ON THE SENSITIVITY OF GROUND MOTION PREDICTION EQUATIONS IN GREECE

Segou M., Voulgaris N., Makropoulos K.

*Department of Geophysics-Geothermics, Faculty of Geology and Geoenvironment,
Panepistimioupoli Zografou, Athens - Greece 15784, msegou@geol.uoa.gr, voulgaris@geol.uoa.gr,
kmacrop@geol.uoa.gr*

Abstract

Ground motion prediction equations, widely known as attenuation relations, are common input for probabilistic and deterministic seismic hazard studies. The construction of a ground motion model to describe such a complex phenomenon as the effects of seismic wave propagation is highly dependable on a number of parameters. The quality and the distribution of strong motion data, which is the original input for the calculation of any ground motion model, can be thought as one of the main parameters that heavily influence the form of ground motion prediction equations. The selected processing scheme, involving significant choices about a series of adjustments and filter specifications, implemented to remove low and high frequency noise, is related with the credibility of the calculated ground motion parameters such as the spectral ordinates. Once a set of response variables for a number of predictors is available, the researcher's interest is related with the mathematical definition of the ground motion model, in terms of selecting the appropriate parameters and the determination of their coefficients of the equation. Another significant part involves the selection of the optimum solver in order to achieve high confidence level coefficients and a computationally inexpensive solution. Each method should be evaluated through statistics but the researcher should bear in mind that residual analysis and statistical errors, although they can adequately represent the efficiency of the mathematical equations, do not always provide information about where our efforts should lie in terms of further improvement. The scope of this paper is to point out the multi-parametric nature of the construction of ground motion prediction equations and how each of the aforementioned development stages influences the credibility of the proposed attenuation relations.

Key words: *processing scheme, corrected acceleration, spectral ordinates, attenuation, mathematical model, earthquake engineering.*

1. Introduction

The estimation of ground motion is an important tool for engineering seismology especially in terms of seismic hazard studies. Since the Long Beach earthquake in 1933 the scientific interest lies in calculating ground motion displacement from the accelerogram. During the last decades many researchers (Converse and Brady, 1992; Boore, 2008, Boore and Bommer, 2005) suggested different processing schemes for deriving corrected time series and spectral ordinates. The credibility of spectral acceleration is directly linked with the construction of ground motion prediction relations. In the effort of deriving an elaborate but subtle processing scheme a number of tests have been conducted and their results helped determining the suggested processing procedure. In addition, a sensitivity

test has been performed partly through statistical analysis, exploring how spectral ordinates can be influenced by the processing scheme.

In this paper we evaluate the multi-parametric nature of ground motion prediction equations construction which involves the determination of the mathematical model and the selection of the appropriate algorithm for solving the strong ground motion equation.

2. Data resources

The database used in this study consists of 434 three-component strong motion records provided – in their uncorrected form- by the internet site of the European Strong Motion Database[1], the Geodynamic Institute of the National Observatory of Athens[2] and ITSAK[3]. Strong motion records had been reprocessed following a subtle processing scheme to ensure minimum influence on the derived spectral ordinates.

During the first steps of the database's development a larger number of strong motion records has been closely reviewed and a quality factor has been assigned by following a number of criteria listed in Douglas (2003) in order to exclude poor quality records from ground motion prediction equation development.

Metadata information plays an important role in compiling a database for engineering seismology studies. In terms of faulting type seismic events are characterized as having normal, strike slip and thrust or reverse rupture. However, site conditions is one of the most critical parameters. Soil site conditions are characterized by following the NEHRP site categorization using the value of the shear wave velocity on the upper 30 m ($Vs30$) as a criterion for assigning a site class. However, the problem lies in the availability of Vs30 since only 42 percent corresponds to a measured geophysical parameter. A list of criteria has been issued by following the standards of NGA project (Darragh and Power, 2005) with some modifications to assign a Vs30 value in sites with missing metadata information.

In the following table we present the distribution of strong motion data through site classes and faulting types, after assigning critical metadata information.

Table 1.

Site class	NEHRP	Normal	Reverse	Strike-Slip	Total
ROCK	A&B	42	24	20	86 (~19.81%)
STIFF SOIL	C	87	27	96	220(`50.69%)
SOFT SOIL	D&E	54	30	54	138(~31.79%)
		183 (~42,16%)	81 (~18,66)	170 (~39,17%)	

---

[1] http://www.isesd.cv.ic.ac.uk

[2] http://www.gein.noa.gr

[3] http://www.itsak.gr

Ψηφιακή Βιβλιοθήκη Θεόφραστος - Τμήμα Γεωλογίας. Α.Π.Θ.

3. Processing Scheme

The final database contains strong motion data from Greece, Italy, Turkey and Iran which have been reprocessed by individual component. Although the majority of the data used comes from Greece, 271 records, in order to achieve a more complete magnitude-distance distribution of the data the final database included a number of events from the aforementioned countries of similar seismotectonic environment. For homogenising the magnitude scales we have used the re-determined epicenter coordinates and body wave magnitude available at the internet site of the International Seismological Center. It is noted that Scordilis (2006) empirical relation had been used for converting into moment magnitude. Strong motion record processing has been performed with Proschema software (Segou and Voulgaris, 2010). Instrument adjustment has been applied for analogue records whereas for digital records this was not necessary due to the flat instrument response over a wide period range. Filter design involved a fourth order bandpass Butterworth followed by phase preserving implementation with varying cut off frequencies in order to sufficiently remove the noise without damaging the frequency content of interest. In some cases before filtering a cosine taper was applied at the edges of the time series to reduce Gibb's effect. Two main points closely related to filtering should be pointed out; the first one is related to the selection of the integration operator since ground motion velocity and displacement should be eventually provided and the second one to the decision whether computations should take place in the time or frequency domain. Concerning the first point results favoured the use of the ideal integrator, with frequency response equal $1/i\omega$ whereas for the second question the authors believe that frequency domain implementation is more consistent. Of course Plancerel theorem supports either choice but since both the noise identification and evaluation of filter implementation afterwards, involve Fourier spectrum representation, frequency domain computation sufficiently deals with the problem in hand. Additionally, frequency domain computation is more time effective which is an important aspect when processing a great number of strong motion records.

Response spectra have been calculated through the analytical solution of Nigam and Jennings (1969) for damping values equal 0%, 2%, 5%, 10% and 20% of the critical damping. In order to avoid uncertainty, related at such an early stage with the orientation of the instrument's transducers, we have calculated response spectra for a damping value equal 5% of the critical damping for the ground motion measure, known as GMRotI50, introduced by Boore et al. (2006), which corresponds to the orientation independent geometrical mean of the two horizontal components. For each type of response spectra the Usable Data Bandwidth (UDB) is provided to ensure further use of credible spectral acceleration values within these limits.

Analytical documentation of the processing scheme followed during database development can be found in Segou et al. (2008).

3.1. The influence of the processing scheme on spectral ordinates

The remaining question is the influence of the decisions described in the last paragraph on the derived spectral ordinates, further used in implementing ground motion prediction equations. In order to demonstrate the differences caused by following another processing scheme we compared intermediate products and spectral ordinates of the suggested processing scheme and the one followed by the European Strong Motion Database. This comparison included almost 250 strong motion records and it has been extended to peak ground acceleration and velocity as well as pseudo-spectral acceleration and velocity for damping values equal 0% and 5% of the critical damping.
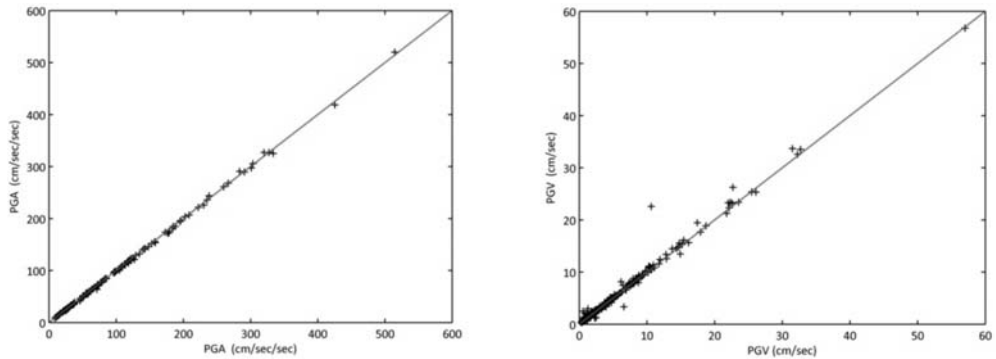
Fig. 1a-b: Influence of processing schemes on the estimation of spectral parameters. Diagram showing peak ground acceleration and velocity derived from the current processing scheme versus the ones from the European Strong Motion Database. It is also noted that the comparison is carried out for the maximum horizontal component.

Figure 1a and 1b correspond to peak ground parameter comparison showing that peak ground velocity is more influenced by the selected processing scheme. The above remark has been in fact anticipated since it is the peak ground velocity's characteristic to emphasize medium to long periods.

As far as spectral parameters (fig. 2-3) are concerned the influence observed is directly linked to the period estimator under consideration. To illustrate the differences we collate response spectra calculated from a digital strong motion record for pseudo-spectral acceleration and velocity. In figure 2a undamped spectral ordinates of ESMD appear having higher values at short periods (up to 1 sec) which can be related with the analytical solution of Chopra (1995) followed by the agency. Although, both analytical solutions, the one of Nigam and Jennings (1969) and Chopra (1995), derive credible response spectra the authors found that the latter solution is often linked with the presence of some irregular peaks in lower damping values. Once again the differences due to the vicinity of filter's cut off frequencies are evident at long periods, especially at the pseudo-spectral velocity response spectra. The diagrams in figure 3 of period estimators at 0.1 sec and 2 sec point out the benefits of using a subtle processing scheme. The strong non linear behaviour of frequency response of elliptical filters near the pass band is responsible for the distortion of the spectral ordinates near the cut off frequencies.

4. The mathematical model of the ground motion prediction equation

The scope of this paragraph is to point out the differences of the mathematical models used in the first stage of the development of a new set of ground motion prediction equations. However, it should be noted that the presentation of the derived equations with the necessary analytical description of the suggested method is beyond the scope of this paper.

The basic mathematical model of attenuation relations can be described by the following form given by Campbell (2003)

$$\ln Y = a + b\,M - c\,\ln D - d\,R + e\,F + f\,S + \varepsilon \tag{1}$$

In the equation above the logarithm of the ground motion is described as a function of magnitude M, R equals the source-to-site distance, D a diversified distance metric whereas F and S correspond to dummy variables for modelling fault type and soil site conditions. In other words in any mathe-
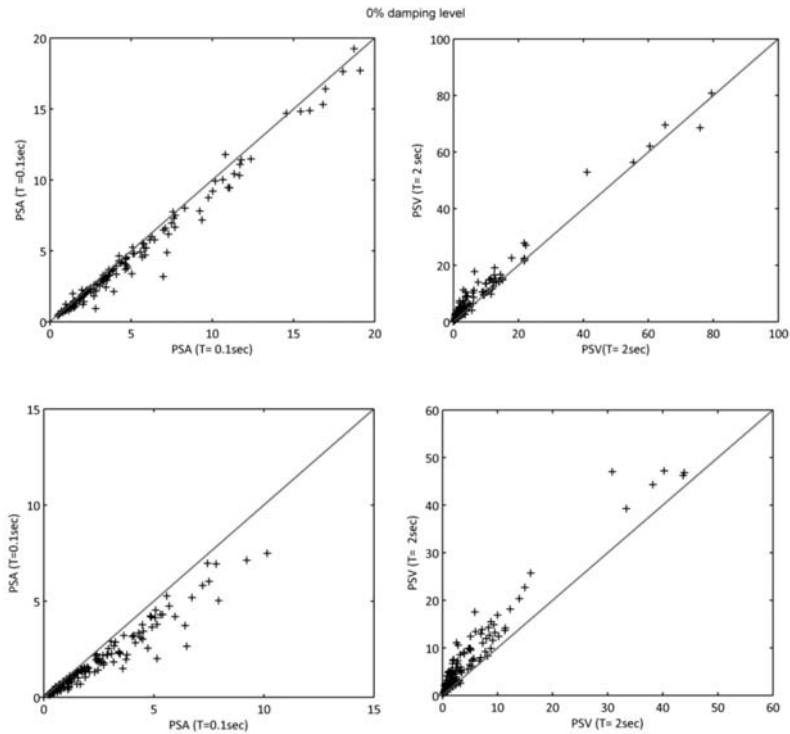
Ψηφιακή Βιβλιοθήκη Θεόφραστος - Τμήμα Γεωλογίας. Α.Π.Θ.

Fig. 2a-b: Influence of processing schemes on the spectral parameters. Comparison of pseudo-spectral acceleration and velocity for 0% and 5% damping, of the current processing scheme and the ESMD.

matical model we can distinguish independent variables, represented by the physical observations, predictors expressing the main modelling parameters, corresponding to magnitude and distance, and the remaining parameters implicitly represent random variables of the model.

However, depending on the available data and recent trends in engineering seismology the described model above can be significantly altered. In this paper we use two types of mathematical models in order to point out the influence of latter.

The first mathematical model (M1) is given by the equation (2)

$$\log Y = a + b\,M + c\,M2 - d\,\log[\,R2 + (d - h)2]\,1/2 + e1\,N + e2\,SS + e3\,RS + f1\,STS + f2\,SFS + \varepsilon\sigma1$$

The main difference with the basic equation (1) is the inclusion of a squared magnitude term and the representation of distance metric in terms of epicentral distance R and focal depth d. The consideration of epicentral distance and depth is important to adequately model the source-to-site distance since we cannot determine a common depth for seismic events due to different geotectonic environments in the broader Greek area. The square magnitude term prevents saturation of ground motion for close distances and large magnitudes. In terms of faulting type, represented with e1…3 dummy variables we found normal and strike slip ground motions very close whereas reverse faulting tends to exhibit stronger ground motions. As far as site conditions, stiff and soft soil sites correspond to amplified, by a constant factor, ground motion.
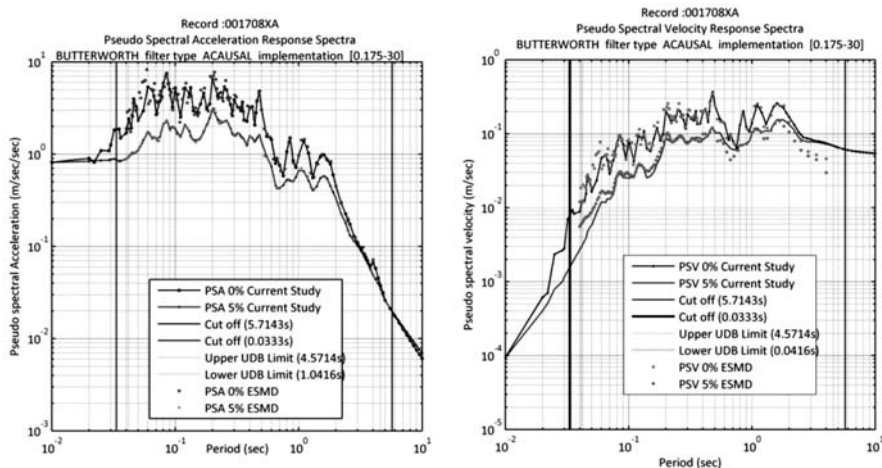
Fig. 3a-b: Influence of processing schemes on the estimation of spectral parameters. Diagrams showing spectral ground motion parameters of the current processing scheme versus the ones from European Strong Motion Database, the pseudo-spectral acceleration and velocity for 0% and 5% damping level and for 0.1 , 1 and 2s period of interest.

The second mathematical model (M2) is given by the equation (3)

$$\log Y = a + b\,M + c\,M2 - d \log[\,R2 + (d - h)2]\,1/2 + e1\,N + e2\,SS + e3\,RS + f\,(M,R,Vs30) + \varepsilon\sigma2$$

The difference between the two mathematical models concerns the term introduced for site conditions. In equation (2) the terms f1 STS, f2 SFS representing a linear amplification model had been replaced by a function f (M,R,Vs30) predicting linear and non-linear amplification factors. To calculate linear and non-linear amplification terms we followed the methodology introduced by Boore and Atkinson (2008) as part of the Next Generation Attenuation project of Pacific Earthquake Engineering Research Institute. Spectral amplification coefficients proposed in Boore and Atkinson (2008) were considered to be an improvement of the ones suggested in Choi and Stewart (2005) and they have been further used in this research study.

It should be noted that there is a clear distinction between the last two equations in terms of modeling soil site conditions. In equation (2) soil conditions are represented through NEHRP site class categories and soil site amplification factors are provided for stiff and soft soil sites corresponding to site class C and D. It is noted that strong motion data of site class A and B were considered as corresponding to rock site condition. Equation (3) allows for continuous values of shear wave velocity measured at the upper 30 m (Vs30) but in any case we should point out the broad range of Vs30 values in each NEHRP site class category.

4.1. The influence of the mathematical model

To determine the influence of the mathematical model we conducted pure error analysis as originally suggested in Douglas and Smit (2003) in magnitude-distance bins separately for each soil site class and faulting type. The length of each bin has been determined equal with 0.2 units of magnitude and 5 km distance. In case the number of earthquakes exceeded three, the error has been determined as the average absolute residuals. No trends in the residuals have been revealed through this practice but this is due to the scarcity of strong motion data inside the bins.
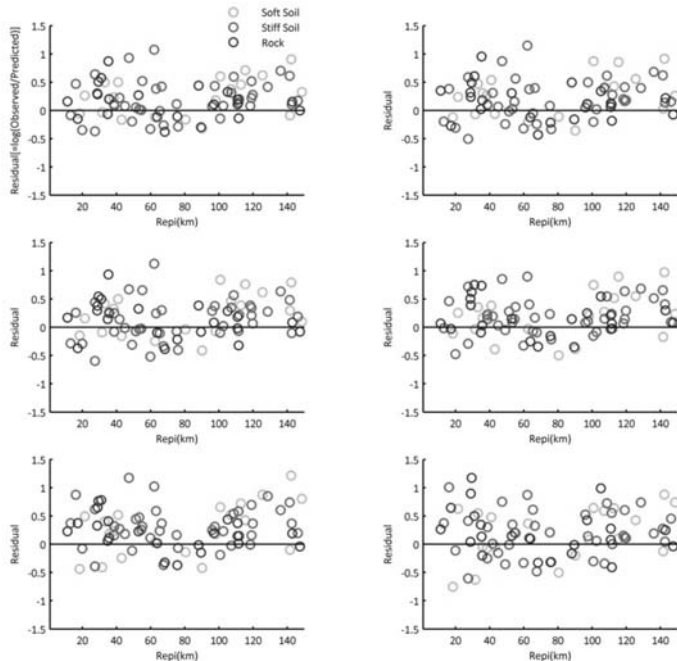
Fig. 4: Residual analysis of the approach considering linear terms for soil site conditions (eq.2). Diagram of residuals versus distance for different soil classes for discrete period estimators.

Additionally, typical residual analysis has been implemented for predicted ground motion estimated from the two mathematical models for spectral acceleration for various period estimators as well as peak ground acceleration. In figures 4 and 5 the residuals are calculated as the difference of the logarithm of the estimated from the logarithm of the predicted ground motion. From the comparison of figures 4 and 5 the suggested mathematical model corresponds to equation (2), since the residuals of the mathematical model proposing a linear term for soil behaviour yields better distributed residuals (fig. 4) than mathematical model of equation (3), incorporating a non linear term for modelling soil behaviour. The reasons behind the difference concerning the distribution of the residuals can be found in site classification and on the quality of the metadata information. Analytically, since there is no measurement in Greece for the Vs30 value of reference site conditions, we considered it to be equal with 760 m/s by following standards of NGA project (Darragh and Power, 2005); nevertheless this is acknowledged to be a possible source of error.

Of course in close relation with the aforementioned observation is the fact that there is a broad range of Vs30 values in each NEHRP site class category but in residual analysis we arbitrary decided that rock, stiff and soft soil conditions correspond to 760 m/s, 520 m/s and 250 m/s respectively.

5. Selecting the appropriate mathematical solver

Based on the maximum likelihood criterion statistical analysis provides a systematic procedure known as regression analysis for the empirical determination of the mathematical model between observed values and model predictors. However in the case under study the mathematical model is expressed through a non-linear equation imposing limitations concerning its solution. Since linearization of equation (2) through a set of linear equations would not adequately represent the
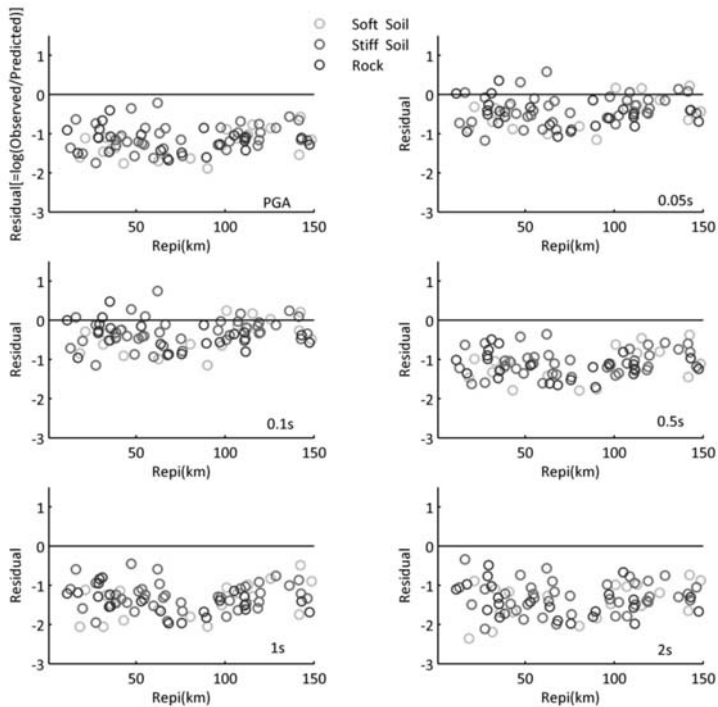
Fig. 5: Residual analysis of the approach considering non linear terms for soil site conditions (eq.3). Diagram of residuals versus distance for different soil classes for discrete period estimators.

problem in hand we had to consider steepest decent and Levenberg - Marquandt as possible solvers for determining the coefficients of equation (2) by regression analysis. The results of combining the aforementioned solvers are given in figure 6. The problem is nearly obvious, not very good magnitude scaling and there is no way to control the coefficients in order to be consistent with their physical meaning. Another disadvantage has been the necessity for providing initial conditions for the model's coefficients in order for regression analysis to begin. This actually implies that the final results are strongly driven by our initial guess.

Another possibility for solving the mathematical equation is related with optimization techniques. Whereas regression analysis provides a unique fit for the model, optimization leads to the determination of the best solution among other good solutions for the given problem. The "problem" corresponds to an objective function (model) describing the minimization of the residuals between predicted and observed ground motion. Following the minimization path there are many possibilities for selecting the appropriate mathematical solver; deterministic algorithms using least squares rule or stochastic ones, like Genetic Algorithms (GA), Simulated Annealing (SA) and Pattern Search (PS), known in programming as direct search methods.

In first place the theoretical aspects of these algorithms have been taken into consideration. Afterwards a crash-test for the efficiency of these algorithms has been designed in order to select the appropriate mathematical solver for equation (2). The aforementioned test involved comparison of the ground motion prediction equations, presented in figure 7, derived with constrained minimization through (1) non-linear least squares (LSQ), (2) non-linear least squares augmented by initial popu-
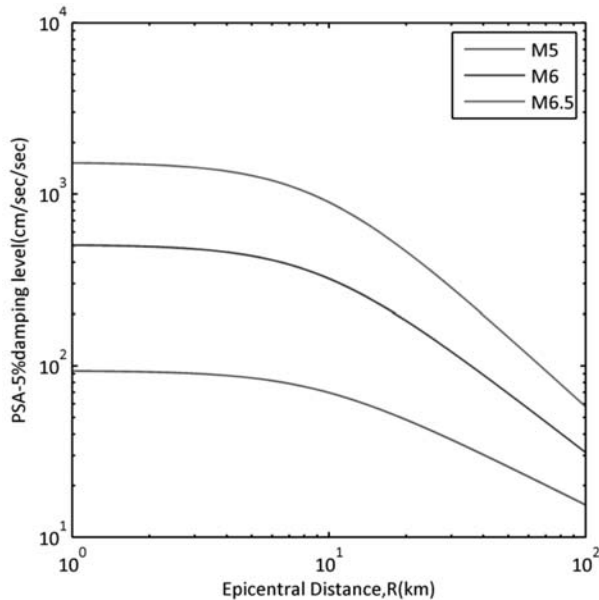
Fig. 6: Coefficient determination through regression analysis. Traditional regression analysis in least squares sense supports the implementation of mixed effects method to determine the coefficients of the mathematical model described by equation 2.

lation development (LSQP), (3) Genetic Algorithm (SGA), (4) Genetic Algorithm augmented by initial population development (GADP), (5) Hybrid Genetic Algorithm (HGA), corresponding to Genetic Algorithm solution refined with deterministic algorithm implementation for local search, (6) deterministic algorithm used in the solver previously suggested (CONMIN), (7) Pattern Search algorithm (PS) and (8) Simulated Annealing (SA).

5.1. The influence of the mathematical solver

In figure 7 the ground motion predictions equations are presented, following the mathematical model of eq. (2) for which the coefficients have been determined each time with a different mathematical solver. The algorithms converge to almost the same solution when a sufficient number of data is present, in other words different algorithms handle data scarcity in different ways. An advantage of using Genetic Algorithms supported by Diverse Population development (GADP) is that magnitude scaling is captured nicely, by taking into consideration even few observations corresponding to close distance range. The ability of optimization techniques to conform to the whole set of the observations and not only to carefully selected convenient observations is considered to be a major strong point.

Even though developed 50 years ago, it was only recently that Genetic Algorithms had been used in solving seismological problems. The stochastic nature expressing the ability to explore natural systems characterized for their complexity constituted their main advantage.

6. Conclusions-Results

In this paper we have focused on the multi-parametric nature of the construction of ground motion prediction equations. The scope had been to determine the possible sources introducing errors in
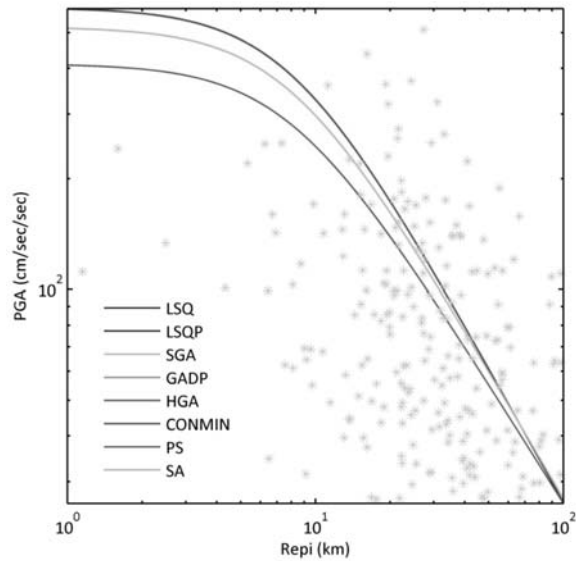
Fig. 7: Coefficient determination through optimization. Under LSQ and LSQP, non linear least square optimization without and with development of initial population, under SGA, GADP and HGA the simple genetic algorithm, the genetic algorithm with development of initial population with Latin Hypercube, and the hybrid genetic algorithm with combined use of deterministic elements for local search, CONMIN the aforementioned algorithm of deterministic nature whereas under PS and SA is noted the pattern search and simulated annealing algorithm. Points in the diagram represent data corresponding only the rock site conditions and normal style-of-faulting, used in the first stage of the determination of coefficients of the mathematical model in equation 2.

the procedure and suggest a course of action in order to derive spectral ground motion prediction equations.

The processing scheme has an important role in deriving credible spectral ordinates and its development should be based on protecting the actual frequency content while removing in the more subtle way the indentified noise.

The distribution of data in terms of magnitude and distance is the controlling factor of the construction of ground motion prediction equations. A dense distribution concerning the type of faulting and soil conditions can easily foreshadow the trends of strong ground motion whereas a poor one may reflect in some cases non typical behaviour especially for spectral acceleration of specific period estimator. The quality of the distribution is closely related with the following selection of the mathematical model.

The mathematical model includes predictors, independent and random variables. In this paper we examined the efficiency behind distance metrics and the modelling term of soil site conditions. Linear terms predicting constant site amplification for stiff and soft soil sites and a combination of linear and non-linear amplification factors has been introduced forming two different mathematical models. Quite explicitly linear modelling of soil conditions provides better ground motion prediction which is also observed when studying the residuals.

Depending on the mathematical model the last important selection is related with the mathematical solver responsible to determine the coefficients of the mathematical model. Regression analysis versus optimization techniques combined with a number of different algorithms such as Genetic Algo-

rithms have been compared. The results of the minimization problem derived from genetic algorithm with stochastic elements such as mutation, survival and crossover augmented by initial population development with Latin Hypercube design have proven to be the most credible.

The systematic analysis of the parameters that influence the construction of strong ground prediction equations pointed out usual pitfalls of the procedure.

7. Acknowledgements

8. References

Boore, D. M. 2008. TSPP—-A Collection of FORTRAN Programs for Processing and Manipulating Time Series. *U.S. Geological Survey* Open-File Report 2008-1111 (Revision 1.6).

Boore, D. M., and Atkinson, G. M. 2008. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra*, 24, 99-138.

Boore, D. M., and Bommer, J. J. 2005. Processing of strong-motion accelerograms: needs,options and consequences. *Soil Dynamics and Earthquake Engineering*, 25, 93-115.

Chopra, A. K. 1995. Dynamics of Structures. Prentice Hall, Englewood Cliffs, NJ.

Converse, A. M., and Brady, A. G. 1992. BAP: Basic Strong-Motion Accelerogram Processing Software; Version 1.0. 92-296A, United States Department of the Interior, U.S. Geological Survey.

Darragh, R., and Power, M. 2005. NGA Documentation.

http:// peer.berkeley.edu/nga/NGA_Documentation.xls, [accessed 11 March 2010].

Douglas, J., and Smit, P. M. 2001. How Accurate Can Strong Ground Motion Attenuation Relations Be? *Bulletin of Seismological Society of America*, 91, 1917-1923.

Douglas J. 2003. What is a poor quality strong-motion record?, *Bull. Earthq. Eng.*, 1,141-156.

Nigam, N. C., and Jennings, P. C. 1969. Calculation of response spectra from strong-motion earthquake records. *Bulletin of Seismological Society of America*, 59, 909-922.

Segou, M., Voulgaris, N., Makropoulos, K. C., and Stavrakakis, G. N. 2008. A review of the Greek strong motion database: needs, improvements and future development. In Proceedings of the 31st general *Assembly of the European Seismological Commission ESC2008*, Hersonissos, Crete, Greece, 7-12 September 2008, 422-427.

Segou, M. and Voulgaris, N. 2010. PROSCHEMA: a Matlab application for processing strong motion records and estimating earthquake engineering parameters, accepted for publication, Computers & Geosciences.