Broad-band ground motions from 3D physics-based numerical simulations using Artificial Neural Networks

by Roberto Paolucci, Filippo Gatti, Maria Infantino, Chiara Smerzini, Ali Güney Özcebe and Marco Stupazzini

Corresponding Author

Maria Infantino, PhD Student

Department of Civil and Environmental Engineering, Politecnico di Milano

Piazza Leonardo da Vinci, 32 – 20133 Milano.

E-mail: maria.infantino@polimi.it.

Abstract

In this paper, a novel strategy to generate broad-band earthquake ground motions from the results of 3D physics-based numerical simulations (PBS) is presented. Physics-based simulated ground motions embody a rigorous seismic wave propagation model (i.e., including source-, path- and site- effects), which is however reliable only in the long period range (typically above 0.75 – 1 s), owing to the limitations posed both by computational constraints and by insufficient knowledge of the medium at short wavelengths. To cope with these limitations, the proposed approach makes use of Artificial Neural Networks (ANN), trained on a set of strong motion records, to predict the response spectral ordinates at short periods. The essence of the procedure is, first, to use the trained ANN to estimate the short period response spectral ordinates using as input the long period ones obtained by the PBS, and, then, to enrich the PBS time-histories at short periods by scaling iteratively their Fourier spectrum, with no phase change, until their response spectrum matches the ANN target spectrum. After several validation checks of the accuracy of the ANN predictions, the case study of the M6.0 Po Plain earthquake of May 29, 2012 is illustrated as a comprehensive example of application of the proposed procedure. The capability of the proposed approach to reproduce in a realistic way the engineering features of earthquake ground motion, including the peak values and their spatial correlation structure, is successfully proved.

Introduction

Earthquake ground motion prediction tools underwent a major development in the recent years, mainly because of the increasing number of strong motion records, especially in the near-field of important earthquakes. This contributed to expand research on ground motion prediction equations (GMPEs), i.e., the empirical models providing peak values of ground motion across the entire frequency band of engineering interest, as a function of magnitude, of suitable measures of source-to-site distance and of site conditions.

Due to simplicity and to limited computational cost, GMPEs are among the most important ingredients of seismic hazard assessment. However, despite their overall effectiveness and ease-of-use, the practical application of GMPEs presents several important shortcomings: (i) they provide only peak values of motion, whereas the use of non-linear time history analyses requiring reliable input motions is becoming more and more relevant within many applications of performance-based seismic design; (ii) although their number is continuously growing, the available records to calibrate a GMPE are still too few to cover the variety of situations, in terms of combinations of magnitude, distance, fault slip distribution, directivity, and shallow geological condition, which may cause a significant variability of ground motions in terms of amplitude, duration and frequency content; (iii) the data-driven calibration of GMPEs implies that the empirical coefficients vary when calibration datasets are updated; (iv) GMPEs encompass generic site conditions, represented for instance by means of the average shear velocity in the top 30 meters, V_{S30}, therefore neglecting the site-specific features, such as surface or buried topographies, basin edges, irregular soil layering, which may critically change the features of ground motion with respect to the generic site response; (v) the point-wise prediction by GMPEs cannot reproduce the spatial correlation structure of the peak values of motion at multiple sites, strongly limiting their use for seismic hazard and risk assessment study at regional scale, such as within large urban areas. As a matter of fact, in such situations additional models describing the spatial correlation of ground motion have to be applied to standard GMPEs (see e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2012).

A variety of procedures was proposed in the past to improve the above limitations of GMPEs and the accuracy of earthquake ground motion prediction (see Douglas and Aochi, 2008, for a comprehensive review). Among such procedures, boosted by the ever increasing availability of parallel high performance computing, 3D physics-based numerical simulations (PBSs) are becoming one of the leading tools to obtain synthetic ground motion time histories, whose use for seismic hazard and engineering applications is subject to growing attention and debate (see e.g. Bradley *et al.*, 2017).

Being based on a more or less detailed spatial discretization of the continuum and on the numerical integration of the seismic wave equation, carried out according to different methods (such as finite differences, finite elements or spectral elements), PBSs require a sufficiently detailed model of the seismic source, of the propagation path, and of the Earth crustal layers. To enjoy the effectiveness of semi-analytical solutions of elastic wave propagation, the shallow Earth's structure is often modelled as a system of horizontal layers (see e.g. Spudich and Xu, 2002; Hisada and Bielak, 2003). In this paper, we will refer only to those approaches where 3D numerical models of the shallow geological layers can be considered.

Physics-based numerical modeling already proved in the recent past to be well suited for global (Graves, 1996; Wald and Graves, 1998; Pitarka *et al.*, 1998; Komatitsch and Tromp, 2002a,b) and regional scale simulations (Bao *et al.*, 1998; Olsen, 2000; Dumbser and Käser, 2006; Day *et al.*, 2008; Tsuda *et al.*, 2011; Smerzini and Villani, 2012; Taborda and Bielak, 2014; Villani *et al.*, 2014; Paolucci *et al.*, 2015; Chaljub *et al.*, 2015; Gatti *et al.*, 2017), making potentially feasible the challenging problem of a multi-scale simulation from the seismic source to the structural response within a single computational model (Mazzieri *et al.*, 2013; Isbiliroglu *et al.*, 2015).

Typically, PBSs are based either on a kinematic description of the co-seismic slip distribution model or on a spontaneous dynamic rupture process. Spatially correlated random field models of slip function parameters (e.g., Herrero and Bernard, 1994; Mai and Beroza, 2003; Crempien and Archuleta, 2015; Anderson, 2015) are often considered to provide a realistic level of complexity of the generated seismic wavefield and enhance its frequency content within physical constraints from seismological observations. However, even in the presence of an ideal seismic source model, exciting the whole frequency spectrum, the accuracy of the PBS in the high-frequency range is limited, on the one hand, by the increased computational burden as the mesh gets finer, and, on the other hand, by the lack of detailed knowledge to construct a geological model with sufficient details also at short wavelengths, especially for complex configurations. As a result, accuracy achieved by PBS is usually bounded up to 1 – 1.5 Hz, although some examples of higher frequency ranges covered by deterministic PBS, with good performance validations against records, have also been published (e.g., Smerzini and Villani, 2012, modeling the M6.3 L'Aquila near-source earthquake ground motion up to 2.5 Hz; Taborda and Bielak, 2014, modeling the M5.4 Chino Hills earthquake up to 4 Hz, Maufroy *et al.*, 2015, simulating a sequence of small earthquakes in the Volvi basin, Greece, up to 4 Hz).

Different recent research works have addressed the high-frequency limitation of PBS, such as in the framework of the Southern California Earthquake Center (SCEC) Broadband Platform, aiming to extend the frequency band of synthetics and to enable PBS to be used with confidence in engineering applications (see Goulet *et al.*, 2015). Broad-band (BB) waveforms are generally produced by a hybrid approach combining low-frequency results from deterministic PBS with high-frequency signals from stochastic approaches, typically through either point- or finite-source methods (e.g., Boore, 2003; Motazedian and Atkinson, 2005) or stochastic Green's function methods (e.g., Kamae *et al.*, 1998; Mai *et al.*, 2010). Hybrid waveforms are then obtained by gluing the low-frequency and high-frequency portions of the spectrum with amplitude and phase matching algorithms (e.g., Mai and Beroza, 2003). Table 1 lists a sample of recently published studies of BB earthquake ground motions based on coupling low-frequency 3D PBS with high-frequency stochastic contributions.

Although it has been applied to many case studies worldwide, the hybrid approach may have some basic drawbacks, which prevent its use especially for regional applications: (i) typically, the low (from PBS) and high (from stochastic) frequency parts turn out to be poorly correlated, being generated through independent methods with different assumptions regarding the source and the propagation medium; (ii) the low and high frequency seismograms are combined around a cross-over frequency f_c, where the corresponding Fourier spectra are multiplied by weighting functions and summed up. Such operation may result in a Fourier spectrum of the hybrid broadband ground motion presenting artificial holes around the cross-over frequency and, to overcome this issue, may require a site-specific calibration of f_c (see e.g. Ameri *et al.*, 2012).

In this paper we propose a novel approach to generate BB ground motions, which couples the results of PBS for a specific earthquake ground motion scenario with the predictions of an Artificial Neural Network (ANN), overcoming some of the main issues of hybrid modeling. The basic steps of the procedure can be summarized as follows: (1) the ANN is trained on a strong motion dataset, to correlate short-period ($T \leq T^*$) spectral ordinates with the long period ones ($T > T^*$), being T^* the threshold period beyond which results of the PBS are supposed to be accurate;

(2) the trained ANN is used to obtain the short period spectral ordinates of the physics-based earthquake ground motion for periods below T* (Figure 1); (3) the PBS long period time histories are enriched at high frequencies with an iterative spectral matching approach, until the response spectrum matches the short period part obtained by the ANN.

A detailed introduction of the procedure, denoted hereafter by ANN2BB, is given in the following chapters, with an application example to the PBS obtained for the M_w6.0 Po Plain earthquake of May 29, 2012 (Paolucci *et al.*, 2015), for which a comprehensive validation exercise can be made, based on more than 30 strong motion records obtained at less than 30 km epicentral distance. Such validation aims at encompassing different key aspects to evaluate the applicability of physics-based earthquake ground motion to engineering practice, not only in terms of the high-frequency content and of the proper attenuation of peak values with distance, but also in terms of the verification of the spatial correlation of peak ground motion values.

Correlation of long and short period spectral ordinates through an ANN trained on a strong motion dataset

Design and training of an ANN

Artificial Neural Networks are generally used to estimate the non-linear relationship between a highly populated vector of input variables and a vector of output unknowns, for the correlation of which fast and closed-form rules cannot easily be applied. As a matter of fact, under mild mathematical conditions, any problem involving a continuous mapping between vector spaces can be approximated to arbitrary precision (i.e. within an error tolerance) by *feed-forward* ANNs which is the most often used type (Cybenko, 1989). Our purpose is to establish through the ANN a correlation between N_{Sa}^{LP} long period response spectral ordinates selected for $T \geq T^{\star}$, being $T^{\star}$ the threshold period corresponding to the range of validity of PBS, with N_{Sa}^{SP} short period response spectral ordinates for $T < T^{\star}$. A high-quality strong ground motion dataset (denoted in the following by SIMBAD, see Smerzini *et al.*, 2014 for details) was used for training. SIMBAD consists of N_{db} ~ 500 three components records from about 130 shallow crustal earthquakes worldwide, roughly homogeneously distributed in the M_W range from 5 to 7.3 and epicentral distance R_{epi} < 35 km. Quantitative information on site characterization, preferably in terms of V_{S30}, is available for all stations.

Two separate ANNs are considered and trained independently, one referring to the geometric mean of the horizontal components and one to the vertical one. As long as the database is updated with new strong motion records, the procedure can ideally be extended by training different ANNs separately, for different homogeneous datasets (such as for different soil classes) and/or for different components of motion (such as fault normal and fault parallel). In our case, the neural network is designed as a two-layers (i.e. nodes are grouped in layers) feed-forward (i.e. the arcs joining nodes are unidirectional, and there are no cycles) neural network with N_n^h sigmoid hidden neurons (the so-called activation functions) and a linear output neuron. The number of nodes in the input layer N_n^i equals the number of input variables N_{Sa}^{LP}. The number of nodes in the output layer N_n^o equals the number of target values N_{Sa}^{SP} . With this kind of configuration, the ANN takes the name of Multi Layer Perceptron (Bishop, 1995; Bishop and Roach, 1992). The backpropagation of error was used in the training phase (McClelland *et al*., 1986). The idea is to propagate the error signal, computed in single teaching step, back to all connected neurons. Back-propagation needs a *teacher* that knows the correct output for any input (supervised learning) and uses gradient descent methods (Levenberg, 1944; Marquardt, 1963) on the error to train the weights. In this work, a built-in neural network fitting tool available in Matlab, namely the package *nftool*, was used. The *nftool* package solves the problem of data fitting using a two-layer feed-forward network trained with the Levenberg-Marquardt algorithm. A simplified sketch of the logic scheme at the basis of the ANN training process is shown in Figure 2.

Referring to Figures 1 and 2, the N_{Sa}^{LP} input parameters are $\backslash\{Log_{10}\left[SA\left(T_j\right)\right]\backslash\}_{j=1}^{N_{Sa}^{LP}}$, where SA is the acceleration response spectral ordinates at period T_j, ranging from the corner period $T^{\star}$(grey line in Figure 1) to 5 s. The outputs are N_{Sa}^{SP} ground motion parameters, specifically, $\backslash\{Log_{10}[SA(T_k)]\backslash\}_{k=1}^{N_{Sa}^{SP}}$, at periods T_k = 0 (i.e. *PGA* = Peak Ground Acceleration), up to $T^{\star}$. Note that the ANN is designed to predict multiple outputs given multiple inputs: specifically, considering T^*=0.75 s, as in this study, the number of outputs and inputs is 20 and 9, respectively, with a sampling equal to T_j = [0.75,0.8:0.1:1.0,1.25:0.25:5.0] s for the input and of T_k = [0,0.05,0.1:0.1:0.7] s. In such conditions, two common sets of weights *w* and biases *b* are iteratively adjusted to map the input to the hidden layer, as well as the hidden layer to the output layer.

As for the training of the ANN, the adopted scheme is based on the random subdivision of the entire dataset of N_{db} input-output data into three subsets (as implemented in Matlab *nftool*): (1) a training set, used to calibrate the adjustable ANN weights; (2) a validation set, made of patterns different from those of the training set and thus used to monitor the accuracy of the ANN model during the training procedure; (3) a test set, not used during ANN training and validation, but needed to evaluate the network capability of generalization in the presence of new data. This distinction helps limiting the problem of overfitting, which is a well-known shortcoming of ANN design. As a matter of fact, even though the error on the training set is driven to a very small value, the network may fail in generalizing the learned training patterns if the patterns of the training set do not sufficiently cover the variety of new situations. An *early stop* criterion was adopted to stop the training phase when the error on the validation set starts growing. In our computations, the training/validation/testing sets were set to 85%/10%/5%. More specifically, before selecting the final network, different ANNs were constructed, for a total of N_{train} = 50, each based on a different training subset randomly extracted among 95% of the records. The final ANN was selected as the one providing the best performance, i.e., the lowest mean square error on the remaining 5% of the dataset.

A number of hidden neurons N_n^h = 30 was assumed, after a parametric analysis proving that this number provides a reasonable compromise in terms of accuracy of the network (see for details Gatti, 2017).

Testing the ANN performance

The performance of the selected ANN in predicting the actual recordings has been evaluated by computing the logarithmic residuals of the response spectral ordinates predicted by the ANN (SA_{ANN}) with respect to the observed ones from SIMBAD dataset (SA_{Obs}), i.e. $\log_{10}(SA_{ANN}/SA_{Obs})$. Figure 3 illustrates the residual bars corresponding to ±1σ for the geometric mean of horizontal components, as a function of *T/T**, for different values of *T**, corresponding to different possible intervals of validity of the PBS results, namely *T**=0.50s (left panel), 0.75 s (center) and 1.0 s (right). The number of input and output parameters (N_{Sa}^{LP}, N_{Sa}^{SP}) in the three cases are (22,6), (20,9) and (17, 11), respectively. Results are shown and compared for the training, validation and test phases. It is shown that, in terms of normalized period, performance is similar for the different values of *T**, with an obvious tendency of larger uncertainties as period gets lower, being more distant than the corner period *T**. In spite of this effect, it is noted that typically the accuracy of *PGA* prediction is higher. When expressed in non-normalized terms, the lower is *T** the more accurate is the prediction. It is worth underling that, with few exceptions, the error of both the validation and test phases is bounded to ±0.3 in $\log_{10}$ scale (i.e., a factor of 2), which corresponds incidentally to the total standard deviation, σ_{log10}, of typical GMPEs (see e.g. Cauzzi *et al*. 2015 derived on a similar database). This suggests that, with respect to standard empirical approaches, the reduction of uncertainty is improved as the period gets close to T^*.

A similar exercise was made for training, validating and testing an ANN to predict short period vertical spectral ordinates, based on the same dataset. Results are shown in Figure 4 and denote, as expected, a slightly worst performance of the vertical ANN with respect to the horizontal one, owing to the generally poor correlation of short vs long period spectral ordinates of vertical ground motions. This is clear especially for the ANN trained for T^*=1 s, with error bars of the validation and testing phases exceeding a factor of 3 (i.e., 0.5 in $\log_{10}$ scale) and with a significant bias on the negative side, showing that, for both the validation and test datasets, the ANN predictions underestimate significantly the observations. However, results get significantly better when decreasing T^* and, already with T^* = 0.75 s, the error bars do not exceed a factor 0.4 in $\log_{10}$ scale and the bias is significantly reduced.

Note that the previous horizontal and vertical ANNs were trained on a dataset (about 500 three-component waveforms), containing strong motion records within relatively limited epicentral distance and magnitude ranges. For this reason, we did not find a significant improvement on the results when distance and magnitude were considered as additional input parameters of the training phase, as it could be in case of training of more general ANNs on wider record datasets. On the other hand, more specific ANNs may be trained on subsets of records, aiming for example at distinguishing between soft and stiff soil conditions and, hence, at providing improved accuracy for site-specific evaluations. A check was made with such objective, as documented in Gatti (2017), but only a slight decrease of performance was found with respect to the ANN trained on the complete dataset, as if the improved classification of records was not sufficient to balance the significant decrease of number of records for each ANN. As a final remark, although we did not make quantitative tests on the minimum number of records needed for robust estimates, our performance checks indicate that stable results are obtained only within the magnitude and distance ranges of the dataset, and extrapolation out of such ranges is not reliable.

The ANN2BB procedure to produce broad-band strong ground motions from 3D physics-based numerical simulations

Based on the tests illustrated in the previous section, different ANNs may be trained for different values of T*, related to the frequency resolution of the numerical model (in this application, T*=0.75 s is considered). Therefore, this first step allows one to compute, for all PBS with range of validity T>T*, a site-specific ANN-based broad-band response spectrum, denoted in the following by ANN2BB, as well as maps of peak values of short period ground motion. Note that, at this stage, such BB response spectrum does not correspond to a specific waveform.

To obtain BB time histories from the ANN2BB spectra, a spectral matching approach is used, similar to those adopted in the engineering practice to adapt a real accelerogram to a prescribed target spectrum (NIST, 2011), where the record is iteratively scaled either in the frequency domain (see e.g. Shahbazian and Pezeshk, 2010) or by wavelet transforms (e.g. Atik and Abrahamson, 2010), with no phase change, until its response spectrum approaches the target within a given tolerance. In our case, instead of a recorded accelerogram, we consider the time history resulting from the physics-based simulation, and, as a target, the ANN2BB spectrum. In this work we selected the scaling in the frequency domain, but other spectral matching procedures can obviously be used.

The difficulty, with respect to the standard spectral matching approach, comes from the low-frequency band-limited nature of the simulated time-history, which implies that the high-frequency content of the waveform, essentially consisting of numerical noise, is not usable for scaling. To overcome this issue, before spectral matching to the desired target ANN2BB spectrum, the high-frequency portion of the simulated waveform was enriched by a stochastic component, by gluing the low and high-frequency parts with the procedure described in Smerzini and Villani (2012). For high-frequency signals, we successfully tested both the Sabetta and Pugliese (1996) and the Boore (2003) approaches, the latter implemented in the code EXSIM (Motazedian and Atkinson, 2005), and selected the result providing the best fit to the target ANN2BB spectrum. Note that, as spectral matching is achieved by scaling only amplitudes, the high-frequency random phases generated in the hybrid step are maintained.

To summarize, the main steps of the ANN2BB procedure are the following:

1) an earthquake ground motion scenario is produced based on 3D PBS, whose accuracy in terms of response spectral ordinates is limited to $T \geq T^*$, owing to mesh discretization issues as well as to limited information on the geological models;
2) an ANN is trained based on a strong motion records dataset to predict short period spectral ordinates ($T < T^*$) based on long period ones ($T \geq T^*$);
3) for each simulated waveform, a ANN2BB response spectrum is computed, the spectral ordinates of which, for $T \geq T^*$, coincide with the simulated ones, while, for $T < T^*$, they are obtained from the ANN. Both horizontal and vertical components can be obtained, although with a lower level of accuracy for the vertical case;
4) the simulated low-frequency waveform is enriched in the high-frequency by a stochastic contribution, characterized by the magnitude and source-to-site distance of the scenario

earthquake under consideration;

5) the hybrid PBS-stochastic waveform is iteratively modified in the frequency domain, with no phase change, until its response spectrum matches the target ANN2BB spectrum.

A case study: broad-band ground motions from the numerical simulations of the May 29 2012 Po Plain earthquake

To test the proposed approach for the generation of BB ground motions and to verify the accuracy of results against observations during recent earthquakes, we considered as a case study the numerical simulation of the M_W6 May 29 2012 Po Plain earthquake, Northern Italy. This earthquake is very meaningful for validation purposes, because of the availability of a significant number of near-source strong-motion records, some of which obtained at very short inter-station distances, as well as of the good knowledge on the complex geologic setting of the Po Plain, which enabled the construction of a robust 3D numerical model including its complex buried morphology. 3D physics-based numerical modelling of ground shaking during the May 29 2012 Po Plain earthquake, has been addressed in a previous work (Paolucci *et al.*, 2015), where the validation of simulated ground motions against recordings has been thoroughly analysed and discussed, limited to the frequency range of design of the numerical mesh.

We aim herein at extending the validation to the simulated BB ground motions, encompassing several aspects of engineering relevance, from the comparison of BB simulated with records at selected near-source sites, as well as the spatial distribution of peak values of ground motion and their spatial correlation features.

Review of the case study and main results

On May 20 and 29 2012, two earthquakes with moment magnitude M_W of 6.1 and 6, respectively, occurred in the Po Plain region, Northern Italy, along a thrust fault system with a nearly East-West strike and dipping to the South (Luzi *et al.* 2013). The May 29 earthquake was extensively recorded by several accelerometric networks, making available a unique dataset of high-quality strong-motion recordings in the near-source region of a major thrust event and within a deep soft sediment basin structure like the Po Plain. More than 30 recordings are available at epicentral distances less than 30 km and have been the basis for the validation of the 3D PBSs.

Referring to Paolucci *et al.* (2015) for a detailed description of the spectral-element model, we limit herein to underline its main features. The model, with an extension of about 74 km x 51 km x 20 km, can propagate up to about 1.5 Hz and includes the following distinctive elements: (i) an *ad hoc* calibrated kinematic source model of the Mirandola fault with a major slip asperity in the up-dip direction; (ii) the 3D velocity model of the Po Plain which accounts for the pronounced irregularity of the base of Quaternary sediments, with thickness varying abruptly in a short distance range from few tens of m in the epicentral area down to several km; (iii) a linear visco-elastic soil model, with frequency proportional quality factor *Q*.

The numerical model was found to predict with satisfactory accuracy, measured through quantitative goodness-of-fit criteria, the most salient features of near-source ground motion, such as, in particular, (i) the strong up-dip directivity effects leading to large fault-normal velocity pulses, (ii) the small-scale variability at short distance from the source, resulting in the out-of-phase motion at stations separated by only 3 km distance, (iii) the prominent trains of surface waves propagating with larger amplitudes in the Northern direction and dominating ground motion already at some 10 km distance from the epicenter, (iv) the spatial distribution of ground uplift on the hanging wall of the fault, in substantial agreement with geodetic measurements, (v) the macroseismic intensity distribution.

Maps of peak values of ground motion

The validation checks quoted in the previous section, and reported in detail by Paolucci *et al.* (2015), were limited to information extracted from the numerical results up to about 1.5 Hz, i.e., the range of validity of the PBS. We consider now additional tests, based on the BB results obtained with the ANN2BB procedure outlined previously.

The spatial variability of peak values of ground motion is first addressed and compared with available observations. To this end, Figure 5 compares the maps of simulated *PGA* (geometric mean of horizontal components) obtained by (a) the ANN2BB procedure (steps 1 to 5 of the previous section), (b) the hybrid PBS-stochastic approach (steps 1 to 4) and (c) the PBS results filtered at 1.5 Hz (only step 1). The Sabetta and Pugliese (1996) approach was considered to produce the stochastic high-frequency portion of motion at step 4. On the same maps of Figure 5, the values of recorded *PGA* are also superimposed, taken from processed ITACA waveforms. In Figure 5d, recorded and simulated (ANN2BB) horizontal *PGA* values are shown as a function of the Joyner-Boore distance, R_{JB}, and compared with the GMPE of Bindi *et al*. (2014), referred to as BI14. The latter was obtained assuming M_W=6.0, reverse focal mechanism and V_{S30}=220 m/s. The following observations can be made:

- the proposed ANN2BB approach provides high-frequency ground motion predictions correlated to the low-frequency motion obtained by PBS. This is made evident by the similarity of the spatial pattern, related to source effects, of Figure 5a (ANN2BB) and Figure 5c (PBS), although PBS values are bounded because of the low frequency range of the simulations. Furthermore, from the comparison between the maps at top of Figure 5, it is apparent that the *PGAs* obtained by the present approach reflect some physical features related to the wave propagation phenomenon itself (directivity, directionality, site conditions, etc.), that are missing from the stochastic approach. Namely, (i) the larger values of peaks on the northern side of the fault are consistent with the up-dip directivity effects, (ii) the pronounced NW-SE alignment of the peak corresponds to the prevailing orientation of the submerged bedrock topography included in the 3D numerical model, thus giving evidence of a complex 3D site effect, as discussed in more detail by Paolucci *et al.* (2015);

- there is an overall good agreement between the spatial distribution of simulated *PGA* and the recorded values, although simulations tend to be lower than records. This is consistent with a similar tendency of underestimation of recorded motions from PBS also in the long period range, as previously noted by Paolucci *et al.* (2015);

- the comparison with the GMPE by BI14 puts in evidence that *PGAs* recorded within the Po Plain lie well below the median empirical prediction. This can be attributed to the reduction of *PGA* values that is usually noted at the surface of deep sedimentary basins (Lanzano *et al*., 2016). It is also noted that the ANN2BB predicted values are below the GMPE results, consistently with records, but their decay with distance is faster, probably due to an overestimation of damping within the shallow soil layers of the numerical model.

Comparison between simulated BBs and recordings

Performance of the ANN2BB approach can be evaluated by checking the BB simulated ground motions. For this purpose, we show in Figure 6, from left to right, the acceleration, velocity and displacement time histories of the NS component of the Mirandola (MRN) station, located at an epicentral distance of 4 km, in one of the areas mostly affected by the earthquake. From top to bottom, the figure shows in sequence the result of PBS, according to Paolucci *et al.* (2015), the stochastic waveform (STO) obtained using the Sabetta and Pugliese (1996) approach, the hybrid

(HYB) waveform obtained by combining PBS at low-frequency and STO at high-frequency, having selected 1.5 Hz as the cross-over frequency for gluing the low and high-frequency parts, the ANN2BB waveform obtained by scaling HYB to the target response spectrum based on the application of ANN to the PBS spectral ordinates. The last row of Figure 6 portrays the recorded (REC) waveform. Comparison is further clarified in Figure 7, in terms of response spectra (left) and Fourier spectra (right) of the waveforms in Figure 6.

It turns out that both the HYB and the ANN2BB waveforms provide a remarkable approximation of recorded ground motion, both in time and frequency domain, enjoying for the MRN station a very good performance of the PBS at long periods, as confirmed by comparison with a larger set of stations, at different distances and azimuths (Figure 8). From this comparison, it is noted that the performance of ANN2BB is less satisfactory at those sites (e.g. MOG0) where the PBS results at long periods do not fit closely the observed values.

The main advantage of ANN2BB vs HYB is that the high-frequency part is related through the ANN to the low-frequency one: therefore, as illustrated in the next section, a good agreement is also expected in terms of the spatial correlation of peak values of ground motion.

Spatial correlation of peak values of ground motion

The most important motivation driving the search for a recipe to produce BB from 3D physics-based simulations using the ANN2BB approach, is that the correlation provided through the ANN between the low- and high-frequency parts of simulated ground motions is expected to ensure a realistic spatial correlation of peaks of ground motion also in the high-frequency range, not covered by the numerical simulations. For this purpose, a standard tool to quantify the spatial variability of a random process of spatially distributed samples is the semivariogram $\gamma(h)$ (Webster and Oliver, 2007) measuring, in general terms, the average dissimilarity of data at inter-station distance h.

Taking advantage of the well-known methods to model the spatial correlation between earthquake ground motion values (see e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2011; Loth and Baker, 2013), the semivariogram *γ(h)* and the corresponding correlation coefficient *ρ(h)* (Webster and Oliver, 2007) can be evaluated through the following steps: (i) computing the semivariogram by the method of moments (Matheron, 1965) under the hypothesis of second order stationarity, (iii) selecting the theoretical model of the semivariogram, (iii) estimating the parameters of the model, referred to as sill (i.e., the variance of the random process) and range (i.e., the inter-station distance at which *γ(h)* tends to the sill, indicating that motions are uncorrelated), by fitting the computed semivariogram values with the functional form chosen at the previous point and (iv) computing the correlation coefficient as the complementary to the semivariogram normalized by the sill. Referring to literature studies for the analytical background (Jayaram and Baker, 2009; Esposito and Iervolino, 2011; 2012), we note that, in this work, the residual terms, on which the semivariogram is computed, are evaluated with respect to an average trend defined as:

$$P(R_{line}) = a + log_{10}(R_{line} + b) \qquad (1)$$

where *P* is the peak parameter of ground motion of interest (e.g., *PGA*) and R_{line} is the closest distance from the surface fault projection of the segment at the top edge of the rupture plane, which was found to be the best distance metrics for the Po Plain simulations (Hashemi *et al.*, 2015), as well as for other case studies of normal and reverse fault earthquakes (Paolucci *et al.*, 2016). Furthermore, *a* and *b* are regression coefficients calibrated either on records or on simulated results.

Figure 9 shows the semivariograms as a function of the inter-station distance from both recorded and simulated ground motions along the NS component at the accelerometric stations illustrated in Figure 5. Symbols denote the semivariogram values associated with different response spectral ordinates, specifically, *PGA*, *SA(0.2s)*, *SA(1.0s)*, *SA(2.0s)*, both for the records (crosses) and for the BB results simulated either through the ANN2BB procedure (open dots) or the HYB procedure (filled squares). The functional form chosen to fit the corresponding semivariogram data is the exponential model (Cressie, 1985), shown by continuous and dashed lines for REC and ANN2BB, respectively. In analogy with previous studies (see e.g. Jayaram and Baker, 2009; Esposito and Iervolino, 2011), to provide a better representation at short separation distances, we have decided to fit manually the semivariograms starting from the least-square estimation. On each subplot of Figure 9 the values of range resulting from the best-fitting model are indicated. Note that larger values of the range, i.e., the inter-station distance at which the correlation coefficient drops to zero, means that correlation is preserved at larger distances.

It turns out that the best-fitting exponential models on records and on the ANN2BB results are in good agreement. In both cases, the value of the range varies between 19 to 25 km, with a relative error between the two range estimates (i.e. from REC and ANN2BB) bounded between 1% (for *SA 0.2s*) and 20% (for *PGA*). This points out that the ANN2BB approach succeeds in reproducing accurately the spatial correlation structure of response spectral ordinates even at short periods. Instead, it is apparent that the application of the HYB procedure produces at short periods (see *PGA* and *SA 0.2s*) a semivariogram which is almost flat, thus denoting a zero correlation coefficient at all interstation distances. As a final remark, it is found that that the trend of ranges obtained with ANN2BB method is increasing with the vibration period, passing from 20 km for *PGA* to 24 km for *SA 2.0s*, in agreement with the other research works previously mentioned.

Although the Po Plain earthquake considered in this work provided one of the widest set of near-source records from moderate-to-large earthquakes worldwide, the number of stations has to be considered limited for the computation of the semivariograms. For this reason, it is not possible to group the stations in order to study possible anisotropies in the features of spatial correlation of ground motion, because the number of stations in each sub-group would be too small. Instead, this is possible when using the results of numerical simulations, because the number of receivers may be made arbitrarily large.

Figure 10 shows the correlation models for *PGA*, left, and *SA 1.0s*, right, obtained from both recordings and ANN2BB results. In addition to the results obtained at the accelerometric stations (solid lines), possible anisotropy patterns have been investigated by considering a sufficiently large set of synthetics receivers located in the Northern and Southern sector with respect to the fault at distances R_{line} lower than 10 km (N and S set, respectively). This figure points out an interesting feature of the ANN2BB simulated waveforms: when considering only receivers with $R_{line} < 10$ km, both in the North and South direction, spatial correlation drops to 0 faster than when the whole set of receivers is considered (i.e., correlation distances are significantly shorter). This is very clear in the intermediate-to-long period range (see e.g. right side of Figure 10, referring to $T = 1$s), while this trend is less evident at short periods (see left side of the figure, referring to *PGA*), although it still appears for the receivers lying on the surface fault projection (Figure 10, left subplot, for $R_{line} < 10$ km, Southern side).

It can be concluded that such spatial anisotropy features of peak values of earthquake ground motion are mainly related to near-source effects. More specifically, proximity to the extended seismic source produces a faster decay of spatial correlation at very short distances, owing to the small-scale spatial variability of ground motion induced by the heterogeneous fault rupture combined with complex site effects related to the approximately NS orientation of the submerged bedrock topography.

Conclusions

In this paper we introduced the ANN2BB procedure, suitable to create realistic BB waveforms from 3D physics-based numerical simulations. It turns out that the performance of this procedure is rather good, provided that the simulations are accurate within a frequency band at least extended to approximately 1.5 Hz, roughly corresponding to $T^* = 0.75$s. In such range, the ANN trained to correlate long period response spectral ordinates ($T \geq T^*$) with those at short periods, was found to provide satisfactory results. The ANN used in this work was trained on a strong motion dataset consisting of about 500 records with moment magnitude from 5 to 7.3 and epicentral distance up to 35 km, but other ANNs can be trained with a similar purpose on wider datasets. Separate ANNs were trained on the geometric mean of the horizontal components and on the vertical components to allow the prediction of three-component ground motions.

An extension of the training dataset is planned to encompass a wider range of magnitude, distance and site conditions. Furthermore, since all ANNs considered in this work are deterministic, i.e., for one set of input spectral ordinates at long period, a single set of output spectral ordinates at short period is provided, the training of stochastic ANNs is also envisioned, by defining weights and biases as random variables.

As a comprehensive validation benchmark, we considered the strong motion records obtained in the near-source region of the May 29, 2012 Po Plain earthquake and the corresponding 3D physics-based numerical simulations carried out by the spectral element code SPEED and illustrated in detail in Paolucci *et al.* (2015). Compared to a standard hybrid approach to produce BB waveforms, consisting of enriching the high-frequency portion of ground motion by a stochastic contribution, the proposed ANN2BB procedure allows one to obtain a similar realistic aspect of the waveform, both in time and frequency domains, but, in addition, it also allows one to obtain maps of short-period peak values of ground motion which reproduce more closely the coupling of source-related and site-related features of earthquake ground motion. And, as a further important asset of the proposed procedure, as also illustrated by a similar application in Thessaloniki (Smerzini and Pitilakis, 2017), it is suitable to portray in a realistic way the spatial correlation features of the peak values of ground motion also at short periods, with the possibility to point out possible spatial anisotropies, typically related to the near-source or complex geology conditions.

To conclude, we remark that, while the correlation structure of the high-frequency peak values is simulated in a satisfactory way, the procedure is not suitable yet to obtain sets of waveforms with realistic spatial coherency features at high-frequency (measured in terms of the coherency operator, see Zerva, 2009), apt for use as input motions for seismic analyses of spatially extended structures. As a matter of fact, the high-frequency stochastic contributions added to the simulated motions need to be re-phased to reproduce properly travelling waveforms. This is probably the single major limitation still existing preventing yet to provide simulated BBs fulfilling all the characteristics of a real earthquake ground motion wavefield.

Data and Resources

Strong-motion recordings of the May 29 2012 Po Plain earthquake were obtained from the ITalian ACcelerometric Archive ITACA, available at http://itaca.mi.ingv.it (last accessed January 2017).

Acknowledgements

This research activity has been funded by the Italian Department of Civil Protection, within the DPC 2014-2016 RELUIS RS2 Project, by Munich Re (Germany) and by swissnuclear (Switzerland), in the framework of research projects focused on the development of advanced physics-based numerical approaches for earthquake ground motion prediction and seismic risk analyses.

References

Akinci, A., H. Aochi, A. Herrero, M. Pischiutta, and D. Karanikas (2017). Physics-Based Broadband Ground-Motion Simulations for Probable Mw ≥ 7.0 Earthquakes in the Marmara Sea Region (Turkey), *Bull. Seism. Soc. of Am.* **107**, no. 3.

Ameri, G., F. Gallovic and F. Pacor (2012). Complexity of the Mw 6.3 2009 L'Aquila (central Italy) earthquake: 2. Broadband strong motion modeling, *J. Geophys. Res.* **117**, no. B04308, 1-18.

Anderson, J. G (2015). The Composite Source Model for Broadband Simulations of Strong Ground Motions, *Seismol. Res. Let.* **86**, no. 1, 68-74.

Atik, L.A., and N. Ambrahamson (2010). An Improved Method for Nonstationary Spectral Matching, *Earthq. Spectra* **26**, no. 3, 601-617

Bao, H., J. Bielak, O. Ghattas, L. Kallivokas, D. O'Hallaron, J. Shewchuk, and J. Xu (1998). Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers. *Comput. Methods Appl. Mech. Engrg.* **152**, no. 1, 85 – 102.

Bindi, D., M. Massa, L. Luzi, G. Ameri, F. Pacor, R. Puglia, and P. Augliera (2014). Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5 to 3.0 s using the RESORCE dataset, *Bull. Earthq. Eng.* **12**, no. 1, 391–430.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.

Bishop, C.M., and C.M. Roach (1992). Fast Curve Fitting using Neural Networks, *Rev Sci Instrum* **63**, 4450.

Boore, D.M. (2003). Simulation of Ground Motion Using the Stochastic Method, *Pure Appl. Geophys*. **160**, no. 3, 635–676.

Bradley, B.A., D. Pettinga, J.W. Baker, and J. Fraser J. (2017). Guidance on the Utilization of

Earthquake-Induced Ground Motion Simulations in Engineering Practice, *Earthq. Spectra* In-Press.

Causse, M., E. Chaljub, F. Cotton, C. Cornou and P.Y.Bard (2009). New approach for coupling k−2 and empirical Green’s functions: application to the blind prediction of broad-band ground motion in the Grenoble basin, *Geophys. J. Int.* **179**, 1627–1644

Cauzzi, C., E. Faccioli, M. Vanini, A. Bianchini (2015). Update predictive equations for broadband (0.01 to 10 s) horizontal response spectra and peak ground motions, based on a global dataset of digital acceleration records, *Bull. Earth. Eng.* **13**, no.6, 1578-612.

Chaljub, E., E. Maufroy, P. Moczo, J. Kristek, F. Hollender, P.Y. Bard, E. Priolo, P. Klin, F. de Martin, Z. Zhang, W. Zhang, and X. Chen (2015). 3-D numerical simulations of earthquake ground motion in sedimentary basins: testing accuracy through stringent models, *Geophys. J. Int.* **201**, no. 1, 90–111.

Crempien, J.G.F., and R. J. Archuleta (2015). UCSB Method for Simulation of Broadband Ground Motion from Kinematic Earthquake Sources, *Seismol. Res. Lett*. **86**, no. 1, 61-67.

Cressie, N. (1985). Fitting variogram models by weighted least squares, *Math. Geol.* **17**, no. 5, 563-585

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems* **2**, no. 4, 303–314.

Day, S.M., R. Graves, J. Bielak, D. Dreger, S. Larsen, K.B. Olsen, A. Pitarka, and L. Ramirez-Guzman (2008). Model for basin effects on long-period response spectra in southern California, *Earthq. Spectra* **24**, no. 1, 257–277.

Douglas, J., and H. Aochi (2008). A Survey of Techniques for Predicting Earthquake Ground Motions for Engineering Purposes, *Surv. Geophys.* **29**, no. 3, 187–220

Dumbser, M., and M. Käser (2006). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - II. The three-dimensional isotropic case, *Geophys. J. Int.* **167**, no. 1, 319–336.

Esposito, S., and I. Iervolino (2011). PGA and PGV Spatial Correlation Models Based on European Multievent Datasets, *Bull. Seismol. Soc. Am*. **101**, no. 5, 2532–2541

Esposito, S., and I. Iervolino (2012). Spatial Correlation of Spectral Acceleration in European Data, *Bull. Seismol. Soc. Am.* **102**, no. 6, 2781–2788

Gatti, F. (2017). Forward physics-based analysis of "source-to-site" seismic scenarios for strong ground motion prediction and seismic vulnerability assessment of critical structures. *PhD Thesis*, CentraleSupélec – Politecnico di Milano, Paris ,France, and Milan, Italy (https://tel.archives-ouvertes.fr/tel-01626230).

Gatti, F., L. De Carvalho Paludo, A. Svay., F. Lopez-Caballero, R. Cottereau ,and D. Clouteau (2017). Investigation of the earthquake ground motion coherence in heterogeneous non-linear soil deposits, *X International Conference on Structural Dynamics*, EURODYN 2017, 10 - 13 September.

Goulet, C.A., N.A. Abrahamson, P.G. Somerville, and K E. Wooddell (2015). The SCEC Broadband Platform Validation Exercise: Methodology for Code Validation in the Context of Seismic-Hazard Analyses, *Seismol. Res. Lett.* **86**, no. 1.

Graves, R. W. and A. Pitarka (2010). Broadband Ground Motion Simulation Using a Hybrid Approach, *Bull. Seismol. Soc. Am.* **100**, no. 5A, 2095–2123.

Graves, R.W. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. Seismol. Soc. Am.* **86**, no. 4, 1091–1106.

Hartzell, S.H. (1978). Earthquakes aftershocks as Green’s functions, *Geophys. Res. Lett.*, **5**, 1–4.

Hashemi, K., I. Mazzieri, R. Paolucci, and C. Smerzini (2015). Spatial variability of near-source seismic ground motion with respect to different distance metrics, with special emphasis on May 29 2012 Po Plain Earthquake, Italy,*7th International Conference on Seismology and Earthquake Engineering*.

Herrero, A. and P. Bernard (1994). A kinematic self-similar rupture process for earthquakes, *Bull. Seismol. Soc. Am.* **84**, no. 4, 1216–1228.

Hisada, Y, and J. Bielak (2003). A theoretical method for computing near-fault ground motions in layered half-spaces considering static offset due to surface faulting, with a physical interpretation of fling step and rupture directivity, *Bull. Seismol. Soc. Am.* **93**, no. 3,1154–1168

Isbiliroglu, Y., R. Taborda, and J. Bielak (2015). Coupled Soil-Structure Interaction Effects of Building Clusters During Earthquakes, *Earthq. Spectra* **31**, no. 1, 463-500.

Iwaki, A., M. Takahiro, N. Morikawa, H. Miyake, and H. Fujiwara (2016). Validation of the Recipe for Broadband Ground Motion Simulations of Japanese Crustal Earthquakes, *Bull. Seismol. Soc. Am*. **106**, no. 5, 2214-2232.

Jayaram, N., and J.W. Baker (2009). Correlation model for spatially distributed ground motion intensities, *Earthq. Eng. Struct. Dynam*. **38**, no. 15, 1687–1708.

Kamae, K., K. Irikura, and A. Pitarka (1998). A technique for simulating strong ground motion using hybrid Green's function, *Bull. Seism. Soc. Am.* **88**, no. 2, 357-367.

Komatitsch, D., and J. Tromp (2002a). Spectral-element simulations of global seismic wave propagation-I. Validation, *Geophys. J. Int.* **149**, no. 2, 390–412.

Komatitsch, D., and J. Tromp (2002b). Spectral-element simulations of global seismic wave propagation-II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys.*

J. Int. **150**, no. 1, 303–318.

Lanzano, G., M. D'Amico, C. Felicetta, R. Puglia, L. Luzi, F. Pacor, and D. Bindi (2016). Ground-Motion Prediction Equations for Region-Specific Probabilistic Seismic-Hazard Analysis, *Bull. Seismol. Soc. Am.* **106**, no. 1, 73–92.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares, *Quart. Appl. Math.* **2**, no. 2, 164–168.

Loth, C., and J.W. Baker (2013). A spatial cross-correlation model of spectral accelerations at multiple periods, *Earthq. Eng. Struct. Dynam.* **42**, no. 3, 397-471

Luzi, L., F. Pacor, G. Ameri, R. Puglia, P. Burrato, M. Massa, P. Augliera, G. Franceschina, S. Lovati, and R. Castro (2013). Overview on the strong-motion data recorded during the May–June 2012 Emilia seismic sequence, *Seism. Res. Lett*. **84**, no. 4, 629–644.

Mai, P.M., and G.C. Beroza (2003). A hybrid method for calculating near-source, broadband seismograms: Application to strong motion prediction, *Phys. Earth Planet In.* **137**, no. 1-4,183–199.

Mai, P.M., W. Imperatori, and K.B. Olsen (2010). Hybrid broadband ground-motion simulations: Combining long-period deterministic synthetics with high-frequency multiple S-to-S backscattering, *Bull. Seismol. Soc. Am.* **100**, no. 5A, 2124–2142.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.* **11**, no. 2, 431–441.

Matheron, G. (1965). Les variables régionalisées et leur estimation, Masson, Paris

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. Seismol. Soc. Am.* **105**, 787-808.

Mazzieri, I., M. Stupazzini, R. Guidotti, and C. Smerzini (2013). SPEED: SPectral Elements in Elastodynamics with Discontinuous Galerkin: a non-conforming approach for 3D multi-scale problems, *Int. J. Numer. Meth. Eng*. **95**, no. 12, 991–1010.

McClelland, J.L., D. E. Rumelhart, and the PDP Research Group (1986). Parallel distributed processing: Explorations in the microstructure of cognition, *Volume 1: Foundations*, The MIT Press, Cambridge.

Mena, B., P.M. Mai, K.B. Olsen, M.D. Purvance and J.N. Brune (2010). Hybrid broadband ground-motion simulation using scattering green's functions: application to large-magnitude events, *Bull. Seismol. Soc. Am*. **100**, 2143–2162.

Motazedian, D., and G.M. Atkinson (2005). Stochastic Finite-Fault Modeling based on dynamic corner frequency, *Bull. Seismol. Soc. Am.* **95**, no. 3, 995–1010.

NIST - National Institute of Standards and Technology (2011). Selecting and Scaling Earthquake Ground Motions for Performing Response-History Analyses, Technical Report NIST GCR 11-917-15, prepared for the Engineering Laboratory of the National Institute of Standards and Technology (NIST) under the National Earthquake Hazards Reduction Program (NEHRP) Earthquake Structural and Engineering Research Contract SB134107CQ0019, Task Order 69220, November 2011

Olsen, K.B. (2000). Site amplification in the Los Angeles basin from three-dimensional modeling of ground motion, *Bull. Seismol. Soc. Am.* **90**, no. 6B, S77–S94.

Paolucci, R., L. Evangelista, I. Mazzieri, E. Schiappapietra (2016). The 3D numerical simulation of near-source ground motion during the Marsica earthquake, central Italy, 100 years later, *Soil Dynam. Earthq. Eng.* **91**, 39-52.

Paolucci, R., I. Mazzieri, and C. Smerzini (2015). Anatomy of strong ground motion: near-source records and 3D physics-based numerical simulations of the Mw 6.0 May 29 2012 Po Plain earthquake, Italy, *Geophys. J. Int.* **203**, 2001–2020.

Pitarka, A., K. Irikura, T. Iwata, and H. Sekiguchi (1998). Three-dimensional simulation of the near-fault ground motion for the 1995 Hyogo-Ken Nanbu (Kobe), Japan, earthquake, *Bull. Seismol. Soc. Am.* **88**, no. 2, 428–440.

Razafindrakoto, H.N.T., B.A. Bradley., R.W. Graves (2016). Broadband Ground Motion Simulation of the 2010- 2011 Canterbury Earthquake Sequence, *2016 NZSEE Conference.*

Ramirez-Guzman, L., R.W. Graves, K.B. Olsen, O.S. Boyd, C. Cramer, S. Hartzell, S. Ni, P. Somerville, R.A. Williams and J. Zhong (2015). Ground-Motion Simulations of 1811-1812 New Madrid Earthquakes, Central United States, *Bull. Seismol. Soc. Am*. **105**, no. 4, 1961-1988.

Roten D., K. B. Olsen, and J. C. Pechmann (2012). 3D Simulations of M 7 Earthquakes on the Wasatch Fault, Utah, Part II: Broadband (0–10 Hz) Ground Motions and Nonlinear Soil Behavior, *Bull. Seism. Soc. Am*. **102**, no. 5, 2008–2030.

Sabetta, F., and A. Pugliese (1996). Estimation of Response Spectra and Simulation of Nonstationary Earthquake Ground Motions, *Bull. Seismol. Soc. Am.* **86**, no. 2, 337–352.

Seyhan, E., J.P. Stewart, and R.W. Graves (2013). Calibration of a Semi-Stochastic Procedure for Simulating High-Frequency Ground Motions, *Earthq. Spectra* **29**, no. 4, 1495–1519.

Shahbazian, A., and S. Pezeshk (2010). Improved Velocity and Displacement Time Histories in Frequency Domain Spectral-Matching Procedures, *Bull. Seism. Soc. Am.* **100**, no. 6, 3213-3223.

Smerzini, C., and M. Villani (2012). Broadband Numerical Simulations in Complex Near-Field Geological Configurations: The Case of the 2009 Mw 6.3 L'Aquila Earthquake, *Bull. Seismol. Soc. Am.* **102**, no. 6, 2436–2451.

Smerzini, C., C. Galasso, I. Iervolino, and R. Paolucci (2014). Ground motion record selection based on broadband spectral compatibility, *Earthq. Spectra* **30**, no. 4, 1427–1448.

Smerzini, C. and Pitilakis K. (2017). Seismic risk assessment at urban scale from 3D physics-based numerical modeling: the case of Thessaloniki, *Bull. Earthq. Eng.,* doi: https://doi.org/10.1007/s10518-017-0287-3.

Spudich, P., and L. Xu (2002). Software for calculating earthquake ground motions from finite faults in vertically varying media, in *International Handbook of Earthquake and Engineering Seismology* W. H. K. Lee, H. Kanamori, P. Jennings, and C. Kisslinger (Editors), Vol. 2, Academic Press, Orlando, Florida.

Taborda, R., and J. Bielak (2014). Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake using different velocity models, *Bull. Seismol. Soc. Am*. **104**, no. 4, 1876–1898.

Tsuda, K., T. Hayakawa, T. Uetake, K. Hikima, R. Tokimitsu, H. Nagumo, and Y. Shiba (2011). Modeling 3D Velocity Structure in the Fault Region of the 2007 Niigataken Chuetu-Oki Earthquake with Folding Structuree, 4t*h IASPEI/IAEE International Symposium-Effects of Surface Geology on Seismic Motion*, 1–11.

Villani, M., E. Faccioli, M. Ordaz, and M. Stupazzini (2014). High-Resolution Seismic Hazard Analysis in a Complex Geological Configuration: The Case of the Sulmona Basin in Central Italy, *Earthq. Spectra* **3**, no. 4, 1801–1824.

Wald, D.J., and R.W. Graves (1998). The seismic response of the Los Angeles basin, California, *Bull. Seismol. Soc. Am.* 88, no. 2, 337–35

Webster, R., and M. A. Oliver (2007). *Geostatistics for Environmental Scientists, Second Edition*, John Wiley & Sons, Ltd

Zerva, A. (2009). *Spatial variation of seismic ground motions Modeling and engineering applications*, CRC Press, Boca Raton.

Full mailing address for each author

Roberto Paolucci

Department of Civil and Environmental Engineering, Politecnico di Milano

Piazza Leonardo da Vinci, 32 – 20133 Milano.

E-mail: roberto.paolucci@polimi.it.

Filippo Gatti

Laboratoire de Mécanique des Sols, Structures et Matériaux, CentraleSupélec - Université Paris-Saclay, 3, rue Joliot Curie, 91190, Gif-Sur-Yvette, France.

Email: filippo.gatti@centralesupelec.fr

Maria Infantino (Corresponding Author)

Department of Civil and Environmental Engineering, Politecnico di Milano

Piazza Leonardo da Vinci, 32 – 20133 Milano.

E-mail: maria.infantino@polimi.it.

Chiara Smerzini

Department of Civil and Environmental Engineering, Politecnico di Milano

Piazza Leonardo da Vinci, 32 – 20133 Milano.

E-mail: chiara.smerzini@polimi.it.

Ali Güney Özcebe

Department of Civil and Environmental Engineering, Politecnico di Milano

Piazza Leonardo da Vinci, 32 – 20133 Milano.
E-mail: aliguney.ozcebe@polimi.it.

Marco Stupazzini
Munich Re, Königinstraße 107, 80802 München, Germany.
E-mail: mstupazzini@munichre.com

List of Figure Captions

Figure 1. Main idea behind the proposed ANN-based approach to generate BB ground motions: for a given ground motion, response spectral ordinates at short periods, i.e., for periods $T \leq T^*$, where T^* is the minimum period of validity of the physics-based numerical model, are computed from the 3D physics-based simulated response spectral ordinates at long periods.

Figure 2. Logic scheme of the ANN training patterns: the long period spectral ordinates (in this case T^* = 0.75 s) represent the teaching inputs, whereas the short period ones are the outputs predicted by the ANN. The number of neurons in the hidden layer is $N_n^h = 30$.

Figure 3. ANN performance in predicting the horizontal components of SIMBAD records (geometric mean of horizontal components), expressed in terms of $\log_{10}(SA_{ANN}/SA_{Obs})$, where SA_{ANN} denotes the response spectral ordinates predicted by the ANN and SA_{Obs} is the observed ones. The performance is estimated at each vibration period T, here normalized with respect to the corner period T^*. The error bars refer to the training (TRN), validation (VLD) and test (TST) set.

Figure 4. Same as Figure 3, but for the ANN trained on the vertical components of records of the SIMBAD dataset.

Figure 5. Map of *PGA* (geometric mean of horizontal components) obtained by a) the proposed ANN2BB approach, b) the hybrid (HYB) approach by combining the PBS with the stochastic signals from the Sabetta and Pugliese (1996) method, c) the PBS filtered at 1.5 Hz. The filled dots superimposed on each map denote the *PGA* values recorded by the available strong-motion stations. d) ANN2BB simulated vs recorded values of *PGA* as a function of the Joyner-Boore distance, R_{JB}, in comparison with the GMPE of Bindi *et al.* (2014), BI14.

Figure 6. From left to right, acceleration, velocity and displacement time histories of NS

component at Mirandola (MRN) station, 29 May 2012 Po Plain earthquake. From the top to the bottom the subpanels show: (i) the physics based numerical simulation (PBS) filtered with f_c = 1.5 Hz; (ii) stochastic waveform (STO) according to Sabetta and Pugliese (1996); (iii) hybrid synthetics (HYB) obtained by coupling PBS at low frequency and STO at high frequency with a cross-over frequency of f_c = 1.5 Hz; (iv) BB synthetics (ANN2BB) resulting by scaling the HYB upon the ANN-based short-period spectral ordinates; (v) records (REC).

Figure 7. Same comparison as in Figure 6 but in terms of acceleration response (left) and Fourier Spectra (right) of NS component at Mirandola (MRN) station.

Figure 8. 2012 Po Plain earthquake simulation**:** comparison between ANN2BB simulations and recordings at four accelerometric stations (MIR08, T0802, BON0 and MOG0) in terms of NS acceleration and velocity time histories (top panels) and acceleration response spectra (bottom). The location of the selected stations is shown in Figure 5.

Figure 9. Semivariograms obtained using records REC (crosses) and the ANN2BB approach (circle) for *PGA* (top left), *SA 0.2s* (top right), *SA 1.0s* (bottom left) and *SA 2.0s* (bottom right). The corresponding best-fitting exponential models are denoted by solid line and dashed line for REC and ANN2BB, respectively. Moreover, for the short period response spectral ordinates (see top panel), the semivariograms (filled squares) and the corresponding best-fitting model (solid line) from HYB results are also shown for comparison.

Figure 10. Spatial correlation models, $\rho(h)$, obtained from the REC and ANN2BB values obtained at the accelerometric stations (solid lines) for *PGA* (left) and *SA 1.0s* (right). The dash and dash dot lines show the correlation models computed using a larger number of ANN2BB receivers located in the Northern (N) and Southern (S) side with respect to the fault at $R_{line} < 1$

Tables

Table 1. Selection of BB earthquake ground motion simulation case studies relying on hybrid approaches*.

Publications	f_c [Hz]	Methods (LF + HF)	Area under study	Validation
Causse *et al*., 2009	1.0	SE + EGF	Grenoble, France	against GMPE
Graves and Pitarka, 2010	1.0	FD + SFF	California, USA	M6.4, Imperial Valley, 1979 M6.9, Loma Prieta 1989 M7.3, Landers, 1992 M6.7, Northridge, 1994
Mena *et al*., 2010	0.5	FD + Sc-GF	San Andreas fault, California, USA	against GMPE
Roten *et al*., 2012	1.0	FD + Sc-GF	Salt Lake City, Utah, USA	against GMPE
Smerzini and Villani, 2012	2.5	SE + SFF	L'Aquila, Italy	M6.3, L'Aquila, 2009
Seyhan *et al*., 2013	1.0	FD + SFF	California, USA	against GMPE
Ramirez-Guzman *et al.* 2015	1.0	FD, FE + SFF, St-GF	New Madrid seismic zone, USA	against GMPE
Iwaki *et al*., 2016	1.0	FD + St-GF	Japan	M6.7, Tottori, 2000 M6.6, Chuetsu, 2004
Razafindrakoto *et al*., 2016	1.0	FD + SFF	Christchurch area, New Zealand	2010-2011 earthquake sequence
Akinci *et al*., 2017	1.0	FD + SFF	Marmara Sea, Turkey	against GMPE

* Low-frequency (LF) methods: FD = Finite Difference; FE = Finite Element, SE = Spectral Element. High-frequency (HF) methods: SFF = stochastic finite-fault (Boore, 2003; Motazedian and Atkinson, 2005; Graves and Pitarka, 2010); EGF = Empirical Green's functions (Hartzell, 1978); Sc-GF = scattering Green's functions (Mai *et al*., 2010); St-GF = stochastic Green's functions (Kamae *et al*., 1998). f_c denotes the cross-over frequency where low frequency (from PBS) and high frequency (stochastic) synthetics are combined.

Figures

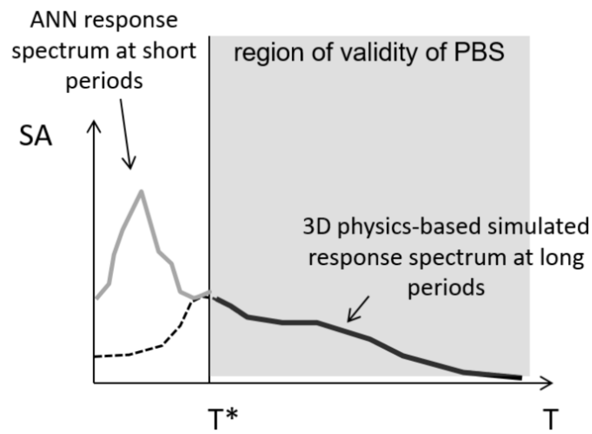

Figure 1. Main idea behind the proposed ANN-based approach to generate BB ground motions: for a given ground motion, response spectral ordinates at short periods, i.e., for periods *T*≤*T**, where *T** is the minimum period of validity of the physics-based numerical model, are computed from the 3D physics-based simulated response spectral ordinates at long periods.

0 km.

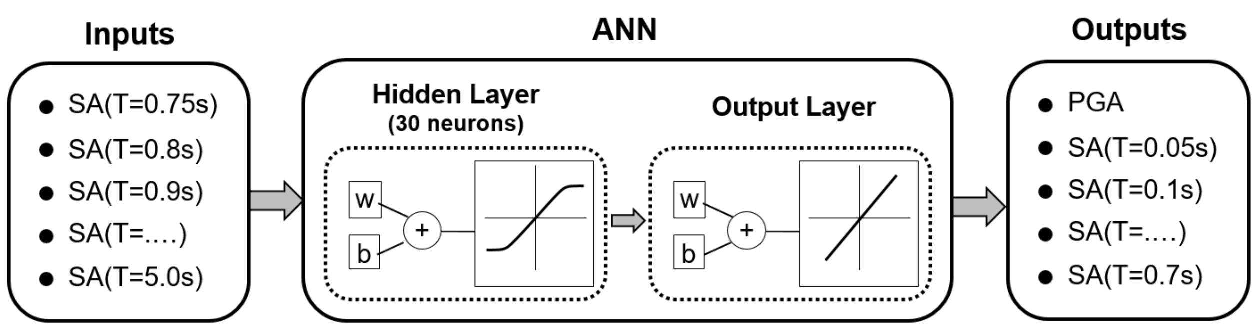

Figure 2. Logic scheme of the ANN training patterns: the long period spectral ordinates (in this case *T** = 0.75 s) represent the teaching inputs, whereas the short period ones are the outputs predicted by the ANN. The number of neurons in the hidden layer is N_n^h = 30.

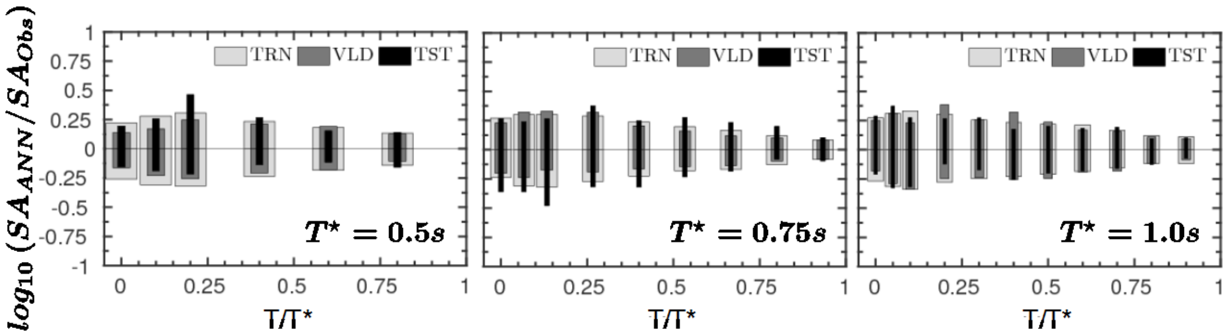

Figure 3. ANN performance in predicting the horizontal components of SIMBAD records (geometric mean of horizontal components), expressed in terms of $\log_{10}(SA_{ANN}/SA_{Obs})$, where SA_{ANN} denotes the response spectral ordinates predicted by the ANN and SA_{Obs} is the observed ones. The performance is estimated at each vibration period *T*, here normalized with respect to the corner period *T**. The error bars refer to the training (TRN), validation (VLD) and test (TST) set.

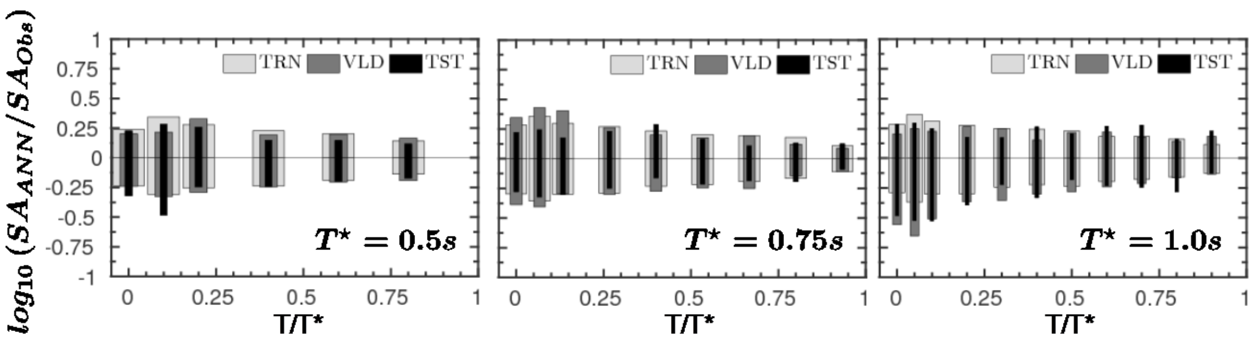

Figure 4. Same as Figure 3, but for the ANN trained on the vertical components of records of the SIMBAD dataset.

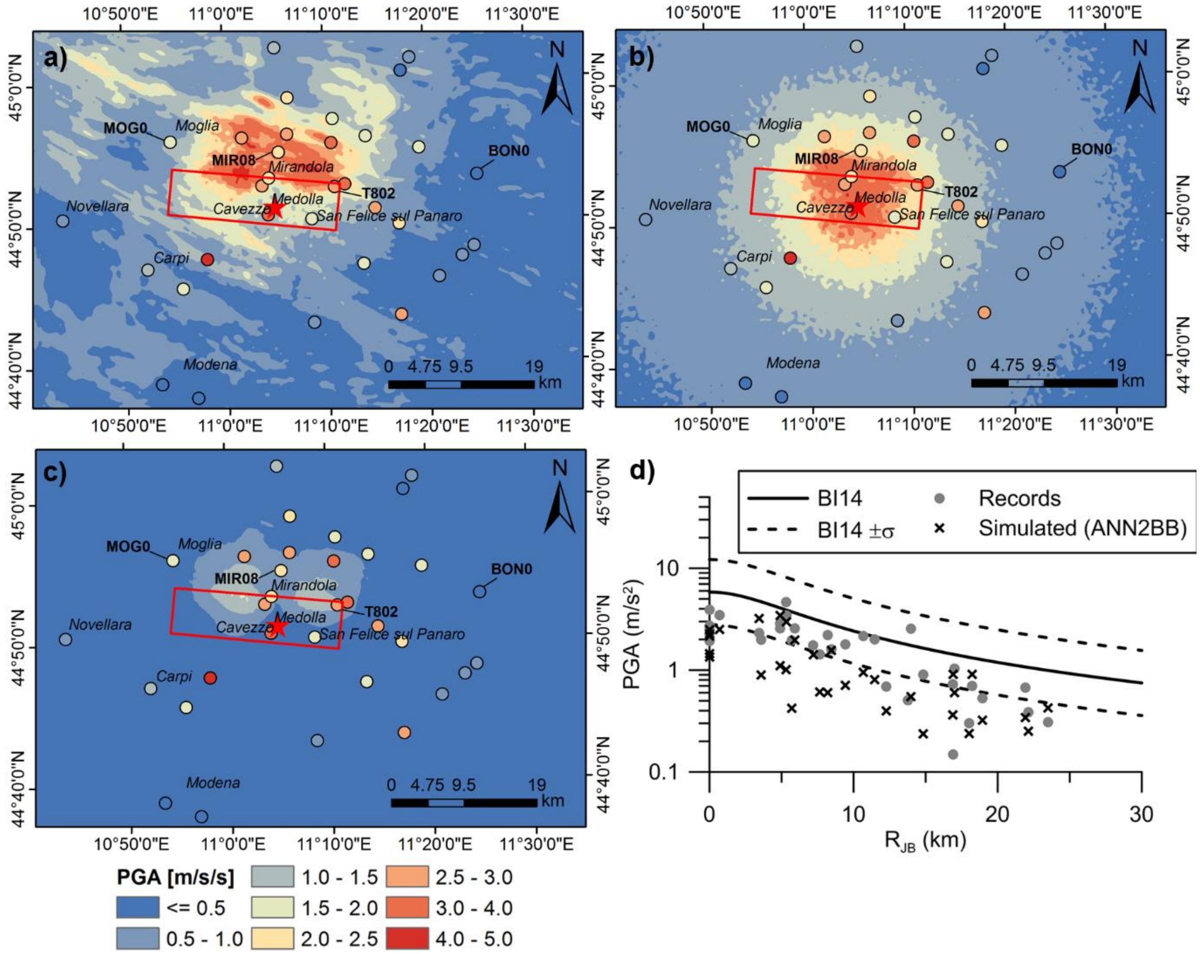

Figure 5. Map of *PGA* (geometric mean of horizontal components) obtained by a) the proposed ANN2BB approach, b) the hybrid (HYB) approach by combining the PBS with the stochastic signals from the Sabetta and Pugliese (1996) method, c) the PBS filtered at 1.5 Hz. The filled dots superimposed on each map denote the *PGA* values recorded by the available strong-motion stations. d) ANN2BB simulated vs recorded values of *PGA* as a function of the Joyner-Boore distance, R_{JB}, in comparison with the GMPE of Bindi *et al.* (2014), BI14.

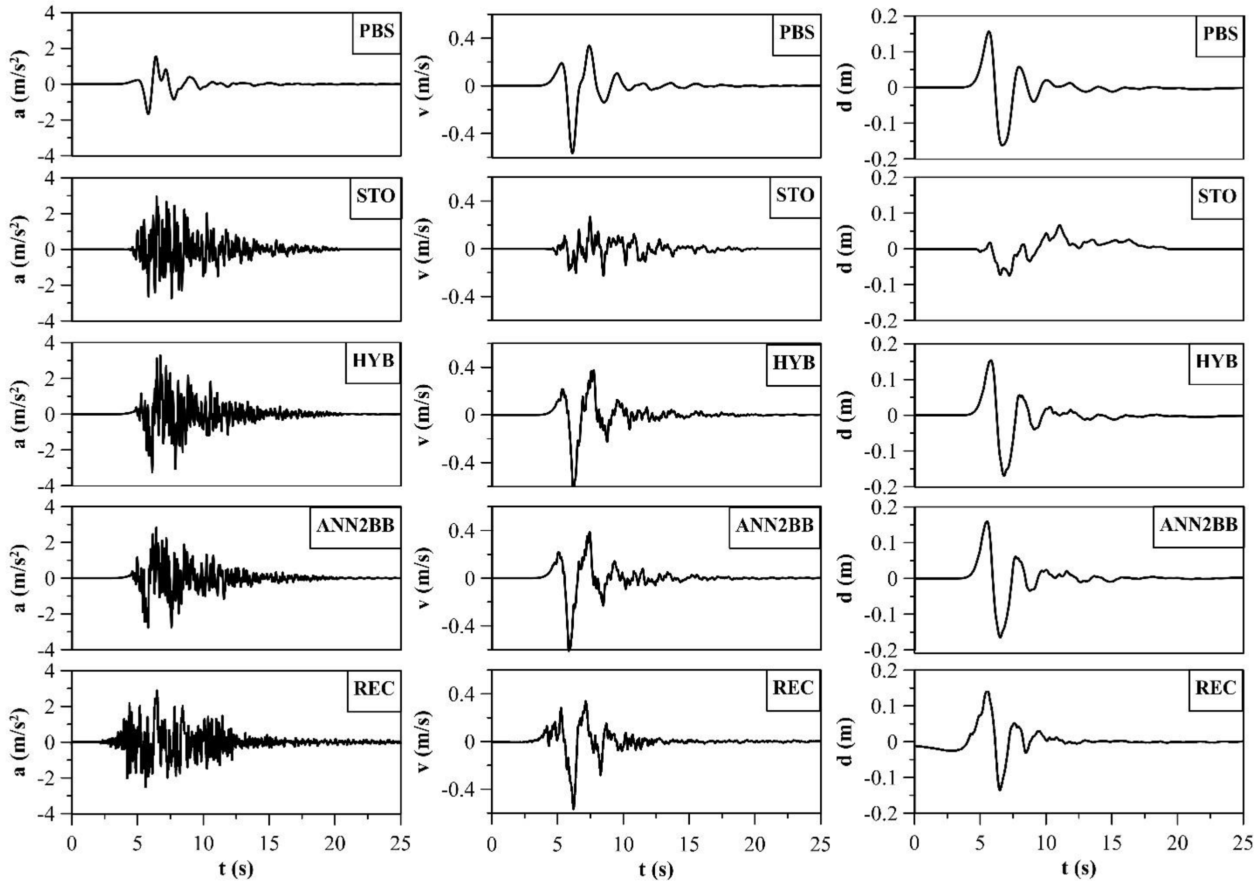

Figure 6. From left to right, acceleration, velocity and displacement time histories of NS component at Mirandola (MRN) station, 29 May 2012 Po Plain earthquake. From the top to the bottom the subpanels show: (i) the physics based numerical simulation (PBS) filtered with f_c = 1.5 Hz; (ii) stochastic waveform (STO) according to Sabetta and Pugliese (1996); (iii) hybrid synthetics (HYB) obtained by coupling PBS at low frequency and STO at high frequency with a cross-over frequency of f_c = 1.5 Hz; (iv) BB synthetics (ANN2BB) resulting by scaling the HYB upon the ANN-based short-period spectral ordinates; (v) records (REC).

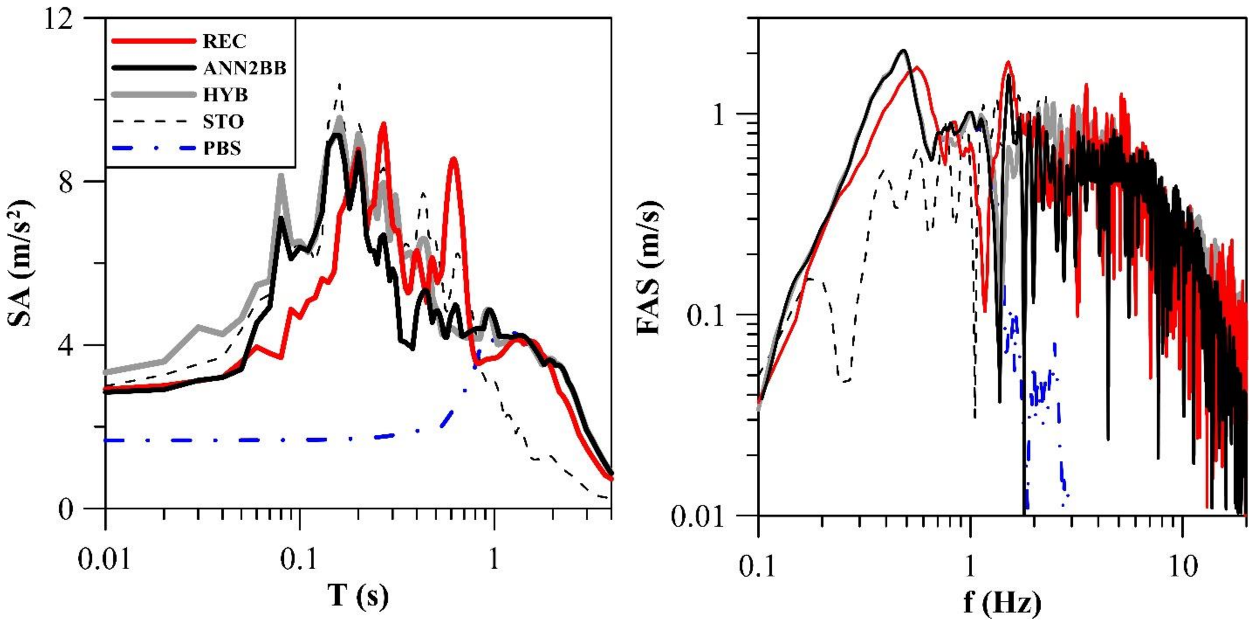

Figure 7. Same comparison as in Figure 6 but in terms of acceleration response (left) and Fourier Spectra (right) of NS component at Mirandola (MRN) station.

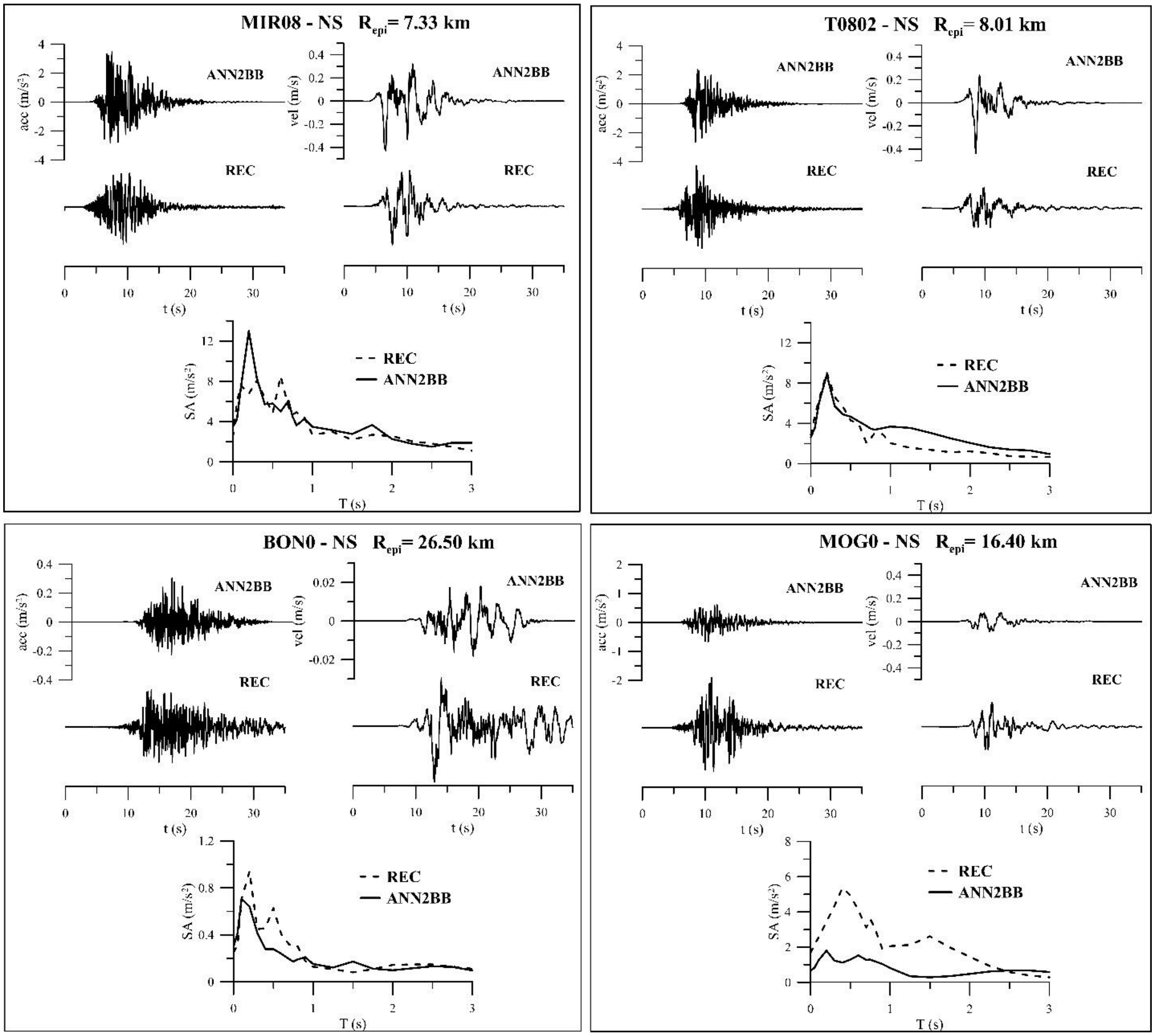

Figure 8. 2012 Po Plain earthquake simulation**:** comparison between ANN2BB simulations and recordings at four accelerometric stations (MIR08, T0802, BON0 and MOG0) in terms of NS acceleration and velocity time histories (top panels) and acceleration response spectra (bottom). The location of the selected stations is shown in Figure 5.

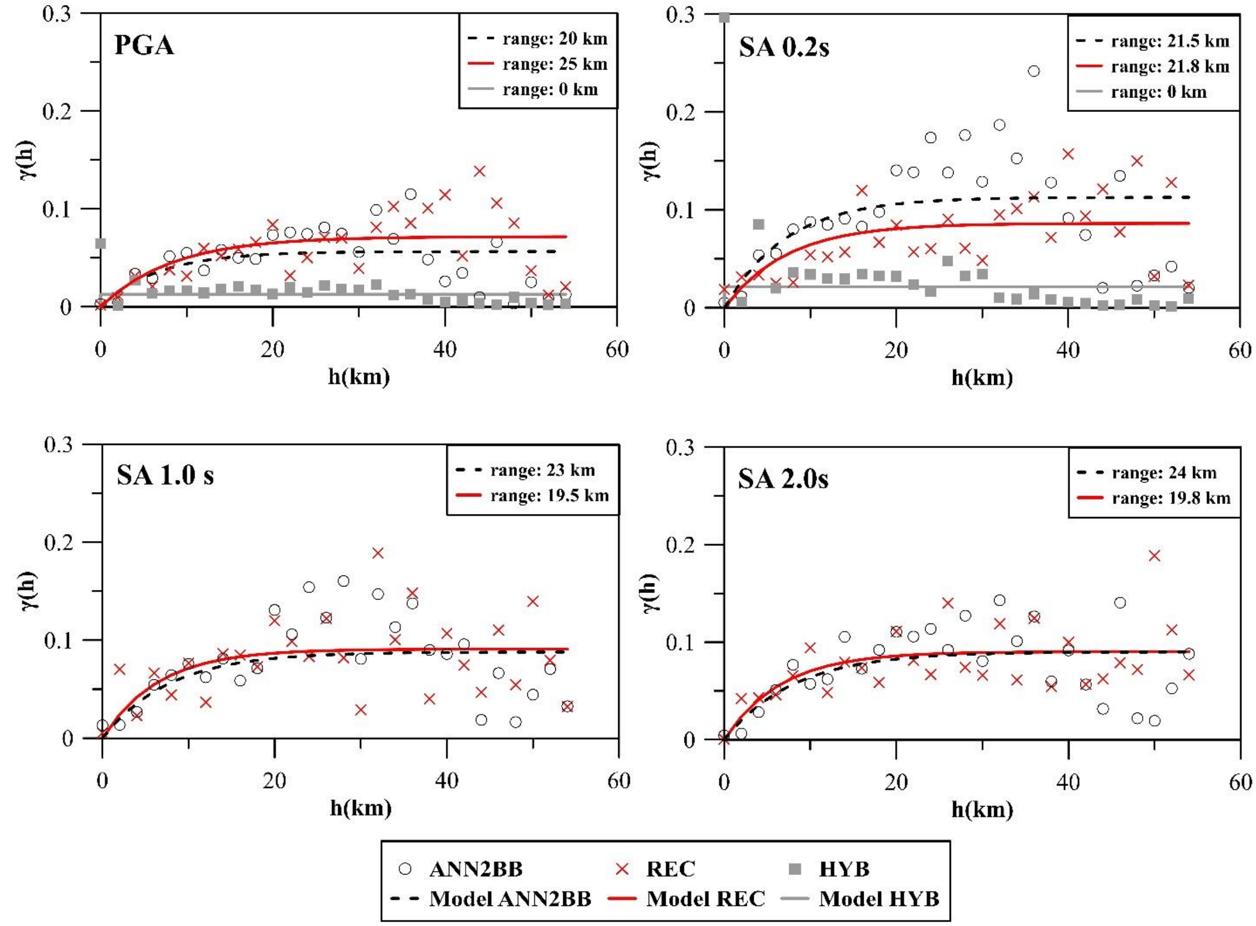

Figure 9. Semivariograms obtained using records REC (crosses) and the ANN2BB approach (circle) for *PGA* (top left), *SA(0.2s)* (top right), *SA(1.0s)* (bottom left) and *SA(2.0s)* (bottom right). The corresponding best-fitting exponential models are denoted by solid line and dashed line for REC and ANN2BB, respectively. Moreover, for the short period response spectral ordinates (see top panel), the semivariograms (filled squares) and the corresponding best-fitting model (solid line) from HYB results are also shown for comparison.

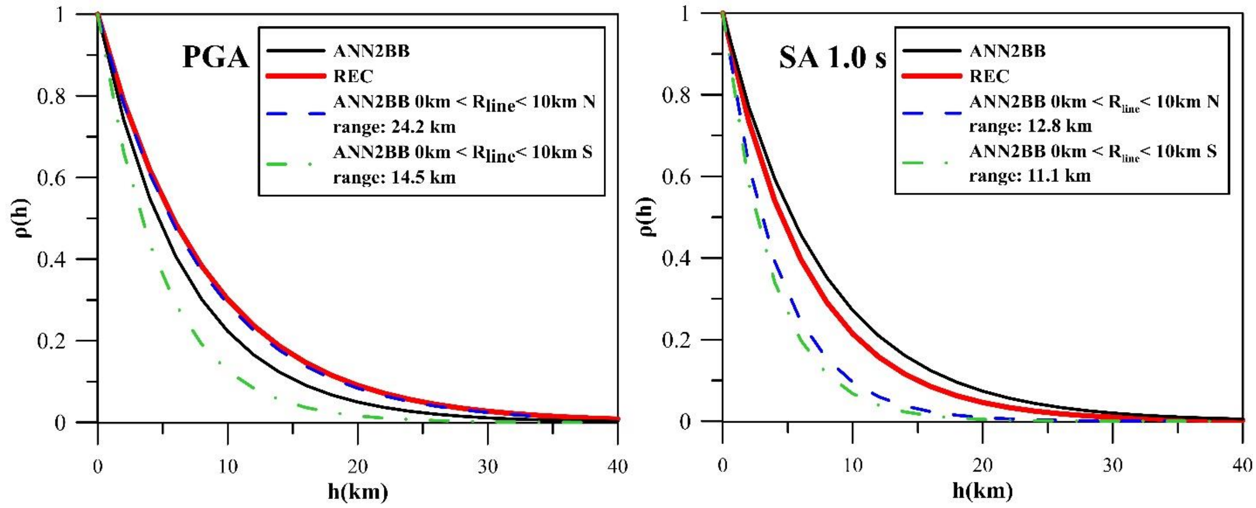

Figure 10. Spatial correlation models, $\rho(h)$, obtained from the REC and ANN2BB values obtained at the accelerometric stations (solid lines). The dash and dash dot lines show the correlation models computed using a larger number of ANN2BB receivers located in the Northern (N) and Southern (S) side with respect to the fault at R_{line} < 1