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

# Validating the Numerical Simulation Approach for Ground Motion Prediction: General Framework and Latest Lessons from the E2VP Project

E. Maufroy[1], E. Chaljub[2], F. Hollender[3], P.-Y. Bard[4], J. Kristek[5], P. Moczo[6], F. De Martin[7], N. Theodoulidis[8], M. Manakou[9], C. Guyonnet-Benaize[10], K. Pitilakis[11], N. Hollard[12]

## ABSTRACT

The Euroseistest Verification and Validation Project (E2VP) is part of a series of complementary benchmarking exercises launched to better assess the ability of numerical simulation to accurately predict seismic ground motion. E2VP targeted more specifically the current, most-advanced numerical methods applied to realistic 3D, linear models of sedimentary basins through a quantitative comparison of the recorded and numerically-simulated ground motions. The target site, located within the Mygdonian basin near Thessaloniki, Greece, has been thoroughly investigated for two decades and a detailed, realistic 3D model has been derived from geological, geophysical and geotechnical investigations, while a dedicated instrumentation provided a significant number of surface and borehole recordings. Verification and validation tests up to a frequency of 4 Hz, much beyond the 0.7 Hz fundamental frequency, have been performed for a set of local, small to moderate magnitude events. For careful and accurate enough computations, the model-to-model differences are smaller than the model-to-observations differences, controlled by uncertainties primarily in the crustal propagation model and source properties, and secondarily in the shallow structure. Additional sensitivty tests illustrate the ability od carefully verified numerical simulation tools to provide an instructive insight at the structure of the so-called "aleatory" variability of ground motion, for both its within- and bewteen-event components.

## Introduction

The rapid development of the simulation codes and computational facilities allowed to consider the use of numerical-simulation tools as a valid option for predicting seismic ground motion, especially for poorly instrumented or moderate-seismicity countries lacking representative earthquake recordings. However, such an approach requires a careful evaluation of the actual performance of numerical simulation codes. This issue has been the topic of a few international studies, including blind prediction tests or comparative exercises, focused on various sites. It

[1] Univ. Grenoble Alpes, ISTerre, CNRS, IRD, IFSTTAR, Grenoble, France, emeline.maufroy@ujf-grenoble.fr
[2] Univ. Grenoble Alpes, ISTerre, CNRS, IRD, IFSTTAR, Grenoble, France, emmanuel.chaljub@ujf-grenoble.fr
[3] CEA, French Alternative Energies and Atomic Energy Commission, Cadarache, France, fabrice.hollender@cea.fr
[4] Univ. Grenoble Alpes, ISTerre, CNRS, IRD, IFSTTAR, Grenoble, France, pierre-yves.bard@ujf-grenoble.fr
[5] Faculty of Math., Physics and Informatics, Comenius Univ., Bratislava, Slovakia, Jozef.Kristek@fmph.uniba.sk
[6] Geophysical Institute, Slovak Academy of Sciences, Bratislava, Slovakia, moczo@fmph.uniba.sk
[7] Bureau de Recherches Géologiques et Minières, Orléans, France, F.DeMartin@brgm.fr
[8] ITSAK, Institute of Engineering Seismology and Earthquake Engineering, Thessaloniki, Greece, ntheo@itsak.gr
[9] Aristotle University of Thessaloniki. Thessaloniki, Greece, manakou@civil.auth.gr
[10] formerly CEA, Cadarache, France, cedric.guyonnet-benaize@hotmail.fr
[11] Aristotle University of Thessaloniki. Thessaloniki, Greece, kpitilak@civil.auth.gr
[12] Univ. Grenoble Alpes, ISTerre, CNRS, IRD, IFSTTAR, Grenoble, France, Niels.Hollard@e.ujf-grenoble.fr

started with the Turkey Flat, California (Cramer 1995), and Ashigara Valley, Japan (e.g., Bard 1992), blind tests focusing on effects of surface sediments, the results of which were presented during the first ESG conference in Odawara (1992). It was followed by more comprehensive comparison exercises on the Osaka/Kobe basin area in Japan (Kawase and Iwata, 1998), and on the Southern California area within the SCEC framework (Day et al. 2001, 2003, 2005; Bielak et al. 2010), which also included the effects of extended sources and regional propagation in the low frequency range (f < 1 Hz). Each of these cases had its own specificities (for instance, very low frequencies for the Osaka and SCEC exercises). A request issued in late 2003 by the French Nuclear Authority (ASN) to perform a 3D, NL simulation of site response for specific sites, was the initial impetus for a dedicated R&D program funded by CEA Cadarache and ILL (Laue-Langevin Institute, an international research centre on neutron science based in Grenoble, and operating the most intense neutron source on Earth). It started with an international benchmarking exercise on the Grenoble basin (Chaljub et al., 2009; Tsuno et al., 2009; Chaljub et al., 2010), and was further deepened through the Euroseistest Verification and Validation Project (E2VP). Considering the lessons of the ESG2006 Grenoble benchmark, the E2VP project was launched in 2007 with two main objectives: (a) a quantitative analysis of the accuracy of current, most-advanced numerical methods applied to realistic 3D models of sedimentary basins, in the linear, small strain domain (3DL verification); (b) a quantitative comparison of the recorded and numerically-simulated ground motions (3DL validation). The selected target site was an extensional graben located in the Mygdonian basin near Thessaloniki, Greece (Figure 1): a detailed, realistic 3D model of the medium had already been derived from a comprehensive set of geological, geophysical and geotechnical investigations, and the site instrumentation installed for about two decades provided a significant number of surface and borehole recordings.

This paper is intended to present a concise overview of the work accomplished since the launching of the E2VP project. This project has been organized in two phases, E2VP1 (2007-2010) and E2VP2 (2012 – 2014). As the main results of the first phase are reported in two recent papers (Chaljub et al., 2015; Maufroy et al., 2015), the present article puts more emphasis on the latest results, while reminding the overall process.

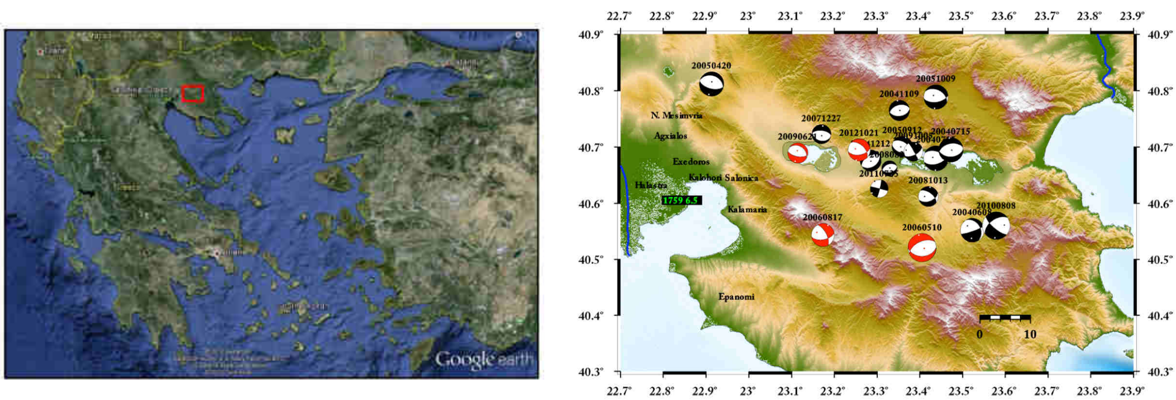


Figure 1 : Location of the Euroseistest site in between the Volvi and Langhada lakes in NorthEastern Greece, together with the location and focal mechanisms of the 19 events used for the validation phase.

Moreover, this 3DL benchmarking exercise has been complemented with two other similar exercises, focusing on the comparison and the assessment of uncertainties associated with a) 1D non-linear simulation codes and b) non-invasive and invasive methods used to derive the seismic velocity profile or seismic parameters within the soil. As these exercises are presented in larger detail elsewhere, only their overall characteristics will be summarized here.

## From E2VP1 to E2VP2 : the main steps

In short, the basic ideas of the project were, on the example of the Euroseistest site, to (1) quantify the "distance" between results of independent models and numerical schemes, and as much as possible to reduce them to the lowest possible level through a careful understanding of the differences; and (2) to compare this "cross-computation distance" to their "distance" to actual measured data for as many real events as possible. The first phase E2VP1 (2007-2010) included a comprehensive series of cross-model verifications, with side computations on canonical models aimed at investigating the accuracy of numerical schemes under very stringent conditions – as detailed in Chaljub et al. (2015) -, and a first round of comparison between observations and simulations for a small number (6) of local events, as reported in Maufroy et al. (2015). The computations were performed up to a frequency of 4 Hz: this remains limited compared to the frequency range of interest in earthquake engineering, but this is significantly higher than all previous similar exercises. It led to a number of lessons and recommendations on the use of the numerical-simulation approach, as listed in Table 1, but it also led to the identification of a few further issues that needed to be addressed in a second phase.

### *3D linear modeling*

The main focus and success of E2VP1 was thus on the use of 3D, linear simulation. The main results are summarized in Figure 2. The code-to-code differences could be drastically reduced by the consideration of dedicated canonical models and stringent goodness-of-fit criteria comparing the waveforms in the time-frequency domain (Kristekova et al., 2009), leading to significant improvements in the numerical schemes (Chaljub et al., 2015). The simulation-to-observation differences could be quantified for only a limited number of events (6) because of the moderate seismicity and the limited extension of the 3D model considered by that time. For those events, the simulated and observed waveforms remain so different that another metrics was adopted to quantify their distance, on the basis of "engineering" parameters. After a careful analysis of the original Anderson's criteria, five parameters (C1 to C5) were selected : pga, the spectral acceleration at intermediate and low frequencies (averages in the [1.5 – 3 Hz] and [0.375 – 0.75 Hz] ranges, respectively), an "energy" indicator (cumulative absolute velocity, CAV), and the Trifunac-Brady duration (RSD). The misfit was computed for each parameter in terms of relative increase or decrease compared to the measured values. Figure 2 indicates that such an "engineering" distance is around 10-25 % between different simulations, while it is in the range 40% to 80% between observations and simulations. These numbers do vary depending on the considered receiver (rock or valley), on the considered event, and on the engineering parameter, but the overall trends are robust, and emphasize both the usefulness of the prior verification part and the difficulty to obtain satisfactory, unbiased numerical predictions of ground motion.

Only very few events could be used for the validation: this is a typical situation for moderate/weak seismicity areas. It was therefore considered useful to include more events [from 6 to 19] in the second phase of the validation exercise (those shown in Figure 1), which led to increase the size of the 3D model, as illustrated in Figure 3a. In addition, the significant distance between observations and simulations was shown to be partly due to uncertainties or errors in source parameters : the misfits on the sole site response component were found lower than those on absolute motion (Maufroy et al., 2015). It was thus decided first to improve as much as possible the location of the 19 selected events, and second to investigate through numerical simulation how the uncertainties in source parameters map on the variability of site-specific ground motion from local earthquakes.

Table 1 : Summary of main learnings from E2VP Phase 1

| | |
|---|---|
| **Main lessons about verification and validation studies** | • Careful verification requires time and often to "go back to basics", while careful validation requires high quality data, i.e., including rich and high quality metadata.<br>• No ground-motion simulation code accounting for wave propagation in complex media can be considered as press-button, neither in the linear, 3D domain, nor in the non-linear 2D - or even 1D - cases. The most common case is that, without iterations and cross-checking, different codes provide significantly different results when applied to the same case study.<br>• Too fast applications of existing codes may yield VERY wrong ground-motion estimates, potentially resulting in raising mistrust in end-users.<br>• Some codes currently used in engineering applications would deserve some significant improvements, or strong warnings on stringent validity limits, while even state-of-the-art codes (predominantly in the "academic" field) deserve constant upgrading. |
| **Main recommendations for a wise use of numerical simulation codes** | • One should never be satisfied with only one computation from one single team, but should request several teams (at least two) with different numerical schemes to perform parallel computations of the same case. Results should be considered as reliable only if they agree beyond some quantitative goodness-of-fit threshold.<br>• These goodness-of-fit criteria should definitely be agreed upon by the engineering community in order to reach an objective of transparent quantitative comparison, which should replace sentences such as "one can see the very good agreement on the figure"…<br>• In the long run, it would be very valuable to assign a specific "quality label" to numerical codes and teams that did accept to run some of the now existing "canonical" cases with their own numerical code, which are freely available on web pages (http://www.sismowine.org/). Maintaining this kind of internet facility in the long run will be beneficial for the whole community.<br>• External peer reviews are always useful in assessing the quality of results derived from highly sophisticated numerical codes.<br>• Comparison with actual data (in-situ earthquake recordings), whenever possible, are always useful. Having sensitive in-situ instrumentation (continuously recording broad-band velocimeters or sensitive accelerometers) proves to be invaluable for checking the reliability of numerical-simulation results. |

### *Non-linear (NL) modeling*

The first phase also included an comparison of 2D, NL simulations on a NS cross-section of Euroseistest. This attempt for a verification of NL codes proved however to be a failure, as code-to-code differences were too large with too many, too poorly identified origins. Yet, it is obvious that NL simulation codes deserve similar verification and validation efforts, especially as they are much more often used in engineering practice than 3D, linear simulation codes. Although non-linear site response was one of the major topics of the two pioneering blind tests initiated in the late 80's for the sites of Ashigara Valley (Japan) and Turkey Flat (California), they were inconclusive because both sites lacked strong motion records. A new benchmarking of 1D NL codes was performed in the last decade, based on the same Turkey fat site that experienced a 0.3g motion during the 2004 Parkfield earthquake, and a few other sites with vertical array data (La Cienega, Caifornia; the KGWH02 KiK-net site in Japan, and Lotung in Taiwan). Its main findings, reported by Kwok et al. (2008) and Stewart and Kwok (2009), emphasized the key importance of the way these codes are used and of the required in-situ measurements. Significant differences between records and predictions have been identified as due to an incorrect velocity profile (although it was derived from redundant borehole measurements), a non-1D soil geometry (non-planar layers), and imperfections / deficiencies in the constitutive models, which were unable to represent the actual degradation curves for shear modulus and damping. The E2VP1 failure and these recent conclusions thus allowed to issue three main recommendations for future benchmarking exercises: a) NL verification should be performed on the simplest

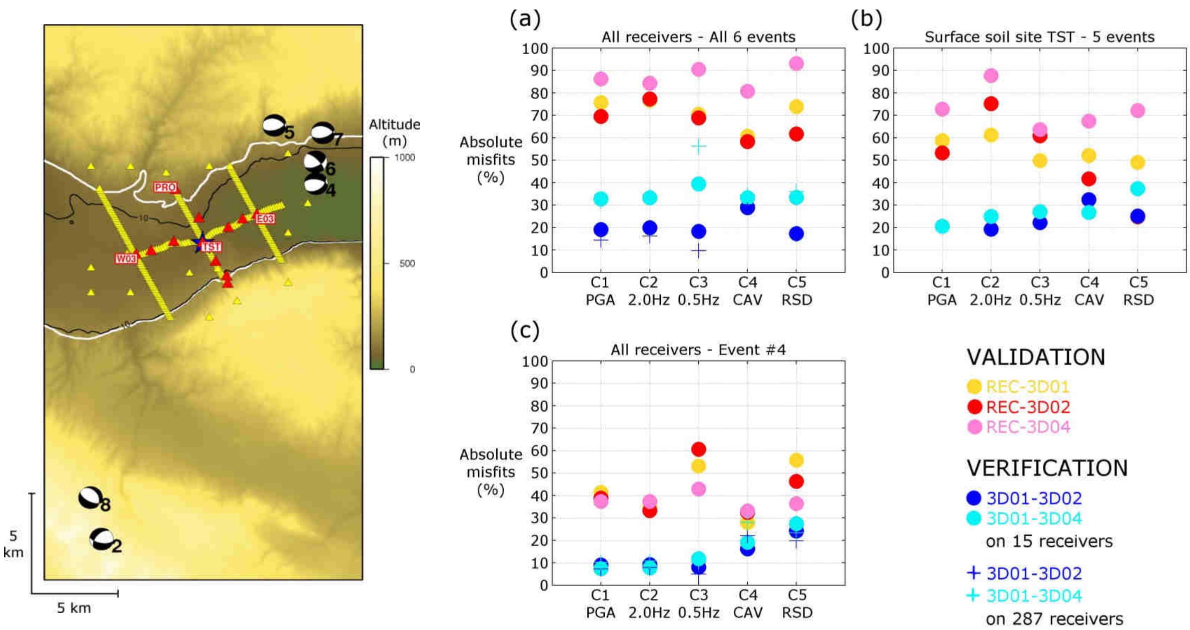


Figure 2 : Summary of horizontal absolute misfits obtained on the E2VP1 evaluation parameters C1 to C5 (see text for their definition) for the verification and validation exercises considering different configurations. Left : localization and focal mechanism of the 6 validation events (beachballs) and of the receivers used for the comparison (red and yellow triangles). Right : (a) average misfits for the 6 selected events at all receivers; (b) average misfits for the 5 events recorded at the central soil site TST; (c) average misfits for the biggest event #4 at all receivers. Synthetics-to-synthetics misfits (verification, blue tones dots) are compared to recordings-to-synthetics misfits (validation, warm tones dots). The verification misfits are computed for either the real array of 15 surface receivers (red triangles, solid circles) or the complete virtual array of 287 receivers (yellow triangles, crosses). A single value per array is obtained by calculating the weighted average of the absolute misfits over the considered receivers (with weights proportional to the value of the corresponding ground-motion parameter).

possible cases (1D soil columns and total stress, water-free, analysis); b) it should be performed on already instrumented sites having recorded large acceleration levels; c) it should be associated with careful in-situ surveys and lab tests designed in tight connection with the needs of the rheological models implemented in the various NL codes.

The second phase, E2VP2 (2012-2014). was thus designed to answer some of the identified issues related with 3D linear modeling, while the PRENOLIN project (Régnier et al., 2015a, b) was designed to start answering the issues about NL modeling according to the E2VP1 lessons, while another benchmarking exercise, named "INTERPACIFIC", was launched for a comparative assessment of the performance of various in-situ geophysical and geotehcnical survey techniques (Hollender et al., 2015).

## Example results of the second phase E2VP2

### *E2VP2 scope and contents*

The second phase of E2VP thus included the following steps

- Improvement of source parameters for an increased number of local events (from 6 to 19)
- Update and extension of the 3D model of the whole Mygdonian basin using all available information (geology, hydrological and geotechnical boreholes, geophysical surveys: seismics, electric resistivity, magneto-telluric, microtremor H/V) to constrain the bedrock geometry, the sedimentary thickness and the seismic velocity. The new model is characterized by the absence of velocity jumps within the sediments, with a "double-gradient" velocity model, characterized by a first linear gradient from 130m/s at surface to 475m/s at an intermediate surface within the basin, and a second linear gradient from this intermediate surface (475 m/s) to the bedrock interface (with a sediment base velocity of 800 m/s). This intermediate surface, corresponding to a gradient change with velocity jump, could be associated to the Mygdonian / Premygdonian (M/P) limit. The corresponding velocity, unit mass and quality factors are listed in Table 2.
- Update of the 3D simulation model (Spectral Element method) with improved meshing and velocity homogeneization strategy, including surface topography and intrinsic attenuation. The meshing was tuned for a maximum frequency of 4 Hz and the associated wavelengths.
- Ground motion simulations for various sets of events and receivers
    - The validation set ("S1") consisting of the 19 selected events, with their actual, improved source parameters (magnitude range : 2.7 – 4.6; distance range : 0 – 30 km), computed at the 15 receivers
    - A second set ("S2") corresponding to 5 real events, taking into account the uncertainty in source location : 125 hypocentral positions were considered for each of the 5 events, by shifting the actual hypocenter by ± 1 km and ± 2 km in each X, Y and Z directions
    - A large set ("S3") of 7*36*5 =1260 virtual events arranged in 7 concentric circles from 2.5 to 30 km, 36 back-azimuths (10° step) and at 5 different depths from 2.5 to 15 km. The corresponding focal mechanisms were randomly generated following a Gaussian distribution around the "average" normal faulting parameters in the Mygdonian basin area : strike = 86°±18°, dip = 52°±15°, rake = -101°±51° The objective was to investi-gate the sensitivity of the site response to the source location in a fully 3D

environment.

- Out of this "S3" set, a subset "S4" was extracted corresponding to a set of 52 really occurred events ("real catalog" see below), which could not however be all used for the validation as the corresponding number of recordings was often too small, and the focal mechanism could not be determined with enough accuracy.

The two latter sets (S2 and S3, including thus S4) were computed for the 15 receivers using the reciprocity theorem. The corresponding hypocenter locations are displayed in Figure 3b.

Table 2: $V_S$, $V_P$, ρ, $Q_S$ and $Q_P$ “anchor” values used to build E2VP2 properties model within the basin.

| | *Vs* | *Vp* | *ρ* | *Qs* | *Qp* |
|---|---|---|---|---|---|
| Surface | 130 | 1500 | 2075 | = Vs/10 | = max [Vp/20, Vs/5] |
| M/P limit | 475 | 2100 | 2130 | | |
| Bedrock top | 800 | 2700 | 2250 | | |

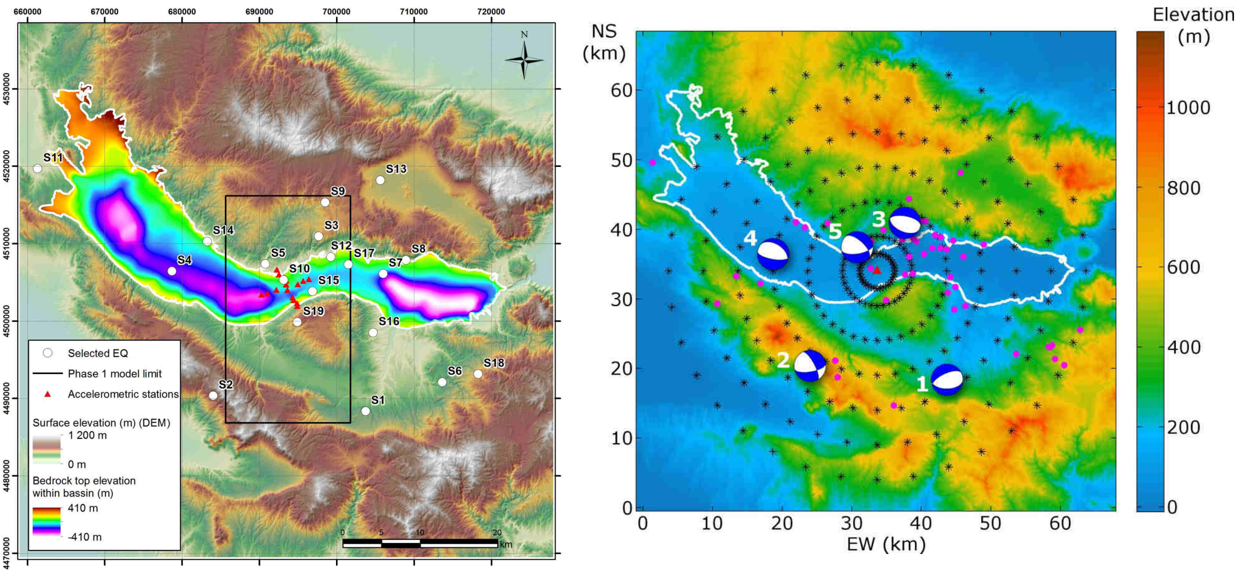


Figure 3: Left, a): Map of the whole model used for E2VP phase 2 modeling (box of 69 x 69 km), with the location of the area of the “phase 1” modeling box, accelerometric stations, modeled earthquakes, DEM et elevation of the top of the bedrock within the basin. Right, b): Location of the "virtual" seismic sources considered in the numerical study. The response of the Mygdonian basin (bold white line) is computed for five real events (beachballs) and 1260 sources (black circular crosses) at the central soil site TST indicated by the red triangle. The real catalog of local events is also shown by the magenta dots.

***New validation results***

The comparison for the 19, relocated events (set "S1") has been found in average slightly improved for rock sites, and slightly deteriorated within the Mygdonian basin, with an overall trend for an overestimation. Results are summarized in Table 3 and Figure 4 for the case of the central site TST, in terms of absolute motion and relative TST0/TST5 amplification. It is found once again that the sole site response ("hybrids") is better estimated: the "hybrids" correspond to synthetics obtained with convolving the actual downhole recording ("TST5") with the Fourier transfer function TST0/TST5 computed for the same event. The significant overestimation in terms of signal amplitude (C1-C3, C4) thus comes mainly for the overestimation of the rock motion, also associated with an underestimation of signal duration (C5): both may come from the absence of scattering in the considered crustal model. When considering all the real receivers, the overall E2VP2 misfit values range between +50 and +150%, to be compared with the +40-80 % of E2VP1 (on only 6 events), while the "site-response only" misfits now range around +20%, while they were around -40% for E2VP1. In short, the new validation phase highlighted once again a strong sensitivity of the validation scores to the source parameters and the associated uncertainties, and the importance of a higher number of rock sites (more than 1 or 2) for a proper "calibration" of the reference motion. The modifications in the basin model have slightly improved the site response estimate.

Table 3: Values of average horizontal misfits on the five engineering parameters Ci between the actual recordings at central soil site TST0 and their numerical predictions. Values in % evaluate the predictions by full synthetics vs. hybrid time histories (i.e. site response only).

| | FULL SYN. | HYBRIDS | FULL SYN. | HYBRIDS | FULL SYN. | HYBRIDS | FULL SYN. | HYBRIDS | FULL SYN. | HYBRIDS |
|---|---|---|---|---|---|---|---|---|---|---|
| | **C1** | **C1** | **C2** | **C2** | **C3** | **C3** | **C4** | **C4** | **C5** | **C5** |
| AVERAGE | 143 | 19 | 129 | 16 | 45 | -22 | 107 | 50 | -69 | 27 |

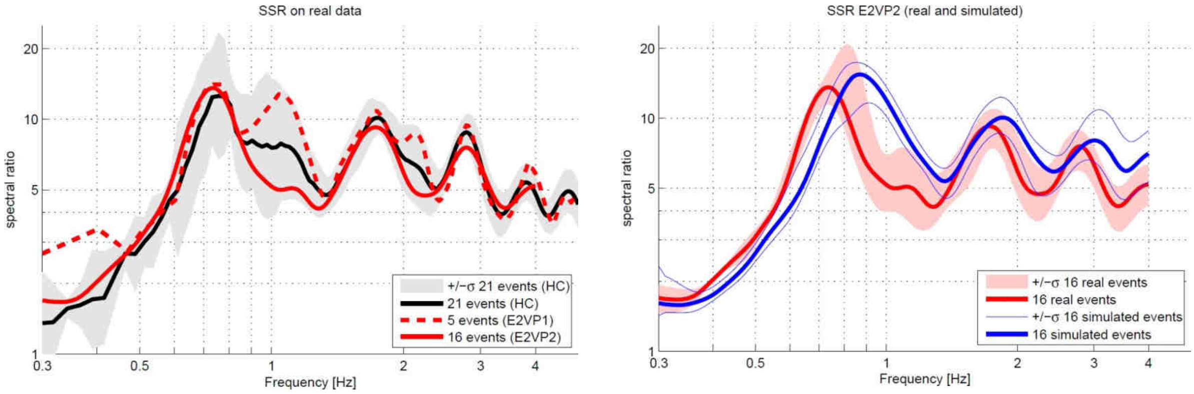


Figure 4: Left : Median of SSR (Standard Spectral Ratios) at TST0 with TST5 as reference station, computed for the actual recordings of 21 events (black line, associated variability shown in gray), for the 5 events selected in E2VP1 (dashed red line) and for the 16 events selected in E2VP2 (solid red line), that were recorded both at TST0 and TST5. Right : The same for the E2VP2 simulations (bold blue line, associated variability shown by thin blue lines) with comparison for the recordings of the same events (bold red line, associated variability in light red)

### *Findings from the sensitivity study*

Given the consistent indications of E2VP phases 1 and 2, the first objective was to evaluate the sensitivity of both the absolute motion and the site response to hypocentral uncertainties: the ± 2 km variability in each x, y and z direction (set "S2") is considered a reasonable and probably minimum estimate of the actual location uncertainty. The results are displayed on Figure 5 for the TST0/TST5 spectral ratios and the 5 considered events shown in Figure 3b. Significant differences appear between the 5 events: the largest variability is found for events S1 and S5, while it is very limited for events S2, S3 and S4. Event S5 turns out to be the shortest epicentral distance to the receiver (epicentral distance 4.5 km, a depth of 11 km), while event S1 is the most distant source, but the most shallow (depth = 5 km, epicentral distance = 18.5 km). Events S2, S3 and S4 have depths around 10km and epicentral distance in the range 8 – 16 km. Validation exercises are thus very difficult for very close events or very shallow, local events, because of the larger sensitivity of the site response in such cases (highly variable incidence and azimuth angles).

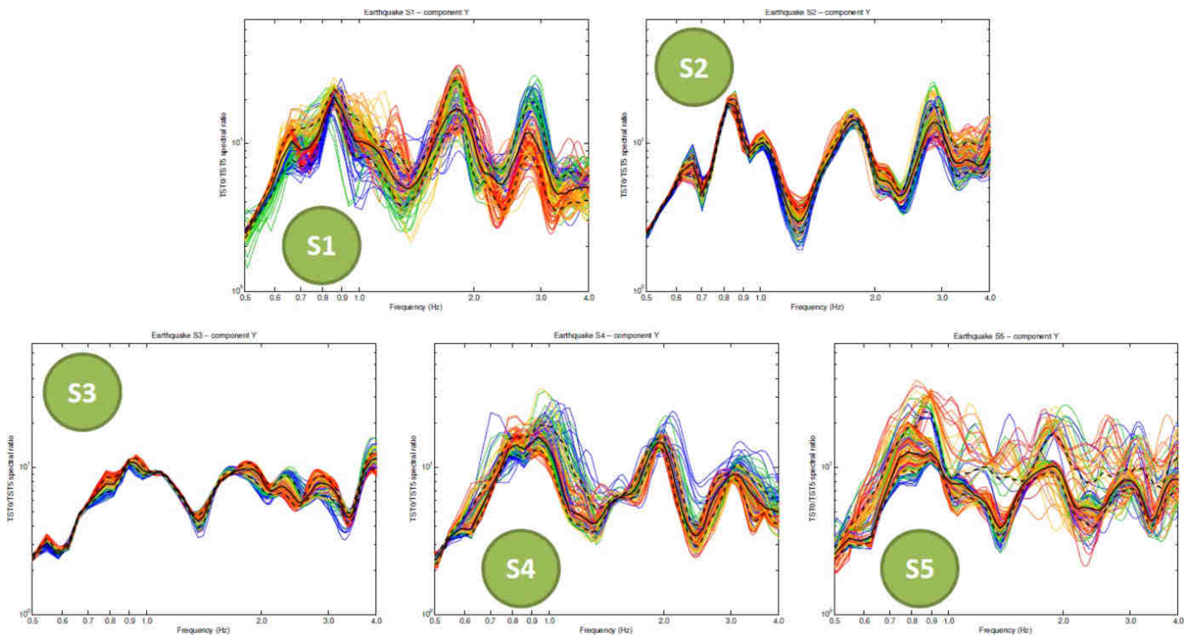


Figure 5: Standard Spectral Ratio TST0 / TST5, computed for the 5 selected real events shown by beachballs in Figure 3b. The variability of the spectral ratio due to hypocenter uncertainty is indicated by the colored lines, from blue for more shallow hypocenters to red for deeper hypocenters. The median ratio is given in each panel by the solid black line, surrounded by the upper (84%) and lower (16%) percentiles as dashed lines.

In parallel to this detailed analysis for a small number of events, the results from the "S3" set (Maufroy et al., 2015, in prep.) provide further indications on the sensitivity of ground motion and site response to the characteristics of the incoming wavefield. In the present case, as partly illustrated in Figure 6, the surface / borehole transfer function and/or response spectral ratio is found to exhibit a combined dependence on source back-azimuth, epicentral distance and depth:

- The amplification is generally found to be slightly larger for shallow, distant sources (Figure 6 upper left); it also exhibits a noticeable dependence on source back-azimuth, being slightly larger at low frequency (f < 1 Hz) for southern events, and at higher frequencies (f > 1 Hz) for northern sources.
- More significantly, the largest variability is found for shallow, far sources while the smallest one consistently corresponds to deep, close sources (Figure 6 bottom). This variability is related to the sensitivity to the back-azimuth of the source
- The variability level is frequency dependent, with a significantly larger value for the intermediate frequency range (1 – 2 Hz), in between the fundamental and first higher 1D resonances, corresponding to the band mostly affected by edge-generated surface waves: their energy strongly depends on the incidence (source-depth) and back-azimuth of the incoming wavefield, in relation with the complex 3D geometry of the sediment-basement interface.
- The variability in the amplitude of the site response is also associated with a significant variability in the duration changes, as measured either through the group delay increase (Sawada et al., 1998; Beauval et al., 2003) or through Trifunac-Brady estimates in various frequency bands (Trifunac and Brady, 1975). For duration however, the key parameter controlling this variability is the source distance, with significantly larger durations for distant sources, corresponding to more oblique incident wavefield associated with a more powerful generation of surface waves on basin edges.
- The variability in site response is also found to be significantly larger when the reference is an outcropping rock at 2-3 km distance from the considered site, while it is minimum when the reference is at depth in the bedrock beneath the considered site. Vertical arrays are thus to be recommended, even though the "reference" motion in deep bedrock may significantly differ from an outcropping rock motion.
- Finally, this comprehensive set of synthetics raised the attention on the need to consider a large set of recordings at a given site (at least several tens, with variable azimuths, depths and distances), in order to have robust estimates of the single-site variability. This should definitely be kept in mind in the derivation of GMPEs.

***Contributions to the understanding of aleatory variability***

The results of the comprehensive set of 1260 virtual sources allowed to further quantify the variability of the site amplification and to compare it with the site specific residual and the associated within-event, single site variability as derived with a GMPE approach using the available catalog (a total of 52 events).

As displayed in Figure 7, the analysis conducted for the central TST station confirms the slight, mean overestimation of the predicted amplification, while the predicted variability (i.e., within event, single site) is significantly smaller than the empirically estimated one, especially around 1 Hz, where the observed response is quite complex and highly variable (see also Figure 4a). This result, which has been recently completed for the full set of 15 sites, does show that carefully verified simulation codes can be used not only to perform deterministic predictions of the ground motion for given scenarios, but also to investigate the structure of the aleatory variability (between-event / within event, single site components).

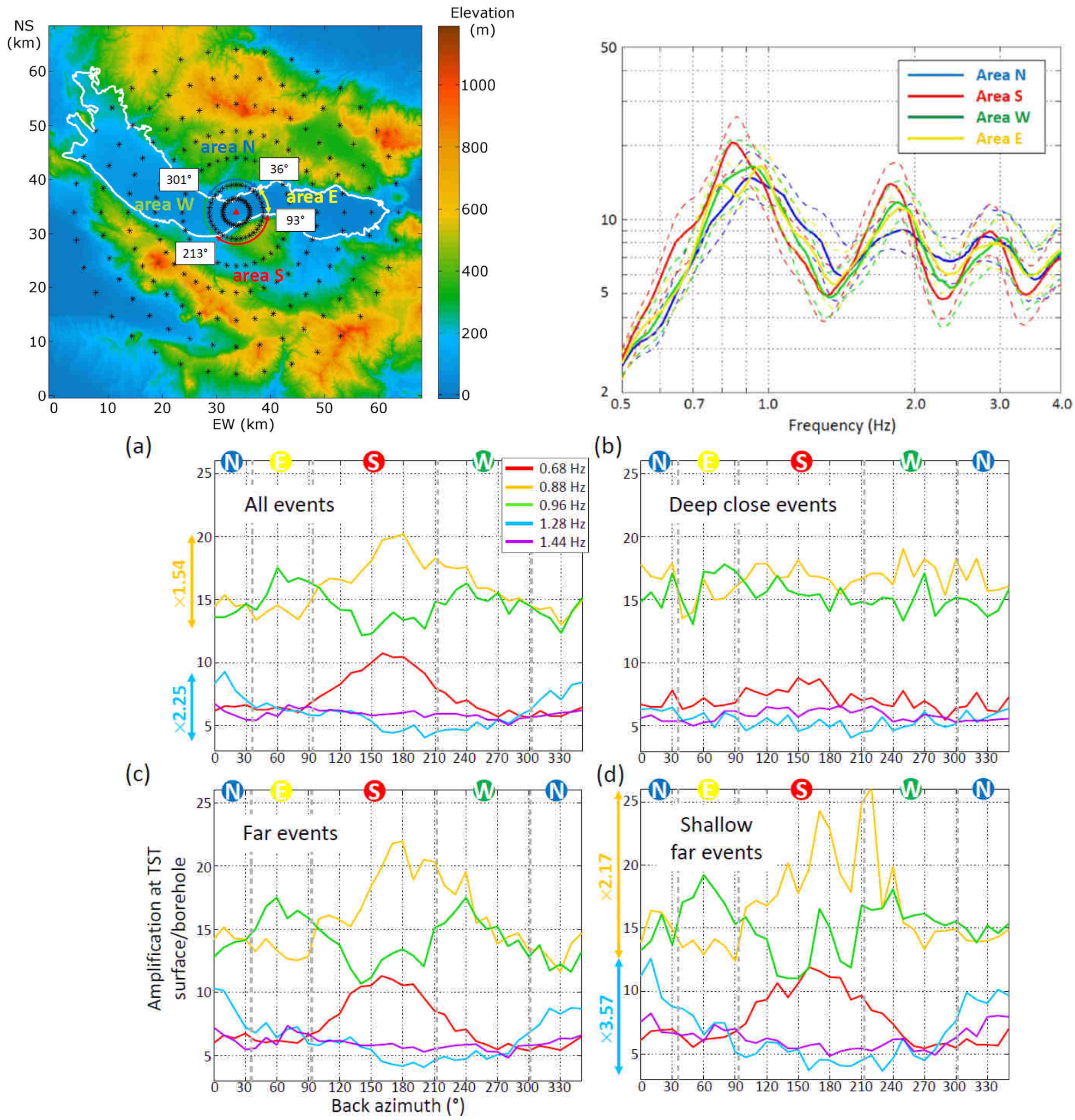

Figure 6: Example impact of the source distance and backazimuth on the average amplification at the central TST site. Top left = Map of the numerical experiment considered in this study, related to the Mygdonian basin, Greece, similar to Figure 3 right. The basin contours are indicated by the bold white line and the surface elevation is given by the color scale. The location of the central receiver TST is shown by the red triangle, and the sources epicenters are shown by the circular setting of black crosses. Ground motions at TST are analyzed by considering 4 back-azimuth areas (N, E, S, and W) as described by the colored circular arrows; the areas are separated by the 4 back-azimuth values at TST (degrees labels) that correspond to the basin edges. Top right = average surface / borehole transfer functions for the 4 different back-azimuth ranges (solid lines = average, dotted lines = average ± one standard deviation. Bottom frames a to d : details on the sensitivity to backazimuth fro five different frequencies (color code), and different event sets, as indicated in the different frames.

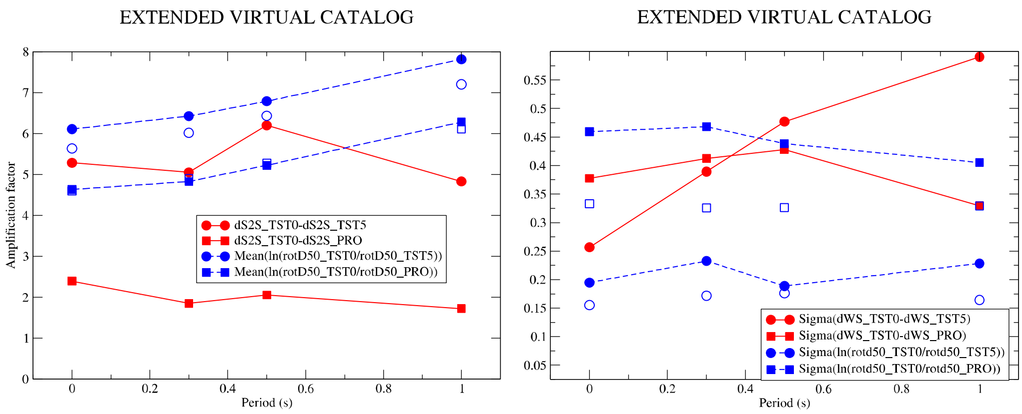


Figure 7 : Comparison of the mean (left) and standard deviation (right) of the observed (red) and simulated (blue) amplification factors at the surface station TST0. The reference station is either the downhole site TST5 (circles) or the outcropping rock site PRO (squares). The filled blue symbols correspond to the comprehensive set of 1260 virtual sources, the red symbols to the real catalog with 52 events, and the open blue symbols to the same simulated 52 events.

This set of synthetic ground motion was therefore also used to derive specific "synthetic" GMPEs in order to investigate the dependence of the aleatory variability σ (and its within-event, φ, and between-event, τ, components) on the explanatory variables and the associated uncertainties. Basically, for each considered source, the numerically derived Green's functions were convolved with adhoc source functions in order to simulate earthquakes with magnitude in the range 2-5. For the subsets corresponding to real events (S1, S2 and S4, the latter being the largest as corresponding to 52 real events), the magnitude was tuned to the real one; for all the other "virtual" events of the S3, the magnitude was assigned arbitrarily, in the limited range 2-5 however to be consistent with the point source assumption. These various sets of scaled synthetics were then used to derive GMPEs using the artificial neural network (ANN) approach described in Derras et al. (2011) and Derras et al. (2014). The "standard" explanatory variables were the moment magnitude Mw, the epicentral dstance $D_{epi}$, the hypocentral depth Z, and the $V_{S30}$ site proxy. Alternative site proxies were also considered (fundamental frequency $f_0$, local sediment thickness h, average sediment velocity $V_{Sh}$), and additional source parameters as well (mainly the source back-azimuth BAZ). The objectives were multifold:

- Compare the within and between- event variability levels between synthetics and real data (set S4)
- Investigate the effect of the size of the data set (small, S4, versus large, S3) on the values of within- and between-event variabilities
- Investigate the effect of controlled uncertainty on magnitude or source localization on the between-event variability
- Investigate the impact of various site proxies ($V_{S30}$, fundamental frequency $f_0$, local sediment thickness h, average sediment velocity $V_{Sh}$,) and of the corresponding uncertainties on the within-event variability
- In such 3D basins, investigate the possible use of other source / site information, such as back-azimuth, in view of somewhat reducing the aleatory variability

The results are still preliminary and will be presented in detail in a later paper. However, the following conclusions can be drawn from the results obtained so far, an example of which is displayed in Figure 8:

i) As partially illustrated (i.e., for one single site on Figure 7), the values of within and between-event variabilities are found somewhat lower on synthetics compared to their values on real data, for comparable data sets (S4 subset).
ii) In addition, the between-event variability is found to be much larger (i.e., by about 50%) for the full S3 set than for the reduced S4 set; the size of the data set also slightly impacts the within-event variability, but to a much smaller extent (around 10-20%). Such findings invite to be cautious when working on too small, regional data sets : both the between- and within-event variabilities may be underestimated, which may in turn also impact the estimates of single-site sigma.
iii) One of the reason of the lower variabilities on synthetics may come from the absence of uncertainty on the source parameters (magnitude and location). Therefore, different levels of uncertainties were artificially introduced in the magnitude and location values, without changing the corresponding synthetics computed with real, unperturbed magnitude and location. Magnitude values were randomly modified using a uniform distribution within [m-Δm, m+Δm], with Δm taken equal successively to 0.1, 0.2, 0.3, 0.4 and 0.5. For each Δm value, ten random set of magnitude values were generated, and GMPEs derived on the corresponding sets of unchanged synthetics, with unchanged locations and distances, and modified magnitude values. The results are illustrated, on the example of the peak ground velocity PGV, in Figure 8 left: as expected, the within-event estimate is left basically unchanged, while the between-event variability exhibits a significant, quasi-linear increase with Δm: it is doubled for Δm = 0.25, and tripled for Δm = 0.5.
iv) Similar observations are obtained concerning the uncertainty on source location (Figure 8 right). The actual locations were randomly modified using a uniform distribution centered on the actual one, with maximum deviations Δl in the x, y and z directions varying from 1 to 10 km. Again, for each Δl value, ten random sets of modified locations were generated to avoid any set-specific bias. Once again, the impact of within-event variability remains basically unchanged, while the between-event variability increases noticeably with source location uncertainty. Considering the average magnitude uncertainty (especially for moderate magnitude events) is at least 0.2, and the average location uncertainty is probably ranging from 2 to 5 km, our results indicate that a non negligible amount (i.e., increase by about 50-100%) of between-event variability comes from source parameter uncertainties. Investing in dense seismological networks for more precise localization could thus allow to reduce sigma, especially for GMPEs including large number of recordings from small magnitude events.
v) The other types of source parameters that were considered for these synthetic GMPEs are the source depth Z and back-azimuth BAZ. It was found that, for this particular data set, the $\tau$ value is signicantly reduced when considering Z (by about 70%), and further reduced (another 60-70%) when considering BAZ. Such results cannot be directly extrapolated to other data sets; they however indicate that the twin parameter set (epicentral distance, depth) performs better in ground motion prediction than the sole hypocentral distance, and that, when considering a single site with a pronounced 3D underground structure, the source back-azimuth should be considered to further reduce the prediction uncertainty.

vi) Finally, it was found that two site proxies, $V_{S30}$ and $f_0$, perform almost equally well to account for site conditions, while the local sediment thickness h and average sediment velocity $V_{Sh}$, which are sometimes proposed for alternative site classifications, perform much more poorly. A similar analysis was performed on the impact of uncertainties in the estimates of these site proxies on the within-event variability φ: it was found – again on this particular data set – that the impact is very small, much smaller than the uncertainties on source parameters. As the number of sites is limited, and the associated geological conditions as well, further investigations are needed to generalize such results.

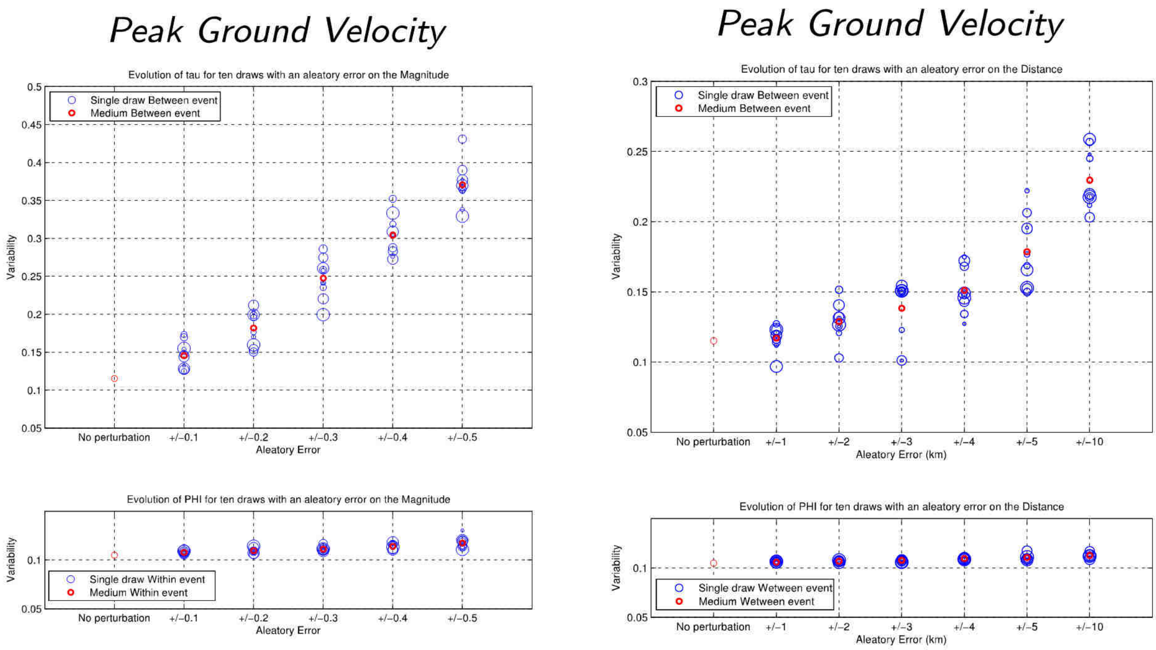


Figure 8 : Influence of the uncertainty on source parameters (magnitude, left; and location, right) on the aleatory variability components (between-event on top and within event on bottom, after Hollard et al., 2015, in preparation). This influence is illustrated here for the PGV. The amount of uncertainties considered for the magnitude values and the soruce location are indicated on the abscissa. Each open blue symbol corresponds to one random generation of perturbated magnitude or location sets, while red circles correspond to the median values for the 10 random sets for one level of magnitude or location uncertainty.

## Other ongoing benchmarking exercises : a short overview

The main objective behind these 3DL benchmarking and modeling studies is the evaluation of the ability of numerical simulation approaches to be used in hazard assessment studies for engineering design purposes. It thus had to be complemented with an evaluation of the ability of NL simulation codes to predict the actual response of soft soils at high strains, and an evaluation of the ability of geophysical and geotechnical survey methods to feed these models with an acceptable accuracy. These are mandatory steps in view of assessing the total uncertainty of the numerical simulation approach.

Concerning the assessment of accuracy of the NL simulation codes, it started with a benchmarking exercise limited to 1D (vertically stratified) soil columns and total stress analysis. The main results obtained so far are summarized in Régnier et al., 2015a and 2015b and will not be reproduced here. The next (uneasy) steps are to extend them to effective stress analysis, and to 2D pr 3D underground structures : this requires again very high-quality data and metadata from carefully selected sites.

Concerning the geophysical survey methods, a comparison of invasive and non-invasive methods was performed within the framework of the "INTERPACIFIC" project on the example of 3 real sites: one on rock (Cadarache, France), one on a stiff, thick soil (Grenoble, France) and one soft soil (Mirandola, Italy). Three different invasive techniques were considered (Down-hole, Cross-hole, and PS Suspension logger), and various teams (up to 6 for one site) performed independent measurements and processing in the same boreholes at each site. Both active (MASW) and passive (AVA, Ambient Vibration Array measurements) non-invasive methods were considered: for each site, the same set of measurements was provided to 14 different teams from Australia, Germany, France, Italy, Japan, Switzerland and USA) who processed them with their own codes and after performing their own data selection amongst the numerous AVA recordings. The main results, presented in more detail in Garofalo et al., 2015a and 2015b, can be summarized as follows (see also Hollender et al., 2015):

- The results were compared in terms of velocity profile $V_S(z)$, travel-time average velocities $V_{SZ}$ with in particular a closer look at the well-known proxy $V_{S30}$, and also in terms of dispersion curves for the non-invasive methods.
- Whatever the considered output, each type of approach (non-invasive and invasive) was found to be associated with some uncertainty. This uncertainty was found larger than expected for invasive methods, and smaller than commonly expected for non-invasive techniques.
- Invasive methods do provide a higher resolution than non-invasive ones in terms of velocity profile $V_S(z)$, but both approaches provide comparable results for "integral" measures such as $V_{SZ}$.
- In particular, the reliability of $V_{S30}$ estimates is found to be comparable for both approaches (Table 4).
- The derivation of dispersion curves from non-invasive measurements (for Rayleigh and possibly also Love waves) was found to be very robust amongst teams and methods: such dispersion curves can thus be considered a reliable and relevant site characterization. It is recommended to couple active and passive non-invasive measurements to extend the frequency range of dispersion curves.
- $V_{S30}$ can be derived directly from the dispersion curves displayed in velocity-wavelength plane. The tricky inversion step may thus be skipped when only VS30 (or travel time average $V_{SZ}$ over any other depth Z) is needed.
- For thick sites, the constraint on bedrock velocity is very poor : invasive methods do not reach the bedrock, and non-invasive techniques usually cannot resolve the underlying bedrock velocity because of too weak energy at low frequency (at least for the "continental" sites that were considered in this exercise).
- For important facilities, it is recommended to use both approaches to couple the benefits of each one : high resolution of invasive measurements, large lateral (and depth) extension of

non-invasive (passive) surveys, and use of borehole information in DC inversion.

- A guidelines document for the use of non-invasive techniques will be issued soon, and plans are being drawn to extend them to all possible methods throughout a major EU-USA collaborative work.

Table 4: Statistics on VS30 estimates from each approach for each of the 3 sites (average, standard-deviation and coefficient of variation). *After Hollender et al., 2015*

| Site | Approach type | $V_{S30}$ average [m/s] | $V_{S30}$ std [m/s] | $V_{S30}$ CoV [-] |
|---|---|---|---|---|
| Mirandola | Invasive | 209 | 12.1 | 0.058 |
| | Non-Invasive | 218 | 16.3 | 0.075 |
| Grenoble | Invasive | 352 | 18.8 | 0.053 |
| | Non-Invasive | 363 | 14.6 | 0.040 |
| Cadarache | Invasive | 1656 | 301 | 0.182 |
| | Non-Invasive | 1591 | 168 | 0.106 |

## Conclusions

The main findings of E2VP1 were confirmed in E2VP2, for both verification and validation aspects:

- The use of numerical simulation codes, even after extremely careful testing and even with the most sophisticated and up-to-date numerical schemes, can still be subject to errors (especially related to the "human factor"): careful use and cross-checking still proves to be mandatory.
- There is no single numerical-modelling method that can be considered the best in terms of accuracy and computational efficiency for all structure-wavefield configurations.
- the very detailed investigations on canonical models allowed identifying the origin of inaccuracies (type of seismic waves vs velocity model smoothness). We thus go on with recommending that any numerical method and code that is intended to be applied for numerical prediction of earthquake ground motion in engineering projects, should be verified through stringent models that would make it possible to test the most important aspects of accuracy. The canonical cases developed within E2VP, which are freely available to the seismological community (http://www.sismowine.org), can serve this purpose.

Most of the new work achieved during E2VP2 was related to validation. The feasibility of such a validation up to the frequency limit considered here (4 Hz) is still a real challenge:

- The site response proves to be very sensitive to the exact position of the source – especially its depth and back azimuth – for very close events and for local, shallow events: as it is unrealistic to expect a precision on localization smaller than 2 km (especially for the depth), it is not recommended to select such events for validation.
- The distance between observations and numerical predictions remain significantly larger than the distance between carefully selected, up-to-date, and carefully implemented numerical

simulation codes. For the prediction of ground motion for expected events with a priori defined source characteristics, the numerical-simulation approach is fully legitimate in the toolbox for site-specific ground-motion estimation.

- In addition, the predictions-to-observations differences are significantly lower when considering only the site amplification, especially when the reference is at depth within a vertical array. This emphasizes the added value of "hybrid" approaches made possible by the availability of down-hole recordings and the invaluable usefulness of in-situ recordings: it seems today very difficult to predict site effects in a complex geometry context with only geological, geophysical and geotechnical information. .
- A comprehensive sensitivity study also showed also that, beyond the deterministic prediction of ground motion for a given earthquake scenario, numerical simulation proves also to be a very useful tool for investigating the structure of the aleatory variability, and identifying some possible ways to reduce it.

## Acknowledgements

We thank A. Kiratzi, Z. Roumelioti from AUTH for their efficient help in relocating the events, and O. Ktenidou for her contribution about aleatory variability. The funding of this project is provided by CEA within the framework of the CASHIMA and SIGMA projects.

## References

Bard, P.-Y. (1992). Discussion session: lessons, issues, needs and prospects, special theme session 5: Turkey Flat and Ashigara Valley experiments, *10th World Conf. on Earthquake Engineering Conf. Proc*., 6985-6988.

Beauval, C., P.-Y. Bard, P. Moczo and J. Kristek, 2003. Quantification of frequency-dependent lengthening of seismic ground motion duration due to local geology: applications to the Volvi area (Greece), *Bull. seism. Soc. Am.,* **93-1**, pp.371-385

Bielak, J., R.W. Graves, K.B. Olsen, R. Taborda, L. Ramírez-Guzmán, S.M. Day, G.P. Ely, D. Roten, T.H. Jordan, P.J. Maechling, J. Urbanic, Y. Cui, and G. Juve (2010). The ShakeOut earthquake scenario: verification of three simulation sets, *Geophys. J. Int*. **180** 375-404. doi: 10.1111/j.1365-246X.2009.04417.x

Chaljub, E., E. Maufroy, P. Moczo, J. Kristek, F. Hollender, P.-Y. Bard, E. Priolo, P. Klin, F. de Martin. Z. Zhang, W. Zhang, X. Chen, 2015. 3-D numerical simulations of earthquake ground motion in sedimentary basins: testing accuracy through stringent models, *Geophysical Journal International* 2015 **201** (1): 90-111 doi: 10.1093/gji/ggu472

Chaljub, E., S. Tsuno and P.-Y. Bard, 2009. Numerical Benchmark of 3D ground motion simulation in the valley of Grenoble, French Alps. *ESG2006* Proceedings, Edited by P.-Y. Bard, E. Chaljub, C. Cornou, F. Cotton and P. Guéguen, LCPC Editions, ISSN 1628-4704, Vol.2, p. 1365-1375.

Chaljub, E.; Moczo, P.; Tsuno, S.; Bard, P.-Y.; Kristek, J.; Kaser, M.; Stupazzini, M. & Kristekova, M. (2010), 'Quantitative comparison of four numerical predictions of 3D ground motion in the Grenoble valley, France', *Bull. seism. Soc. Am*. **100**(4), 1427-1455.

Cramer, C. H. (1995), 'Weak-motion observations and modeling for the Turkey Flat, U.S., Site-Effects Test Area near Parkfield, California', *Bull. Seismol. Soc. Am*. **85**(2), 440-451.

Day, S.M., J. Bielak, D. Dreger, R. Graves, S. Larsen, K.B. Olsen, and A. Pitarka (2001). *Tests of 3D elastodynamic codes: final report for lifelines project 1a01*, tech. rep., Pacific Earthquake Engineering Research Center and Southern California Earthquake Center.

Day, S.M., J. Bielak, D. Dreger, R. Graves, S. Larsen, K.B. Olsen, and A. Pitarka (2003). *Tests of 3D elastodynamic codes: final report for lifelines project 1a02*, tech. rep., Pacific Earthquake Engineering Research Center.

Day, S.M., J. Bielak, D. Dreger, R. Graves, S. Larsen, K.B. Olsen, and A. Pitarka (2005). *3D ground motion simulations in basins: final report for lifelines project 1a03*, tech. rep., Pacific Earthquake Engineering Research Center.

Derras, B., P.Y. Bard, F. Cotton and A. Bekkouche, 2012. Adapting the neural network approach to PGA prediction: an example based on the KiK-net data, *Bull. seism. Soc. Am.*, Vol. **102**, No. 4, pp. 1446-1461, August 2012, doi: 10.1785/0120110088.

Derras, B., P.-Y. Bard and F. Cotton, 2014. Towards fully data driven ground-motion prediction models for Europe, in press and on line, *Bulletin of Earthquake Engineering*, **12**-1, p. 495-516. DOI 10.1007/s10518-013-9481-0

Garofalo, F., S. Foti, F. Hollender, P.Y. Bard, C. Cornou, B.R. Cox, M. Ohrnberger, D. Sicilia, M. Asten, G. Di Giulio, T. Forbriger, B. Guillier, K. Hayashi, A. Martin, S. Matsushima, D. Mercerat, V. Poggi and H. Yamanaka, 2015. InterPACIFIC project: comparison of invasive and non-invasive methods for seismic site characterization. Part I: Intra-comparison of surface wave methods. Submitted to *SDEE*, June 2015.

Garofalo, F., S. Foti, F. Hollender, P.-Y. Bard, C. Cornou, B.R. Cox, A. Dechamp, M. Ohrnberger, D. Sicilia and C. Vergniault, 2015. Interpacific project : comparison of invasive and non-invasive methods for seuismic site characterization. Part II : intercomparison between surface-wave and borehole methods. Submitted to *SDEE*, June 2015.

Hollender, F., C. Cornou, B. Cox, F. Garofalo, S. Foti, P.-Y. Bard, M. Ohrnberger, A. Dechamp, C. Vergniault and D. Sicilia, 2015. Dans la catégorie des méthodes de caractérisation des VS30 et classes de sol, le gagnant est… : Principaux résultats du benchmark "InterPacific", Proceedings of the AFPS2015 National Colloquium, Paris-Marne-la-Vallée, November 30-December 2, 2015, 11 pages.

Kawase, H., and T. Iwata (1998). A report on submitted results of the simultaneous simulation for Kobe, in *The effects of surface geology on seismic motion, recent progress and new horizon on ESG study*, vol. 3, Balkema/CRC Press.

Kristeková, M.; Kristek, J. & Moczo, P. (2009). Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophys. J. Int.* **178**(2), 813—825.

Kwok, A.O., Stewart, J.P., Hashash, Y.M., 2008. Nonlinear ground-response analysis of Turkey Flat shallow stiff-soil site to strong ground motion. *Bull. Seismol. Soc. Am.* **98**, 331–343

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, *Bull. seism. Soc. Am*., v. **105**, p. 1398-1418, doi:10.1785/0120140228.

Maufroy, E., E. Chaljub, N. Theodoulidis, Z. Roumelioti, F. Hollender and P.-Y. Bard, 2015. Soruce-related variability of site amplification in the Mygdonian basin, Greece, from acceleromertic recordings and 3D numerical simulations. In preparation (August 2015).

Régnier, J., L.F. Bonilla, P.Y. Bard, H. Kawase, E. Bertrand, F. Hollender, M. Marot and D. Sicilia (2015a). PRENOLIN Project: a benchmark on numerical simulation of 1D non-linear site effect. 1 – verification phase based on canonical cases. *6ICEGE*, Christchurch, New-Zealand, November 2015.

Régnier, J., L.F. Bonilla, P.Y. Bard, H. Kawase, E. Bertrand, F. Hollender, M. Marot, D. Sicilia and A. Nozu (2015b). PRENOLIN Project: a benchmark on numerical simulation of 1D non-linear site effect. 2 – Results of the validation phase. *6ICEGE*, Christchurch, New-Zealand, November 2015.

Sawada, S. (1998). Phase characteristics on site amplification of layered ground with irregular interface, in *The Effects of Surface Geology on Seismic Motion*, K. Irikura, K. Kudo, H. Okada, and T. Sasatani (Editors), Balkema, Rotterdam, 1009–1013.

Stewart, J., Kwok, A., 2009. Nonlinear Seismic Ground Response Analysis: Protocols and Verification Against Array Data. *PEER Annu. Meet.* San Franc.-Present. 84.

Trifunac, M.D., and A.G. Brady (1975). On the correlation of seismic intensity scales with the peaks of recorded strong ground motion, *Bull. Seism. Soc. Am.* **65**.1 139-162.

Tsuno, S. E. Chaljub and P.-Y. Bard, 2009. Grenoble simulation benchmark: comparison of results and main learnings. *ESG2006* Proceedings, Edited by P.-Y. Bard, E. Chaljub, C. Cornou, F. Cotton and P. Guéguen, LCPC Editions, ISSN 1628-4704, Vol.2, p. 1377-1433.1436.