MULTIVARIATE BAYESIAN REGRESSION ANALYSIS APPLIED TO PSEUDO-ACCELERATION ATENUATTION RELATIONSHIPS

D. Arroyo[1] and M. Ordaz[2]

[1] *Associate Professor, Dpto. De Materiales, Universidad Autónoma Metropolitana, México DF, México*
[2] *Professor, Instituto de Ingeniería, Universidad Nacional Autónoma de México, México DF, México*
Email: aresda@correo.azc.uam.mx , MOrdaz@iingen.unam.mx

ABSTRACT :

The application of a linear multivariate Bayesian regression model to compute pseudoacceleration (*SA)* attenuation relationships is presented. The model is able to include the correlation between observations for a given earthquake, the correlation between *SA* ordinates of different periods and the correlation between coefficients of the regression model. Through comparisons between results obtained with the least-squares method and the one-stage maximum-likelihood method, the advantages of the Bayesian model are discussed.

KEYWORDS: Bayesian analysis, attenuation relations, ground motions

1. INTRODUCTION

In the past, empirical attenuation relationships of *SA* were fit to available data trough the least squares method. Several authors observed that in some cases the attenuation of *SA* with distance could not be correctly determined with this method since it disregarded the correlation between observations recorded at different sites for a given earthquake (Campbell 1981 and Joyner and Boore). Indeed, the two-stage regression method and the one-stage maximum likelihood methods were developed to solve this problem (Joyner and Boore 1993).

Also, attenuation relationships have been derived through univariate Bayesian analysis (Veneziano and Heidari 1985, Ordaz *et al.* 1994 and Reyes 1999). Ordaz *et al.* (1994) discussed the advantages of the Bayesian analysis with respect to the least squares method. The results that we present in this paper can be considered an extension of the original work of Ordaz *et al.* (1994), since we also used a Bayesian approach. Nevertheless, our model is more general and is able to include the correlation between observations recorded at different sites for a given earthquake, the correlation between *SA* ordinates at different periods, and the correlation between regression coefficients.

In this paper we briefly compare the results obtained with results obtained with the least-squares and the one-stage maximum likelihood methods. For the comparisons we use as benchmark a set of synthetic *SA* spectra with predefined statistical parameters.

In order to fulfill the length requirements of the conference we have only included some basic findings of several results that we have obtained with the Bayesian model presented. Other issues such as convergence of the Gibbs sampling method, a sound discussion on how the prior information should be defined when working with actual ground motions, differences between multivariate and univariate analysis and their possible implications on the estimation of seismic parameters through attenuation relationships will be presented in two manuscripts which are in preparation for a specialized journal.

2. THE REGRESSION MODEL

For a given *T*, the standard shape shown in equation (1) was adopted as attenuation model for regression analysis.

$$y(T) = \alpha_1(T) + \alpha_2(T)(M_w - 6) + \alpha_3(T)(M_w - 6)^2 + \alpha_4(T)\ln(R) + \alpha_5(T)R \qquad (2.1)$$

Where *T* stands for structural period, *y* is the natural logarithm of *SA*(*T*), M_w is the moment magnitude, *R* is

some measure of the distance to the fault area and α_i (T) are the coefficients determined by regression analysis. Although in this paper we have used the function shown in equation (2.1) as attenuation model, the procedure presented can be readily applied to other linear forms of attenuation relationships.

The multivariate regression model considered is $Y = X\,\alpha^T + E$, where symbol ^{T} stands for transpose, $\boldsymbol{Y}$ is a known n_o x n_T matrix which includes n_o observations of $y(T)$ for the n_T periods considered, $\boldsymbol{X}$ is a known n_o x n_p matrix which is composed by the n_o observations of the n_p parameters considered in the model (note that according to equation (2.1) the elements of the first column of the matrix $\boldsymbol{X}$ are equal to unity), $\boldsymbol{\alpha}$ is a unknown n_T x n_p matrix which comprises the coefficients determined by regression analysis (each row of $\boldsymbol{\alpha}$ contains the $\alpha_i(T)$ coefficients for a given T) and $\boldsymbol{E}$ is a unknown n_o x n_T matrix which is comprised by the regression residuals.

It is assumed that the elements $\boldsymbol{E}$ are correlated, normally distributed random variables with zero mean. The correlation between elements of $\boldsymbol{E}$ is defined through an unknown $n_o\,n_T$ x $n_o\,n_T$ matrix $\Omega = \Phi \otimes \Sigma$, where $\boldsymbol{\Phi}$ is an unknown n_o x n_o matrix which accounts for the correlation between the rows of $\boldsymbol{Y}$, $\boldsymbol{\Sigma}$ is an unknown n_T x n_T matrix which accounts for the correlation between spectral ordinates, and the symbol $\otimes$ stands for Kronecker product.

3. THE ONE-STAGE MAXIMUM LIKELIHOOD METHOD

For the model described in last section, the likelihood of $\boldsymbol{Y}$ is defined in equation (3.1), where Tr denotes trace and the symbol $\propto$ stands for proportionality, since we have omitted the normalization constant.

$$L\left(Y|\alpha,\Sigma,\Phi,X\right) \propto |\Sigma|^{-n_0/2}\,|\Phi|^{-n_T/2}\,\exp\left\{-\frac{1}{2}Tr\left[\Phi^{-1}\left(Y-X\alpha^T\right)\Sigma^{-1}\left(Y-X\alpha^T\right)^T\right]\right\} \qquad (3.1)$$

Following Joyner and Boore (1993) we considered that the elements of $\boldsymbol{E}$, say ε_{ij}, can be expressed as the sum of earthquake-to-earthquake variability (ε_e) and record-to-record variability (ε_r). In addition, the following considerations were made:

1) For a given earthquake and a given site, the coefficient of correlation between residuals for two different periods, say T_1 and T_2, is equal to ρ_{T_1,T_2}
2) For a given earthquake the coefficient of correlation between residuals for the same period at different sites is equal to γ_e.
3) For a given earthquake the coefficient of correlation between residuals for two different periods, say T_1 and T_2, at different sites is equal to $\gamma_e\,\rho_{T_1,T_2}$.
4) Residuals related to different earthquakes are independent.

According to these assumptions Φ is a block diagonal matrix defined in (3.2) where n_e is the number of earthquakes considered, and the squared submatrix ϕ_i related to earthquake i is defined in (3.3). The size of ϕ_i is the number of records of the earthquake i. Note that γ_e is equal to parameter γ in Joyner and Boore (1993) while ρ_{T_1,T_2} is the coefficient of correlation between spectral ordinates $SA(T)$ for a given pair of periods, namely T_1 and T_2. In summary, the same structure of matrix Φ used by Joyner and Boore (1993) for the univariate case was used in the multivariate case.

$$\Phi = \begin{bmatrix} \phi_1 & 0 & \cdots & 0 \\ 0 & \phi_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{n_e} \end{bmatrix} \qquad (3.2)$$

$$\phi_i = \begin{bmatrix} 1 & \gamma_e & \cdots & \gamma_e \\ \gamma_e & 1 & \cdots & \gamma_e \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_e & \gamma_e & \cdots & 1 \end{bmatrix} \qquad (3.3)$$

For a given γ_e the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ which maximize the likelihood are the well known weighted least squares estimators defined by equations (3.4) and (3.5).

$$\hat{\alpha} = \left(X^T \Phi^{-1} X\right)^{-1} X^T \Phi^{-1} Y \qquad (3.4)$$

$$\hat{\Sigma} = \frac{\left(Y - X\,\hat{\alpha}^T\right)^T \Phi^{-1} \left(Y - X\,\hat{\alpha}^T\right)}{n_0} \qquad (3.5)$$

In the maximum likelihood method, the value of γ_e which maximizes the likelihood is found iteratively and the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ for the regression analysis are the values related to the γ_e of maximum likelihood.

4. THE BAYESIAN MODEL

In the Bayesian approach $\boldsymbol{\alpha}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are regarded as matrix random variables with known joint prior density $p(\boldsymbol{\alpha}, \boldsymbol{\Sigma}, \boldsymbol{\Phi})$. This prior density is updated through Bayes' theorem and hence the posterior density is given by the product between the likelihood and the prior density.

In standard Bayesian analysis, three types of $p(\boldsymbol{\alpha},\boldsymbol{\Sigma}, \boldsymbol{\Phi})$ are commonly used: vague densities, conjugate densities and generalized conjugate densities. Vague densities are used when prior knowledge about parameters is diffuse, while conjugate and generalized conjugate densities are used when prior information about parameters is available. A more detailed description of each family of probability density functions and their implications in the regression analysis can be found elsewhere (see for example, Broemeling 1985, Rowe 2002).

In this paper we adopted a generalized conjugate probability density function as basic density. However, in order to keep the structure of $\boldsymbol{\Phi}$ shown in equation (3.2) we used a scalar beta density for γ_e, thus the prior density of $\boldsymbol{\Phi}$ is not of standard form. These densities were chosen since we found that they were suitable to properly include our prior knowledge of the parameters in the regression analysis.

The prior joint probability density used is given by $p(\alpha_V, \Sigma, \Phi) = p(\alpha_V)\,p(\Sigma)\,p(\Phi)$, where $\boldsymbol{\alpha_V}$= vec($\boldsymbol{\alpha}$).

Following Rowe (2002), we assume that the prior density of $\boldsymbol{\alpha_V}$ is the Normal density defined in equation (4.1) with mean $\boldsymbol{\alpha_{V0}}$ and covariance matrix $\boldsymbol{\Delta}$. Thus, $\boldsymbol{\alpha_{V0}}$=vec($\boldsymbol{\alpha_0}$) where $\boldsymbol{\alpha_0}$ is the prior expected value of $\boldsymbol{\alpha}$ and the positive $n_T\,n_P$ x $n_T\,n_P$ matrix $\boldsymbol{\Delta}$ is the prior covariance matrix of $\boldsymbol{\alpha_{V0}}$.

$$p(\alpha_V) \propto |\Delta|^{-1/2} \exp\left\{-\frac{1}{2}(\alpha_V - \alpha_{V0})^T \Delta^{-1} (\alpha_V - \alpha_{V0})\right\} \qquad (4.1)$$

For $\boldsymbol{\Sigma}$ we used as prior density the inverted Wishart density (Rowe 2002) shown in equation (4.2) with parameters ν and $\boldsymbol{Q}$.

$$p(\Sigma) \propto |\Sigma|^{-\nu/2} \exp\left\{-\frac{1}{2} Tr\left[\Sigma^{-1} Q\right]\right\} \qquad (4.2)$$

The positive n_T x n_T matrix $\boldsymbol{Q}$ can be computed from the prior expected value of $\boldsymbol{\Sigma}$ as it is shown in equation (4.3):

$$Q = (\nu - 2n_T - 2)\Sigma_0 \qquad (4.3)$$

where $\boldsymbol{\Sigma_0}$ is the prior expected value of $\boldsymbol{\Sigma}$. Note that $\boldsymbol{Q}$ also depends on the scalar ν, which is a measure of our degree of certainty on Σ_0. In order to give a finite value to the variance of the elements of $\boldsymbol{\Sigma}$ the value of ν should be greater than $(2n_T+4)$; the larger the value of ν, the greater the degree of certainty on Σ_0.

In Bayesian analysis usually an inverted Wishart density is also used for Φ (Rowe 2002). Nevertheless, if it is desired that $\boldsymbol{\Phi}$ has the structure shown in equation (3.2) the inverted Wishart density cannot be used. We note that, for a given set of data, $\boldsymbol{\Phi}$ is only a function of γ_e, hence we decided to use a scalar beta density for γ_e according to equation (4.4).

$$p(\gamma_e) \propto \gamma_e^{a-1}(1-\gamma_e)^{b-1} \qquad (4.4)$$

where parameters a and b can be computed from the prior expected value of γ_e and its prior standard deviation.

In summary, the prior information about regression parameters is included in the analysis through $\boldsymbol{\alpha_{V0}}$, $\boldsymbol{\Delta}$, $\boldsymbol{Q}$, ν, a and b, which are known as hyperparameters, and equations (4.1) to (4.4).

Multiplying the prior joint density function by the likelihood function we obtain the posterior joint density of the regression parameters:

$$p(\alpha, \Sigma, \gamma_e | X, Y) \propto |\Sigma|^{-(n_o+\nu)/2} |\Phi|^{-n_T/2} \gamma_e^{a-1}(1-\gamma_e)^{b-1} \exp\left\{-\frac{1}{2}(\alpha_V - \alpha_{V0})^T \Delta^{-1}(\alpha_V - \alpha_{V0})\right\}$$
$$\exp\left\{-\frac{1}{2} Tr\left[\Sigma^{-1}\left((Y - X\alpha^T)^T \Phi^{-1}(Y - X\alpha^T) + Q\right)\right]\right\} \qquad (4.5)$$

This joint density should be marginalized in order to obtain the posterior marginal expected values of $\boldsymbol{\alpha}$, $\boldsymbol{\Sigma}$ and γ_e. However, for this density, it is not possible to obtain marginal distributions in an analytical closed form. Nevertheless, marginal expected values can be numerically computed through the stochastic integration method known as Gibbs sampling.

5. SYNTHETIC DATA

In order to assess the performance of the least-squares method, the one-stage maximum likelihood method and the Bayesian method we generated different sets of synthetic $SA(T)$ spectra with predefined statistical properties. We considered 25 structural periods (T) ranging between 0 and 5.0 seconds.

We generated 6 sets of synthetic spectra assuming the number of earthquakes shown in table 1. It was considered that the moment magnitude (M_w) of the events follows a modified Gutenberg-Richter distribution, with minimum value of M_w=6, maximum value of M_w=8.2 and β=2.0

Also, we considered that each earthquake was recorded at the number of sites shown in table 1. We assume that R followed a uniform distribution between 250 and 400 Km. Hence, the number of records for a given set is the product between the number of earthquakes and the number of sites.

Table 1 Synthetic sets of *SA(T)* spectra used

Set	Earthquakes	Sites	n_o	Set	Earthquakes	Sites	n_o
1	3	4	12	4	20	10	200
2	5	10	50	5	50	10	500
3	10	10	100	6	100	10	1000

The predefined statistical properties of the set of ground motions were $\boldsymbol{\alpha}_p$, $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Phi}_p$. We set $\boldsymbol{\alpha}_p$ as the value of the attenuation model proposed by Reyes (1999) for the CU station of Mexico City.

The diagonal terms of $\boldsymbol{\Sigma}_p$ were set as the variance of the residuals (σ^2) related to the Reyes (1999) model, while the off-diagonal terms were computed through the equation proposed by Baker and Cornell (2006) to estimate the coefficient of correlation ρ_{T_1,T_2}. For $\boldsymbol{\Phi}_p$ we used the structure shown in equation (3.2) with γ_e=0.2234 which is a value that we infer from the results presented by Joyner and Boore (1993).

Thus, given the number of earthquakes and the number of records shown in table 1, a n_ox n_T matrix random variate from the matrix normal distribution was generated. The mean value of the distribution was computed with $\boldsymbol{\alpha}_p$ with a matrix covariance defined by $\Phi_p \otimes \Sigma_p$. Note that regardless of the set considered, $\boldsymbol{\alpha}_p$, $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Phi}_p$ represent the statistical properties of the entire population of *SA(T)* spectra; hence these parameters were used as benchmarks for the regression analysis presented in the following sections.

6. RESULTS FOR ONE-STAGE MAXIMUM LIKELIHOOD METHOD

Figure 1 shows a comparison between regression parameters obtained with the one-stage maximum-likelihood method and the benchmarks. Although not shown, the results were very similar to those observed for the least squares method. In general, the one-stage maximum likelihood method is not able to attain the benchmark values, except for $\alpha_2(T)$ and σ.

7. RESULTS FOR THE BAYESIAN METHOD

7.1 Prior Information

The elements of matrix $\boldsymbol{\alpha}_0$ were set as follows. With the amplitude Fourier spectra defined by the Brune's model and using random vibration theory, we obtained *SA* spectra related to several values of M_w and R (McGuire and Hanks 1980). Then, we applied the least squares method to compute the value of $\boldsymbol{\alpha}_0$. This implies that *a priori* we believe that the attenuation of *SA* spectra could be properly characterized by the Brune's model.

Since the terms α_1 and α_5 in equation (2.1) depend on site effects (Ordaz *et al.* 1994) we assigned to their variance a large value with respect to their prior expected value. This implies that $\alpha_1(T)$ and $\alpha_5(T)$ could attain any value which yields the best fit in the regression analysis. In the case of α_2, α_3 and α_4 we assigned to their covariance a value that implies a coefficient of variation of 1.7, as it was used in a previous study (Ordaz *et al.* 1994).

We set the diagonal elements of Σ_0 equal to 0.49, which means that *a priori* we believe that the standard deviation of the residuals is equal to 0.7 independently from T. In addition, the off-diagonal terms were defined through the coefficient of correlation shown in equation (7.1):

$$\rho^0_{T_1,T_2} = \exp\left(-q\left|T_1 - T_2\right|\right) \qquad (7.1)$$

According to equation (7.1) the coefficient correlation varies from unity when periods are equal to nearly 0.05 when the difference between periods is about 3 seconds and for q=1.0. It must be acknowledged that equation (7.1) is quite arbitrary. This correlation structure was created only for the synthetic example presented in this paper. When working with actual ground motions it could be more reasonable to use the correlation coefficients defined by Baker and Cornell (2006) as prior values. In the example presented in this paper we decided not to use the equations of Baker and Cornell (2006) as prior information since they were used to generate the synthetic data.

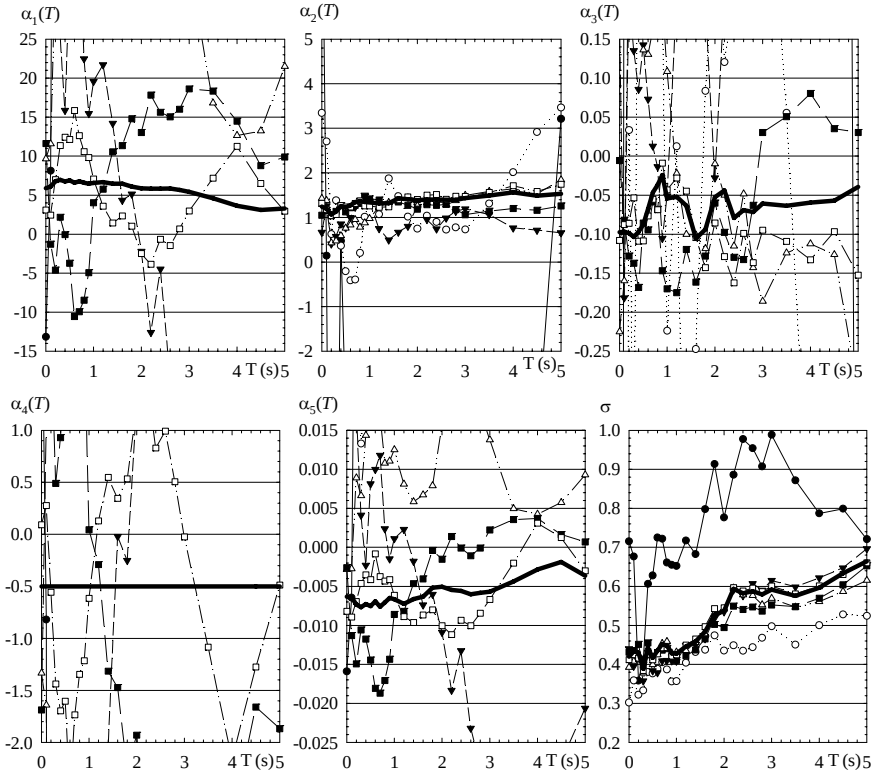

Figure 1 Results for maximum likelihood method (Symbols: black circles n_o=12, white circles n_o=50, black triangles n_o=100, white triangles n_o=200, black squares n_o=500, white squares n_o=1000 and thick line benchmark)

The degree of certainty of Σ_0 depends on parameter ν; as it has been discussed, the larger the value of ν, the larger the degree of certainty on Σ_0. The posterior value of Σ can be expressed as a weighted average between prior information and the conditional weighted least squares estimate. The weighting factors are $\nu/(n+\nu)$ for the prior information and $n/(n+\nu)$ for the conditional weighted least squares estimate. In the computations we have used a value equal to 57 (the minimum required for n_T=25 in order to give a finite value to the covariance of Σ_0). It means that we believe that Σ_0 is uncertain.

Finally, our prior information of γ_e is vague, so we have set a=b=1.5, which is a very flat density, with expected value of 0.5, in order to force γ_e to take the value which yields the best fit to data.

7.2 Results

The results obtained with the Bayesian model are presented in figure 2. Conversely to what happened with the other models, the Bayesian was able to attain benchmark values, except for parameter α_3(T). Note that with n_o between 100 and 200 very accurate estimates of the regression parameters are observed. Although not shown, small values of standard deviations are observed in the Bayesian model, which is an advantage over the other

methods, since scatters of the Bayesian regression parameters are smaller than those observed with other methods.

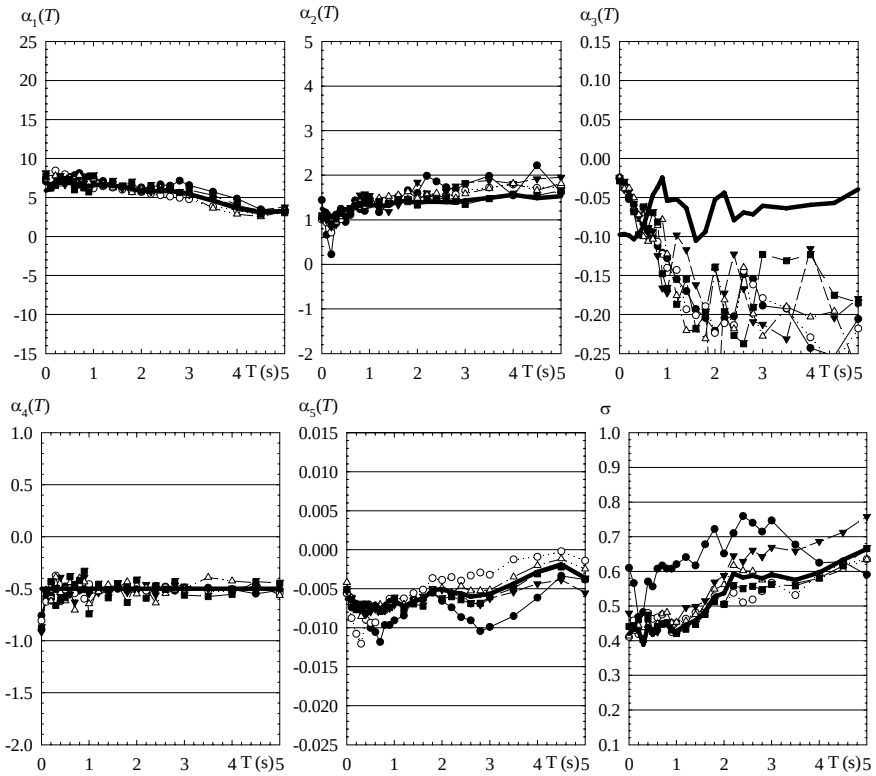

Figure 2 Results for Bayesian method (Symbols: black circles n_o=12, white circles n_o=50, black triangles n_o=100, white triangles n_o=200, black squares n_o=500, white squares n_o=1000 and thick line benchmark)

We noted that the information contained in data was not enough to completely define parameter $\alpha_3(T)$; in other words, parameter $\alpha_3(T)$ has little effect on y and almost any value could have been used. Hence, the Bayesian method, instead of leading to any value of $\alpha_3(T)$ (such as the other methods) leads to values of $\alpha_3(T)$ that are close to its prior value.

Also, it was observed that accurate estimates of γ_e were obtained for n_o greater than 50. In spite of our use of a prior value of γ_e=0.5 the data shifted the prior value to the correct value of γ_e. On the other hand, to obtain accurate estimates of γ_e values of n_o greater than 200 were required for the one-stage maximum likelihood method.

8. DISCUSSION AND CONCLUSIONS

We noted that while values of M and R lie in the ranges observed in the sample, results for the three methods yield the same level of accuracy, even when some coefficients seem theoretically unacceptable. However, we decided to compare the predicted SA spectra, obtained with coefficients associated to n_o=200 and different regression methods, with the corresponding benchmark spectra. We choose M=7 and four different values of R: 200 and 500 Km. These values are out of the range of data contained in the synthetic set; thus, this comparison can be regarded as an evaluation of the possibility of extrapolating the results from the regression analysis. The results are summarized in figure 3. As can be observed, only the Bayesian regression yields acceptable results. On the other hand, the least squares and maximum likelihood methods might lead to very inaccurate results. Note that for R=200, large differences are observed for the least-squares and maximum-likelihood methods, in spite of the fact that this distance is only 20% lower than the minimum value of R included in the synthetic set.

We have presented a linear multivariate Bayesian regression method which includes the correlation between observations for a given earthquake, the correlation between *SA* ordinates at different periods, and the correlation between coefficients of the regression model.

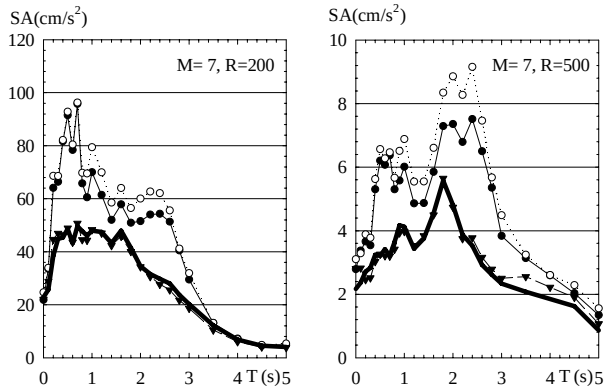

Figure 3 Comparison between *SA* spectra obtained through different methods. (Symbols: black circles least squares method, white circles maximum likelihood method, black triangles Bayesian method and thick line benchmark value)

Through comparisons between results obtained with the least squares and the maximum likelihood methods we have shown than multiple solutions close to minimum error could exist and that the Bayesian method could be used to obtain regression parameters consistent with seismological theory. In addition, for the synthetic example presented, it is shown that attenuation relationships obtained through Bayesian analysis yield more accurate results than other methods when attenuation relationships are extrapolated. However, the Bayesian method requires significantly more analytical and computational work than traditional methods.

REFERENCES

Baker, J. W. and Cornell, A. C. (2006). Correlation of response spectral values for multi-component ground motions, *Bull. Seism. Soc. Am.* **96**, 215-227

Broemeling, L. D. (1984). Bayesian analysis of linear models, Marcel Dekker Inc., New York, 454 pp.

Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra, *Bull. Seism. Soc. Am.* **73(6)**, 1865-1894.

Campbell, K. W. (1981). Near-source attenuation of peak horizontal acceleration, *Bull. Seism. Soc. Am.* **71**, 2039-2070

Joyner, W. B. and Boore, D. M. (1993). Methods for regression analysis of strong-motion data, *Bull. Seism. Soc. Am.* **83(2)**, 469-487.

Ordaz, M., Singh, S. K. and Arciniega, A. (1994). Bayesian attenuation regressions: an application to Mexico City, *Geophys. J. Int.* **117**, 335-344.

Reyes, C. (1999). El estado límite de servicio en el diseño sísmico de edificios, Ph. D. Thesis, School of Engineering, UNAM.

Rowe, D. B. (2002). Multivariate Bayesian statistics: models for source separation and signal unmixing, Chapman & Hall/CRC, New York, 329 pp.