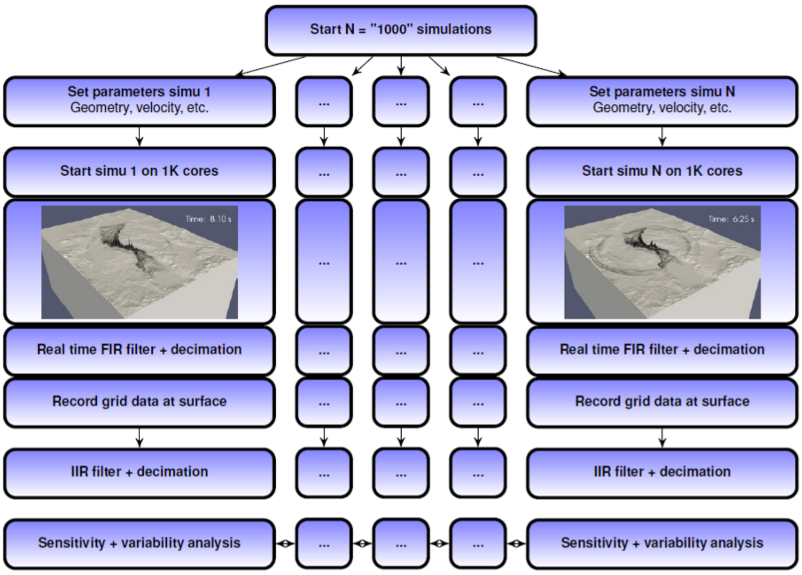

Figure 1 *Concept of the runs that led to a probabilistic approach for seismic hazard targeting millions of core scalability.*

Methods

We are using and developing a 3D seismic wave field simulation code, so called EFISPEC3D (De Martin 2011. http://efispec.free.fr), based on the spectral-element method (SEM) (Maday and Patera 1989; Komatitsch and Vilotte 1998). Our implementation is being widely used for computing seismic wave propagation in complex geological media (De Martin *et al*. 2013; Matsushima *et al*. 2014; Chaljub *et al*. 2015; Maufroy *et al*. 2015). An animation of wave propagation performed for this study is available at http://efispec.free.fr/videos/e2vp2.shaheenII.mp4.

Using the Shaheen II CrayXC40 computer at KAUST, we performed independent simulations to check the stages 1° and 2° of our methodology but also to efficiently prepare the step 3° concerning the data management and transposition needed by the UQ processing (Figure 1 "sensitivity + variability analysis"). The UQ processing of the 3D wave propagation will be tackled by the Monte Carlo (MC) estimates and by the model reduction spectral methods. Among spectral methods, we will give priority to the polynomials chaos method which is widely used for several years and has shown advantageous benefits such as its quick convergence for regular process (Le Maître and Knio 2010; Sochala and Le Maître 2013). First results of UQ processing of seismic 1D wave propagation have been achieved independently by Sochala and De Martin, *submitted*.

Concerning the HPC side, to get to the optimal number of cores to be used for each simulation, we did both strong and weak scaling analyses. The strong scaling exhibits a good speedup up to 8,000+ cores (Intel e5-2698v3) on Shaheen II (see Figure 2). The loss of a factor two inside the node (i.e., 2 CPU sockets) is probably due to a saturation of the QPI link, but further investigations would be needed to confirm this hypothesis.

For each simulation, we save the time series of the three components of the ground motion at every Gauss-Lobatto-Legendre (GLL) points (used as receivers) of the free surface (i.e., 12,297,150 GLL points). During one simulation of 30 seconds of seismogram (150,001 time steps), the time series at each receiver are low-pass filtered by a finite-impulse response (FIR) to avoid time aliasing and saved every 10 time steps to get 2.1 TB of seismograms (instead of 21 TB without the FIR filter). This writing operation is optimized on the Lustre file system following the tuning strategy provided by Paciucci *et al*. 2016. At the end of each simulation, the receivers' time series are low-pass filtered by an infinite impulse response (Butterworth design) and decimated by a factor 15, directly in EFISPEC3D. As a consequence, after the finite and infinite impulse responses filtering and decimation, the total disk space is decreased by a factor 10x15 = 150. The file containing the 12,297,150 three-component seismograms weights 137 GB.

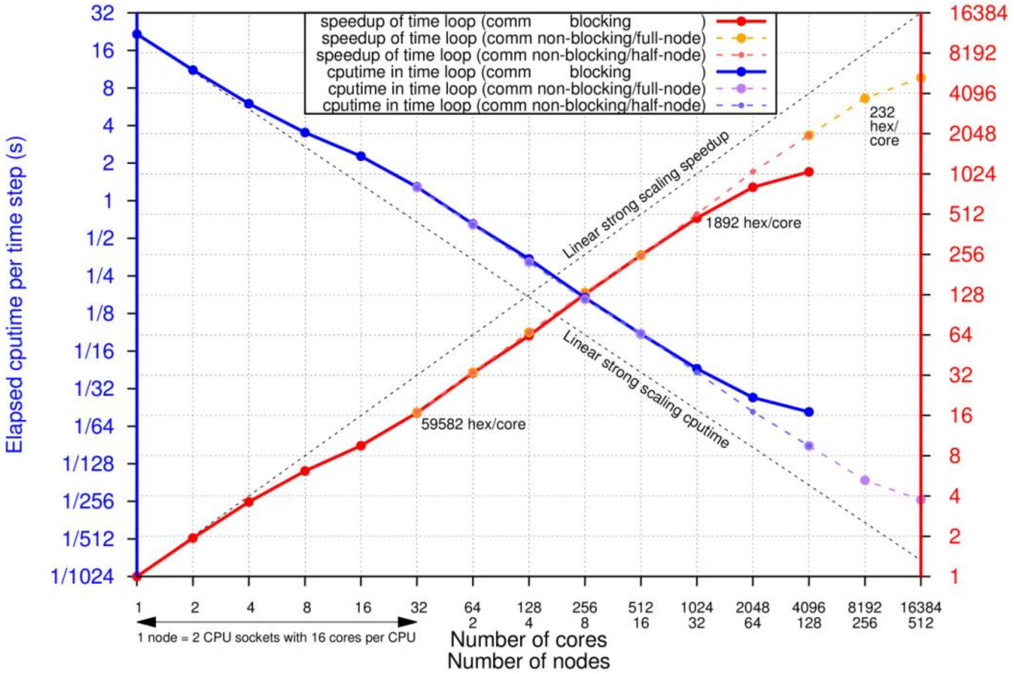

Figure 2 *Strong scaling speedup (red curves) and strong scaling cputime (blue curves) of EFISPEC3D obtained on the supercomputer ShaheenII@KAUST. The finite-element mesh is composed by 1,906,624 hexahedron elements.*

To validate much further our methodology (stages 1° and 2°) and in order to obtain convergences of the MC estimates (stage 3°), we have performed 500+ simulations on Shaheen II supercomputer (representing approximately 5 million core hours, i.e., 10,000 core hours per simulation). Each simulation have been launched on 2,048 cores and lasted 4h30min. A total of about 70 TB have been generated (500+ simulations times 137 GB of seismograms per simulation). Figure 3 shows the variance-based global analysis conducted on 1D wave propagation (Sochala and De Martin, *submitted*) that we intend to perform on each receivers of the 3D SEM simulations to answer the key question at stake: what is the spatiotemporal variability and sensitivity of seismograms at the surface with respect to seismological parameters?

Conclusions

The main idea of the probabilistic approach introduced in this abstract, so called big data analysis, is really to take advantage of coming computing facilities by considering their main features (high core counts, multi memory hierarchies, large file system) and limitation (usual lack of application scalability) and to pair this with mathematical developments as uncertainties quantification judiciously adapted to our geophysical data. By targeting an integrated workflow, we shown that the final time to solution can be greatly improved together with an optimal usage of the large super computer and the associated new capabilities which are not so obvious to handle for any scientific community. On the geophysical side of the project, we shall obtain interesting results for probabilistic seismic hazard assessment.

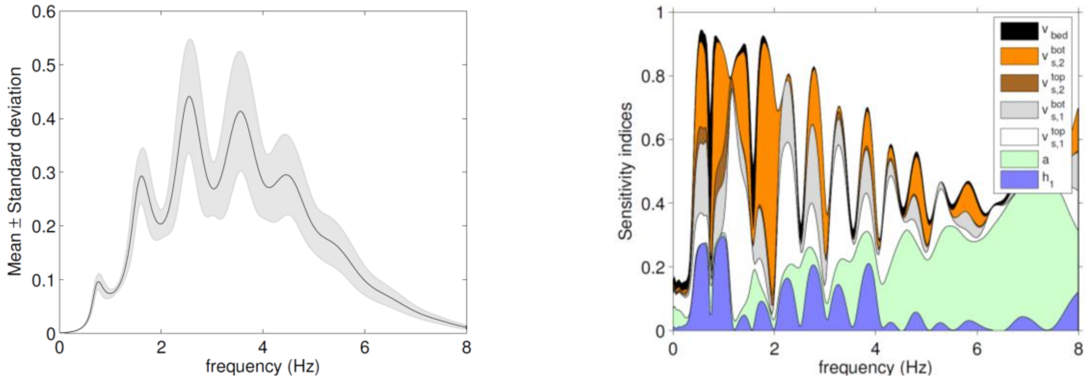

Figure 3 *Left: mean motion (solid line) and standard deviation (grey shading) in the frequency domain. Right: first order sensitivity indices (each colour represents an uncertain input parameter). (Courtesy of Sochala and De Martin, submitted).*

References

Chaljub, E., Maufroy, E., Moczo, P., Kristek, J., Hollender, F., Bard, P-Y., Priolo, E., Klin, P., De Martin, F., Zhang, Z., Zhang, W., and Chen, X. (2015) 3-D numerical simulations of earthquake ground motion in sedimentary basins: testing accuracy through stringent models. Geophysical Journal International, 201(1), 90-111.

De Martin, F. (2011) Verification of a spectral-element method code for the southern California earthquake center LOH. 3 viscoelastic Case. Bulletin of the Seismological Society of America, 101(6), 2855-2865.

De Martin, F., S. Matsushima, H. Kawase (2013) Impact of geometric effects on near-surface Green's functions. Bulletin of the Seismological Society of America, vol. 103(6): 3289-3304, doi: 10.1785/0120130039

Komatitsch, D., & Vilotte, J. P. (1998) The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. Bulletin of the seismological society of America, 88(2), 368-392.

Le Maître, O., Knio, O.M. (2010) Spectral methods for uncertainty quantification, Springer.

Maday, Y., & Patera, A. T. (1989) Spectral element methods for the incompressible Navier-Stokes equations. In IN: State-of-the-art surveys on computational mechanics (A90-47176 21-64). New York, American Society of Mechanical Engineers, 1989, p. 71-143. Research supported by DARPA. (Vol. 1, pp. 71-143).

Matsushima, S., T. Hirokawa, F. De Martin, H. Kawase, F. J. Sánchez-Sesma (2014) The Effect of Lateral Heterogeneity on Horizontal-to-Vertical Spectral Ratio of Microtremors Inferred from Observation and Synthetics, BSSA, vol. 104(1):381-393, doi:10.1785/0120120321

Maufroy, E., E. Chaljub, F. Hollender, J. Kristek, P. Moczo, P. Klin, E. Priolo, A. Iwaki, T. Iwata, V. Etienne, F. De Martin, N. Theodoulidis, M. Manakou, C. Guyonnet-Benaize, K. Pitilakis, and P.-Y. Bard (2015) Earthquake ground motion in the Mygdonian basin, Greece: the E2VP verification and validation of 3D numerical simulation up to 4 Hz. BSSA, 105(3).

Maufroy E., E. Chaljub, N. P. Theodoulidis, Z. Roumelioti, F. Hollender, P.-Y. Bard, F. De Martin, C. Guyonnet-Benaize and L. Margerin (2017) Source-related variability of site response in the Mygdonian basin (Greece) from accelerometric recordings and 3D numerical simulations, BSSA, 107(2).

Paciucci, G, De Martin, F. Thierry, P. (2016) Enhancing Scalability and Performance of Parallel File Systems. http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/enhancing-scalability-and-performance-white-paper.pdf

Pitilakis, K., Roumelioti, Z., Raptakis, D., Manakou, M., Liakakis, K., Anastasiadis, A., and Pitilakis, D. (2013) The euroseistest strong-motion database and web portal. Seismological Research Letters, 84(5):796–804.

Sochala, P., Le Maître, O. (2013) Polynomial Chaos expansion for subsurface flows with uncertain soil parameters, AWR, 62,139-154.

Sochala, P. and De Martin, F. (submitted) Surrogate combining harmonic decomposition and polynomial chaos for seismic full waveform in uncertain medium. Computational Geosciences.