FINAL TECHNICAL REPORT

AWARD NUMBER: 08HQGR0086

Uniform hazard spectra for site-specific applications including uncertainties in site-response

PI: Adrian Rodriguez-Marek
Report co-author: Gonzalo Montalva
Washington State University
Dept. of Civil and Environmental Engineering
Sloan Hall 101
Pullman, WA 99164-2910
509-335-7088 (phone)
509-335-7632 (fax)
adrian@wsu.edu

February 2010

Research supported by the U. S. Geological Survey (USGS), Department of Interior, under USGS award number 08HQGR0086. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. Government

1 Introduction and Objectives

1.1 Introduction

In current practice, the estimation of seismic hazard is performed within the context of probabilistic seismic hazard assessment (PSHA). PSHA results are given in terms of ground motion parameters with a certain probability of exceedance over a set time interval, or equivalently, ground motion parameters with a preset return period. For example, the International Building Code, IBC (ICC 2006) bases the design response spectra on spectral accelerations with a return period of about 2475 years which corresponds to a probability of exceedance of 2% in 50 years. A uniform hazard spectrum (UHS) is a spectrum build with spectral accelerations that have a uniform probability of exceedance for all periods. Code-based design spectra are build with the objective it represents a UHS. In practice, code-based design spectra are typically constructed from spectral accelerations obtained from USGS seismic hazard maps, which are developed using PSHA for a baseline site category. However, most design scenarios involve site conditions different than the baseline case. In these cases, site response analyses must be applied to obtain site-specific hazard estimates.

Site response can be accounted for by modifying the baseline UHS either with site response analyses or using code-prescribed period-, intensity-, and site-dependent spectral amplification factors. Neither of these approaches ensures that the resulting response spectrum is a UHS, or, more generally, that the uncertainty predicted for ground motions at a reference site is rigorously propagated in a site response analysis (Goulet and Stewart 2009). There are two contributions to ground motion variability for site-specific applications. First, there is variability (both epistemic uncertainty and aleatoric variability) that results from the variability of the reference-site ground motions. The propagation of this variability to the site-specific site conditions can not be captured by a deterministically-computed site factor, in particular because soil nonlinearity necessarily implies that site response is not the same for all ground motion scenarios that contribute to the hazard at one return period. A second source of variability results from the site response process. Part of this variability is aleatoric in nature (related to variability in site response to different input motions). In a prediction exercise, however, there is also epistemic uncertainty in site response that results from uncertainties on the model used for site-

amplification analysis (model uncertainty) and the lack of knowledge of the soil profile at a given site (measurement uncertainty).

Site response is a function of the input motion and the soil properties at a site. The measured value of a soil property is always considered an approximation to its true value due to limitations of the measuring and sampling techniques and the limited volume of material that can be sampled in a project. The engineer that analyzes a geotechnical problem often fills the gaps left by site exploration and laboratory testing with inferences based on judgment, and often must trust the results of tests that only approximately measure the properties needed for an analysis. In summary, the properties of soils and rocks should be considered to be random variables.

This report presents two alternative methodologies to compute site-specific spectra that incorporate measurement uncertainties. The first methodology is based on site-specific site response analyses; the second methodology is based on site factors quantified based on the generic site categories used by current building codes. These methodologies can have a direct impact in seismic hazard analysis. The effect of uncertainties in measured or estimated soil properties on site amplification will be properly quantified. In particular, the effect of these uncertainties on code-based site amplification factors will be studied. This will result in recommendations on site-amplification factors that can maintain the uniform hazard properties of the reference-rock spectrum.

1.2 Background

Site response may be estimated either by direct measurements or indirectly, meaning that soil properties and profile geometry are evaluated and from them an estimate of local amplification is computed. In engineering design, site effects can be accounted by site amplification factors or site-specific analyses. Direct measurements of site effects are obtained by comparing the ground motions on soil to that on a rock reference site, and have led to the calculation (or estimation) of amplification factors, which are then used by building codes (e.g., Borcherdt, 1994; Dobry et al., 2000). These amplification factors are linked to site classes which are most commonly defined by the average shear-wave velocity in the upper 30 m (Vs30). Indirect estimation of site effects enables the designer to perform site-specific analyses, which have the advantages of increased flexibility (can be used on sites different than those were recorded ground motions are available), and potential reduction of uncertainty due to detailed characterization of the subsoil.

Disadvantage of the site-specific method are the uncertainties that are inherent to the testing on which it relies upon, and to the physical model of the site response phenomena.

Bazurro and Cornell (2004a and b) established a framework to incorporate site response into PSHA in a rigorous fashion. The incorporation of local site conditions into the estimation of seismic risk is achieved by means of a frequency dependent site amplification factor (*AF*) given by

$$AF = \frac{S_a^s(f)}{S_a^r(f)} \tag{1.1}$$

where $S_a^s(f)$ is the spectral acceleration at the soil surface, $S_a^r(f)$ is the spectral acceleration at a reference site (i.e. rock), and f is the frequency. The site response problem is defined as the computation of $S_a^s(f)$ and $\sigma_{lnS_a^s}(f)$ for given $S_a^r(f)$ and $\sigma_{lnS_a^r}(f)$. This computation requires the definition of the soil profile (i.e. layering), the soil properties, and the input ground motion in terms of its spectral acceleration.

In Bazzurro and Cornell's procedure, the site amplification factor is mapped as a function of $S_a^r(f)$ for a given frequency. Assuming that linear fitting in the log-log space is statistically appropriate (see Bazurro and Cornell 2004a), *AF* is expressed as

$$lnAF(f) \approx c_0 + c_1 * \left(\overline{lnS_a^r(f)} + \varepsilon_{lnS_a^r(f)} * \sigma_{lnS_a^r(f)} \right) + \varepsilon_{lnAF(f)} * \sigma_{lnAF(f)} \tag{1.2}$$

where c_0, and c_1 are coefficients of the regression, $\overline{lnS_a^r(f)}$ is the median value of the logarithm of the spectral acceleration on rock (predicted by ground motion prediction equations, GMPEs); $\varepsilon_{lnAF}(f)$ is a standard normal variable that represents the deviation of the amplification factor from its median value; $\sigma_{lnAF}(f)$ is the standard deviation of the amplification factor; $\varepsilon_{lnS_a^r}(f)$ is a standard normal variable that represents the deviation of the log-spectral acceleration in rock from its median value; and $\sigma_{lnS_a^r}(f)$ is the standard deviation of the spectral acceleration in rock in natural log space. Bazurro and Cornell (2004b) indicate that the standard deviation of the amplification factor is not significantly dependent on the input ground motion. This statement

will be put to the test in this study. Equation 1.2 along with Equation 1.1 can be used to estimate $S_a^s(f)$ for a given $S_a^r(f)$ - $\sigma_{lnS_a^r}(f)$ pair:

$$\ln \overline{S_a^s(f)} \approx c_0 + (c_1 + 1)\ln S_a^r(f) \tag{1.3}$$

where $\ln \overline{S_a^s(f)}$ is the median estimate of the log-spectral acceleration in soil. The variability in the surface motion is obtained from (Bazurro and Cornell 2004b):

$$\sigma_{lnS_a^s(f)} \approx \sqrt{(c_1+1)^2\left(\sigma_{lnS_a^s(f)}\right)^2 + \left(\sigma_{lnAF(f)}\right)^2} \tag{1.4}.$$

Equation 1.4 assumes that the residuals of the amplification factor and the residuals of the input motion are uncorrelated. As pointed out by Bazurro and Cornell (2004b), McGuire had indicated that these parameters are mildly negatively correlated. Again, this statement will be put to the test in this study.

In Bazurro and Cornell's approach, the coefficients for the *AF* (Equation 1.2) are obtained by fitting Equation 1.2 to a series of site response analyses (Bazurro and Cornell 2004a). When site conditions are not known with sufficient detail or when site response analyses are not warranted in design, the *AF* can be estimated using Vs30. This parameter is a valid estimator for site response because of its relation with site period (Dobry et al. 1976), as shown by the early work of Borcherdt (1994). However, site factors are usually applied deterministically. This approach has been criticized by Goulet and Stewart (2009) for underestimating the surface ground motions computed probabilistically for site-specific cases.

Bazurro and Cornell indicate that both S_a^r and its variability are obtained from GMPEs. However, estimates of uncertainty from GMPEs include component of site-to-site variability that are not applicable to site-specific analysis. The basic premise of the proposed methodologies is that uncertainty has to be properly partitioned into the uncertainty that results from event-to-event variability, uncertainty that results from inter-station variability, and the intra-station variability. An estimate for this partition is obtained using the KiK-net database and the results are presented in Section 2. Section 3 presents the methodology used to propagate the uncertainties in input parameters to estimates of ground motion uncertainty. This methodology

uses a random field model to quantify the variability in shear-wave velocity profiles. The proposed methodologies for site response analyses are discussed in Section 4. Finally, the most important conclusions of this study are summarized in Section 5.

2 Site-specific ground motion variability

GMPE equations are usually given in the form of:

$$\ln(y) = \ln(\bar{y}) + \varepsilon \tag{2.1}$$

where y is a ground motion parameter, which in this study is the spectral acceleration at 5% damping; $\bar{y}$ is the median value of y predicted as a function of site and earthquake parameters; and ε is a random variable that quantifies aleatoric ground motion variability and is generally assumed to be normally distributed with standard deviation σ_{tot}. The total uncertainty can be separated into various components. Generally, it is separated into intra-event and inter-event terms (this is the case for all of the NGA equation, see Power et al. 2008). This decomposition is possible if multiple records are available for each earthquake. Conceptually, uncertainty could be decomposed into additional components. For example, Joyner and Boore (1993) proposed its decomposition into an *earthquake-to-earthquake* term, a *site-to-site* term, and an additional term that groups the remaining uncertainty:

$$\varepsilon = \varepsilon_{e\text{-}e} + \varepsilon_{s\text{-}s} + \varepsilon_{o} \tag{2.2}$$

For any given earthquake, $\varepsilon_{e\text{-}e}$ takes a given value (that is unknown in a prediction exercise) and for the ensemble of earthquakes in the database the variable becomes a random variable with a given standard deviation $\sigma_{e\text{-}e}$. Similarly, for any given site, $\varepsilon_{s\text{-}s}$ takes a given value (e.g. the "site term") and for the ensemble of sites in the database it becomes a random variable with zero mean and standard deviation $\sigma_{s\text{-}s}$. The residual variability ε_{o} is due to possible variations in the site term from earthquake to earthquake, to measurement uncertainty, or to inherent variability (e.g. variability that can not be captured by the current parameterization).

In this work, we postulate that when performing site specific analyses, the $\varepsilon_{s\text{-}s}$ term must be replaced by a deterministic site-term. However, since the computation of this site term involves

epistemic uncertainty, this uncertainty must also be included into the prediction of ground motion. Hence, for a site-specific analysis Equation 2.2 becomes:

$$\ln(y) = \ln(\bar{y}) + \eta_{site} + \varepsilon_{e\text{-}e} + \varepsilon_o \tag{2.3}$$

The site term η_{site} in Equation 2.3 can be estimated using site response analyses. Note that this site term will have both a median value ($\bar{\eta}_{site}$) and an uncertainty ($\sigma_{\eta site}$). Both must be computed using the site response analyses. The aleatoric uncertainty for a given site ($\varepsilon_{e\text{-}e} + \varepsilon_o$) can be estimated using single-station standard deviations (e.g. Atkinson 2006). The KiK-net ground motion database (http:///www.kik.bosai.go.jp; see also Kinoshita 1998; Aoi et al. 2000; Fujiwara et al. 2004) was used to obtain preliminary estimates of single-station standard deviation. These results are presented below.

2.1 Estimates of Single Station standard deviation

Estimates of single-station standard deviation were obtained using the KiK-net ground motion data base. A GMPE was first developed using both ground motion records at the surface and ground motions recorded at the bottom of boreholes. The borehole records were included to also obtain estimates of the variability in the amplification function between borehole and the surface.

KiK-net database

The KiK-net database will be used in this research. Extensive processing of the KiK-net database was done by G. Pousse and is described Pousse et al. (2005) and Cotton et al. (2008). The KiK-net network is located in Japan. Each station in this network has two 3-component accelerographs, one at the surface and another at depth (most sensors are located at a depth of -100m or -200m). The instruments have a 24 bit analog-to-digital converter with a sampling frequency of 200 Hz (Fujiwara et al. 2004). All records between 1996 and October 2004 with $M_{JMA} > 4$ have been downloaded. As a preliminary check to avoid subduction related records, only events with depth less than 25 km were analyzed. The M_{JMA} magnitude was converted to seismic moment magnitude using the Fukushima (1996) relationship (Cotton et al. 2008). Closest distance to the rupture was computed for all recordings. This rupture is assumed to correspond to the hypocentral distance for small to moderate earthquakes or when the source dimensions

remain unknown. The magnitude-distance sampling of the database is shown in Figure 2.1. A visual inspection was performed on ground motion data to check for glitches and to keep only the main event if multiple events were recorded in the time series. The signals were band-pass filtered between 0.25 and 25 Hz with a Butterworth filter. This filtering was performed in the time domain with four poles and two passes using SAC2000 routines (Goldstein et al. 2003). The longest usable period of our empirical model is less than 3 s. For a more detailed description of the data processing, please refer to Pousse et al. (2005) and Cotton et al. (2008). Shear-wave velocity profiles at each of the ground motion stations were obtained using downhole shear-wave velocity measurements.

The objective of this study is not to develop an GMPE that serves as a predictive tool in the future, but to develop a baseline relationship for the subsequent analysis of residuals in the database. For this reason, only records from stations that recorded 5 or more earthquakes were considered. The distribution of records is shown in Figure 2.1

Ground Motion Prediction Equation Model

Spectral accelerations at any given period (y) are assumed to be defined by:

$$\ln(y_{ij}) = \bar{y}_{ij} + \varepsilon_{\sigma,ij} + \varepsilon_{\tau,ij} \tag{2.4}$$

where the subscript i denotes *event i*, and the subscript j denotes *record j*. ε_σ and ε_τ are the intra-event and inter-event terms and are assumed to be normally distributed random variables with zero mean and standard deviation σ and τ, respectively. The total standard deviation is hence given by $\sigma_{tot} = (\sigma^2 + \tau^2)^{0.5}$.

The functional form selected was the one used by Boore and Atkinson (2008). The choice was made purely due to the fact that it is one of the simplest of those used in the NGA set. The functional form for the median estimate is given by:

$$\ln(\bar{y}) = \mathrm{Fm} + \mathrm{Fd} + \mathrm{F_{site}}\,(S_{surface}) + [\mathrm{F_{100}}\,(S_{100}) + \mathrm{F_{200}}(S_{100})]\,(1 - S_{surface}) \tag{2.5}$$

where

$$\mathrm{Fd} = [c_1 + c_2(\mathrm{M} - M_{ref})]\,\ln(\mathrm{R}/R_{ref}) + c_3\,(\mathrm{R} - R_{ref})$$

$$\mathrm{R} = \mathrm{sqrt}\,(\mathrm{R}^2 + h^2)$$

$$Fm = e_1 + e_5(M - M_h) + e_6 (M - M_h)^2 \qquad \text{for } M < M_h$$

$$Fm = e_1 + e_7 (M - M_h) \qquad \text{for } M > M_h$$

$$F_{site} = b_{lin} \ln(Vs30/V_{ref})$$

$$F_{100} = a_{100} + b_{100} \ln(Vs30/ V_{ref}) + c_{100} \ln(V_{s,hole} / 3000)$$

$$F_{200} = a_{200} + b_{200} \ln(Vs30/ V_{ref}) + c_{200} \ln(V_{s,hole} / 3000)$$

The predictor variables are:

M	Moment Magnitude
R	Closest distance to the fault (See Cotton et al. 2008)
Vs30	Average shear-wave velocity over the upper 30 m.
$V_{s,hole}$	Shear-wave velocity at the depth of the instrument.
$S_{surface}$	Flag = 1 for a record at the surface, 0 otherwise
S_{100}	Flag = 1 if the borehole instrument is at a depth of 150m or less, 0 otherwise
S_{20}:	Flag = 1 if the borehole instrument is at a depth larger than 150m, 0otherwise

and c_1, c_2, c_3, e_1, e_5, e_6, e_7, b_{lin}, a_{100}, b_{100}, c_{100}, a_{200}, b_{200}, c_{200}, h, and M_h are model parameters and M_{ref}, V_{ref}, and R_{ref} are the reference magnitude, average shear-wave velocity, and distance.

The use of a single predictive equation for surface and borehole records implies that both distance and magnitude dependency are the same for both sets of records. The difference between surface and borehole is captured by the site factor and the borehole factors. These factors are multiplicative factors with respect to a reference motion recorded at a surface site with Vs30 = V_{ref}, where V_{ref} was chosen to be 760 m/s.

The most important limitation of the proposed functional form is that site amplification is assumed to be linear. This appears to be the case for most ground motion stations but not for all. However, nonlinearity could not be constrained using purely empirical methodology and at this stage we chose to ignore it.

The regression methodology uses the random effects model of Abrahamson and Youngs (1981). The most important points are:

- Only surface data is used to constrain the site amplification term (parameter b_{lin})
- Inter-event standard deviations in the random effects loop are computed using all data (surface and borehole). That implies that the inter-event terms are assumed to be the same for the surface and the borehole (the same assumption was made by Abrahamson and Silva (2008) in their NGA GMPE by assuming that inter-event terms for soil and rock stations are the same).
- Intra-event standard deviations are assumed to be independent of magnitude.

The last assumption is somewhat limiting and can be reduced in further analyses. Because of this assumption, the dependency of the intra-event term on depth, Vs30, or magnitude had to be obtained from an analysis of residuals. Parameters of the GMPE for selected periods are given in Table 2.1

Analysis of residuals

The estimates of single-station standard deviation were obtained by performing an analysis of residuals on ground motion stations that recorded 15 or more earthquakes. This rendered a total of 44 stations. Standard deviations at each station were averaged to obtain an estimate of the single-station standard deviation. A more in-depth analysis of these standard deviations is currently being undertaken under USGS Award G10AP00029. The resulting estimates of single-station standard deviation are shown in Figure 2.2 and Table 2.2.

Remarkably, the single station standard deviation estimates from the KiK-net database are similar to those obtained by Atkinson (2006) using a Southern California data set. Other studies obtained values for single-station standard deviation that are remarkable similar to those obtained in this study (Chen and Tsai 2002, Abrahamson, personal comm.). For the KiK-net database, single-station standard deviations are on average 25% lower than the total standard deviation estimated from the entire database. This percentage may vary from region to region as the total standard deviation from the aforementioned studies also varies. It is also important to note that the estimates of standard deviation in this study are obtained using a large range of earthquake magnitudes. Most GMPE indicate that standard deviation is magnitude dependent, decreasing for increasing magnitude (e.g. Abrahamson and Silva 2008, Chiou and Youngs 2008, Idriss 2008).

Since the single-station standard deviation also includes the inter-event term, it is expected that its value would be lower if magnitude dependence were included.

Note that the standard deviations obtained at the borehole using the entire data set are lower than those obtained at the surface (Figure 2.2). However, this reduction in standard deviation from surface to depth is not observed when looking at single-station estimates of standard deviation (Figure 2.2, Table 2.2). Single-station standard deviation at depth are lower for short periods (up to 4% lower for T=0.3s), and higher for periods higher than 1 s (up to 7% higher). This is an indication that an important contributor to uncertainty in ground motion predictions is due to site-to-site variability and is likely due to improper characterization of the site amplification term. A corollary to the observation that surface and depth single-station sigma are similar is that single station estimates of standard deviation at the surface are in general applicable to larger depths. This is important given that most ground motion station networks do not have borehole instruments.

3 Uncertainty in site response analyses

The site term in Equation 2.3 (η_{site}) can be estimated using a site response analysis code. The GMPE already accounts for site response through a Vs30 dependent term, and η_{site} is the residual with respect to this site term. Hence, the site term would be obtained from the difference of an amplification factor obtained from site amplification analysis with respect to a reference site and the Vs30 term of the GMPE. As previously indicated, both the median value and the uncertainty of the site term must be computed. The uncertainty in the site term results from uncertainties in the input parameters (i.e. input motion variability, uncertainty on the shear-wave velocity profile, and uncertainty on the nonlinear behavior of the soils) and the model uncertainty (or bias) that results from possible errors in the selected site-amplification code. In this study, we focus on the components of uncertainty related to the input parameters. Note that while these uncertainties are formally epistemic uncertainties, it is easier to treat them within the aleatoric uncertainty terms described in Equation 2.3. Model uncertainties can be treated within a logic tree approach.

Input motion uncertainty can be quantified by the standard deviation term in a GMPE. As discussed in Section 2, for a site-specific analysis this standard deviation is best estimated using single-station standard deviations. This is paramount to assuming that the reference motion (e.g.,

the motion at bedrock) has no site-to-site uncertainty and that all the site-to-site variability is due to the particularities of site response not captured by the site term in the GMPE.

Uncertainty in the shear-wave velocity of the soil profile can be quantified by a random field model (e.g. Toro 1995). The following section describes the development of a random field model using the KiK-net database to constrain the parameters of the model. Once the input variability is constrained, a Monte Carlo simulation approach can be used to compute how input variability propagates to the estimates of site response.

3.1 Random field model

A random field model to capture the variability in shear-wave velocity profiles is developed as part of this study. This model is based on the model published by Toro (1995). An alternate model using Markov Chains is also proposed. The parameters for both models are calibrated using the extensive shear-wave velocity database associated with the KiK-net network. Note that while one particular application of the model is presented here (e.g., the modeling of the KiK-net shear-wave velocity database), the proposed models can be calibrated using any existing dataset.

Modified EPRI

A model developed by Toro (EPRI 1993) has been modified to match the characteristics of the KiK-net database. The model is described herein. Let the natural logarithm of shear-wave velocity as a function of depth be defined by a random field V(z) where z is depth. Assume that

$$V(z) = t(z) + \varepsilon(z) \tag{3.1}$$

where $t(z)$ is the median value of the logarithm of the shear-wave velocity, and $\varepsilon(z)$ is a Gaussian random process with zero mean, standard deviation $\sigma_\varepsilon(z)$ and auto-correlation $\rho(\Delta z,z)$ given by:

$$\rho(z,\Delta z) = (1-\rho_S) * \exp(-\Delta z/\kappa) + \rho_S \tag{3.2}$$

where ρ_S and κ are given by

$$\rho_S(z) = \rho_o - \exp(-(z-z_o)/b) \geq 0 \tag{3.3}$$

$$\kappa = \kappa_o + (\kappa_{max} - \kappa_o) * (z/z_{max}) \geq 0 \tag{3.4}$$

where ρ_0, z_0, b, κ_o, κ_{max}, and z_{max} are fitting parameters, and where κ is limited to κ_{max} ($\kappa < \kappa_{max}$). The parameters fitted to the entire KiK-net data set and for subsets based on Vs30 site categories are shown in Table 3.1. The parameters were obtained using the Levenberg-Marquardt algorithm for nonlinear regression method implemented into Matlab. The geostatistics of the KiK-net database are shown in Figure 3.1.

The geostatistical model in Equations 3.1 to 3.4 can be used in conjunction with a routine to generate artificial shear-wave velocity profiles that replicate the statistics of the underlying data. The first step for generating artificial profiles is to determine the depth to the bottom of the profile. By definition, the bottom of the profile corresponds to the location of the elastic half space in site-response analyses. For the KiK-net database, the dept to the bottom of the profile follows a lognormal distribution. Parameters of the lognormal model for the KiK-net database are shown in Table 3.2.

Once the bottom of the profile is defined, the next step consists in the generation of the layering for each profile. Profile layering is assumed to follow a nonstationary Poisson model (EPRI 1993) with parameter λ^{-1}, where λ is a model parameter that represents the frequency of layer transitions at each depth. It is assumed that λ varies with depth as given by:

$$\lambda(z) = a(b+z)^{-c} \tag{3.5}$$

where a, b, and c are parameters that can be constrained with data. Table 3.3 shows the regression results for the coefficients of Equation 3.5 for each Vs30 based site category, and Figure 3.2 shows the variation of λ with depth for the generic case (i.e. the entire database). The thickness of each layer is then generated using:

$$\exp[-\lambda(z)\, h]=1-u \tag{3.6}$$

where u is a standard normal random variable and h is profile thickness. A realization of u is obtained using a random number generator and for each realization of u a profile thickness h is obtained. Once the layering is generated for each profile using Equation 3.6, the shear-wave velocities are generated using the statistical model given in Equations 3.1 to 3.4. The model generation routine is described in EPRI (1993) and is summarized in Figure 3.3.

One layer lag Gaussian Model

Figure 3.4 shows the correlation structure that results when considering only adjacent layers for profiles in the KiK-net shear-wave velocity database. The strong correlation of shear-wave velocity values between adjacent layers indicates the suitability of a model that uses only this correlation as correlation structure. Moreover, for the KiK-net data set and subsets based on Vs30 classes the correlation function is nearly independent of lag distance; that is, the random field that represents shear-wave velocity is stationary. Figure 3.5 shows the correlation structure resulting from considering only layers that are separated by two layers. Note that the correlation coefficients reduce drastically, but the correlation structure could be also modeled using a constant correlation coefficient. Table 3.4 shows the stationary correlation values for each category considering one-layer lag.

Markov Chain

A model based on the Markov Chain concept is also proposed for the generation of the random profiles. The model is constrained by the statistics and correlation structure of the underlying database (the KiK-net database in this case). The first step of the model is the generation of shear-wave velocity values adequately correlated with layer depth (Figure 3.6). The second step is the assemblage of the correlated pairs into random shear-wave velocity profiles.

This model uses the fact that shear-wave velocities of consecutive layers in the KiK-net database are strongly correlated. The generation of the correlated shear-wave velocity-layer depth pairs (Vs-depth pairs) is done by means of rank correlation coefficients (RCC), as opposed to simply using linear correlation coefficient (ρ).The selected RCC for this study was Kendall's τ. The procedure for the generation of the correlated variables is the following.

a) Compute the RCC (i.e. Kendall's τ) for the data set.
b) Transform that coefficient to Pearson's ρ using

$$\rho = \sin(\tau \pi / 2) \tag{3.7}$$

c) Generate a set of pairs of correlated normal Gaussian random variables using the correlation coefficient computed in (b),

d) Obtain the CDF for each variable.

The CDF's obtained in (d) can be inverted to obtain the marginal distribution constrained by the shear-wave velocity and layer depth that are constrained by the database. The marginal distribution for shear-wave velocities in the KiK-net database can be modeled by a lognormal distribution, with mean and standard deviation of the associated normal distribution of 6.2 and 0.84 respectively. The marginal distribution for layer depth can be modeled by a lognormal distribution, with mean and standard deviation of the associated normal distribution of 3 and 1.3 respectively. χ^2 goodness of fit test confirmed this fit to be appropriate. The result of this approach is a set of randomly generated pairs of shear-wave velocity and layer depth that follow the same correlation structure as the original database (Figure 3.6).

The second step of the Markov Chain model generation routine consists in the use of the generated Vs-depth pairs to construct realistic shear-wave velocity profiles. This is achieved by first creating depth and Vs bins that span the entire range of possible shear-wave velocity and depth values. A *state* is then defined as a combination of a Vs bin and depth bin. As a simple example, we could define depth and Vs bins as:

depth bin 1 = [0, 50]

depth bin 2 = [50, 150]

Vs bin 1 = [360, 560]

Vs bin 2 = [560, 760]

All possible *states* (i.e. d1V1, d1V2, d2V1, and d2V2) define the *state vector* (P_i), which in this simple example would be a 4 by 1 vector. For a given layer, the state vector takes a value. In this example, the state vector for layers *i* would have 3 zeros and a 1 in the position that corresponds to the Vs-depth combination of layer *i*.

A first order Markov chain methodology is then used for the profile generation. The procedure consists on determining the transition matrix (T), generating an initial Vs and layer depth pair, and multiplying the transition matrix times the state vector, which gives the next state vector as given by

$$P_i = T * P_{i-1} \tag{3.8}$$

where P is the state vector and T the transition matrix. In matrix form equation 3.8 takes the form

$$\underbrace{\begin{pmatrix} state_1 \\ state_2 \\ \vdots \\ state_n \end{pmatrix}}_{P_i} = \underbrace{\begin{pmatrix} P[state_1|state_1] & P[state_1|state_1] & \cdots & P[state_1|state_n] \\ P[state_2|state_1] & P[state_2|state2] & \cdots & P[state_2|state_n] \\ \vdots & \vdots & \ddots & \vdots \\ P[state_n|state_1] & P[state_n|state_2] & \cdots & P[state_n|state_n] \end{pmatrix}}_{T} \underbrace{\begin{pmatrix} state_1 \\ state_2 \\ \vdots \\ state_n \end{pmatrix}}_{P_{i-1}} \tag{3.9}$$

The transition matrix corresponds to the conditional probability that the Vs-layer depth pair in layer *i* be any state given a current state (represented by $P_{i\text{-}1}$). The first step in constructing the transition matrix is defining the desired accuracy, which is given by the number of Vs-depth bins. In this study shear-wave velocity and layer depths were divided into 6 bins each, giving 36 possible states (the length of the state vector). One additional case was added to include the probability that given $P_{i\text{-}1}$ the profile ends at that depth, thus the transition matrix is a square matrix of size 37 and the state vector is 37 by 1. The process is initiated by selecting an initial state using the surface shear-wave velocity distribution. Equation 3.9 is then applied generating a state vector where each element corresponds to the probability of each state (e.g. Vs-depth bin) occurring. This probability vector is used to randomly select a state vector with only one non-zero state. This process is repeated iteratively until all layers within a profile have been assigned a state (Vs-layer depth range). The specific Vs and depth values of each layer are calculated sampling the Vs-depth pairs computed in step one within the limits set by the corresponding state. This procedure is summarized in Figure 3.7.

Evaluation for results

Three artificially generated sets of 600 profiles each were generated using each of the three models proposed in this study. The statistics of the KiK-net database and those of the artificially generated data sets are shown in Figures 3.8 to 3.9. These figures illustrate the remarkably good agreement between the statistics of the KiK-net database and those of the three sets of randomly generated profiles. The Gaussian (modified EPRI) model reproduces particularly well the statistics of the recorded shear-wave velocity profiles. The misfit of the Markov chain method to the measured correlation functions could be avoided by increasing the size of the transformation

matrix, that is increasing the number of bins in which depth and/or shear-wave velocities are divided. Moreover, because the transformation matrix (T) is constructed using the database, the increase in T's dimension would produce correlation functions that are closer to the measured ones until measured and generated curves would overlap.

Figure 3.10 shows the comparison of ratios of response spectra (RRS) for the three artificially generated sets of profiles and the set of measured shear-wave velocity profiles of the KiK-net database. The RRS were computed using equivalent linear analyses using the software SHAKE (for details, see Montalva 2010). There is disagreement between the Markov and stationary Gaussian models results with those calculated with the same input motions and the KiK-net database sites. This discrepancy can be attributed to the way the artificial sites were generated; that is, based on the values of Vs and depth of the previous layer, disregarding the correlation with layers further than one layer away. This dependence can be included in the Gaussian model by using a 3 by 3 covariance matrix instead of the correlation coefficient (equivalent to using a 2 by 2 covariance matrix), or by using a Markov chain of order 2. Without these modifications the RRS computed with the profiles generated by the model based on Toro's work (Toro 1995; EPRI 1993) gives the closest results to the RRS computed for the measured shear-wave velocity profiles. Variability of RRS as reproduced by the three models is shown in Figure 3.11. The comparison with the variability observed in KiK-net database sites shows relatively good agreement in all three cases, the closest results are found for the non-stationary Gaussian model.

3.2 Monte Carlo simulation approach

An estimate of the site term (η_{site}) and its uncertainty ($\sigma_{\eta,site}$) can be obtained by means of a site response analysis code (e.g. Shake; Schnabel et al. 1972) along with a Monte Carlo simulation approach. Monte Carlo simulation has been used in the past for propagating uncertainties in site response analysis (e.g. Baturay and Stewart, 2003) and in general in various problems where input uncertainty must be propagated through a nonlinear system. The Monte Carlo simulation approach adopted herein makes use of the Latin Hypercube sampling technique to reduce computational cost. Two types of analyses were conducted. In the first case, we assume that a shear-wave velocity profile has been measured but there is an associated measurement uncertainty. We call this the site-specific approach. In the second case, we consider that a range

for the average shear-wave velocity of the site is known (i.e. site class), but details of the profile are missing.

The Monte Carlo simulation approach consists of the following steps:

1) Ground Motion Selection

The variables that were considered to play a role in site response analyses are ground motion characteristics, ground motion intensity, and the site profile, which in this study is parameterized by Vs30. A random sample of 100 ground motions from the NGA database was selected. These ground motions satisfy certain general criteria on magnitude and distance form the fault ($6 < Mw < 8$; $R_{jb} < 100$ km, and $Vs30 > 650$ m/s). A subset of 12 ground motions with response spectra that preserved the median and standard deviation of the 100 ground motions was selected.

Input ground motions were scaled prior their use in ground response analyses. The intent of the scaling was to provide an ensemble of time histories with varying intensities. Linear scaling was applied so that peak ground accelerations were 0.1, 0.2, 0.3, 0.4, and 0.5 g for all time histories.

2) Profile Generation

The fist type of profile generation consists on a set of randomly generated shear-wave velocity profiles obtained by assuming that the measured average Vs of each profile has a normal distribution centered on the recorded value with a coefficient of variation of 6%. This coefficient of variation was reported by Marosi and Hiltunen (2004) and Moss (2008) as a typical SASW measurement error in the average Vs over the upper 30 m (see Table 4.1). Profiles were then generated by introducing random variability around the Vs profile obtained from the measured dispersion curve, and ensuring that the generated Vs maintains a coefficient of variation of 6%. Note that this methodology for generating random profiles assumes no spatial correlation other than the trend given by the measured Vs profile. Moreover, this methodology does not assure that each realization of the Vs profile is compatible with the measured dispersion curve. While this is a simplification, parametric studies showed that the simplified model captured well uncertainties in estimated RRS values.

The second type of profile generation makes use of the three random field methods outline in Section 3.1. For each of them 100,000 sites were simulated for Site C (360 m/s < Vs30 < 760 m/s) and Site D (180 m/s < Vs30 < 360 m/s).

From each of the 6 generated samples (of 100,000 sites each), 6 groups with different standard deviations of their Vs30 values were selected. The mean value of Vs30 was kept constant for each of the sub-groups of each site class. The Vs30 sampling was done using a beta distribution that has the bounds of the site category (C or D), and has the desired mean and standard deviation. The distribution is fully characterized by parameters "*p*" and "*q*" and is obtained using:

$$p = \hat{\mu}\left(\frac{\hat{\mu}(1-\hat{\mu})}{\hat{\sigma}^2} - 1\right) \tag{3.10}$$

$$q = (1-\hat{\mu})\left(\frac{\hat{\mu}(1-\hat{\mu})}{\hat{\sigma}^2} - 1\right) \tag{3.11}$$

where *a* and *b* are the Vs30 bounds of each site class and $\hat{\mu}$ and $\hat{\sigma}$ are given by:

$$\hat{\mu} = \frac{\bar{x} - a}{b - a} \text{ and } \hat{\sigma}^2 = \frac{s^2}{(b-a)^2}$$

This characterization makes it possible to fix the mean (μ), and then solve for the values of *p* and *q* for a desired standard deviation value (σ).Latin Hypercube sampling was then used for the selection of 100 sites from the 100,000 sites sample. Using the distributions (defined by *p* and *q*) with custom standard deviations, the 6 groups for each method and soil class were created.

3) Site Response Analyses

All the combinations of sites groups, ground motions (each scaled to 5 different intensities), and site classes were analyzed for each random field methodology. Acknowledging the limitations of the linear-equivalent analysis methods but considering its widespread use and its suitability for very large number of analyses, the analyses were done using Shake91 (Schnabel et al. 1972, Idriss and Sun 1992). Details of the site response analyses are given in Montalva (2010).

4 Proposed Methodologies for computing site-specific hazard

The proposed methodology for computing site specific hazard is based on the Bazurro and Cornell (2004b) methodology described in Section 1.2. This methodology is based on the estimation of ground motions (both median and standard deviation) at a reference rock site using GMPEs, and modifying this records through site response analyses (Equations 1.1 to 1.4).

We propose herein that median ground motions at the reference stations be estimated using GMPEs, but the standard deviation estimates from GMPEs must be modified or replaced by single-station estimates of standard deviation. As has been previously suggested (e.g., Joyner and Boore 1993; see also Section 2 of this report), the standard deviation includes a variability that reflects uncertainty on the site term. This uncertainty is replaced in site-specific analyses by an estimate of the site term and hence should not be included in the estimate of the reference ground motion. While for a given site the site term is a deterministic value, there is an epistemic uncertainty associated with its estimation (σ_{lnAF}) that can be quantified using the methodology presented in Section 3.

Details on the methodology are presented below. The case where it is assumed that enough information exists to conduct site response analysis is treated separately from the case in which the amplification factor is estimated using correlation with Vs30 based site categories.

4.1 Site-specific site response analyses

This methodology applies when site specific site response analyses are desired and enough information is available on the profile to merit site-specific analyses. As described by Bazurro and Cornell 2004b (and summarized in Chapter 1), the spectral acceleration at the soil surface is estimated using Equations 1.1 and 1.3 with coefficients obtained from site response analyses. We propose that the reference ground motion (lnS_a^r) be estimated using GMPEs corresponding to the basement rock condition at the site. The standard deviation of the reference ground motions, $\sigma_{lnS_a^r(f)}$, however, should be obtained from single-station estimates of standard deviation. Section 2 presents an analysis of single-station standard deviation for the KiK-net database, and the results are summarized in Table 2.2. Similar studies in other regions have resulted in similar estimates of single-station standard deviation (e.g. Chen and Tsai 2002, Atkinson 2006).

An important consideration is whether single-station standard deviations are dependent on site conditions. Figure 4.1 shows the estimated single-station standard deviations at selected periods from 44 stations in the KiK-net database that recorded 15 or more earthquakes. Observe that there is no clear trend in the value of single-station standard deviation with the Vs30 value. Similar observations were made when plotting as a function of other parameterizations of the site profile. These results appear to indicate that estimates of single station standard deviation that are made at any station are not necessarily biased as a function of the site profile. Note, however, that the value of the single-station standard deviation can vary significantly from one station to the other (Figure 4.1).

Caution should be used when considering the values of single station standard deviation listed in Table 2.2 because they are generally obtained from low magnitude earthquakes and it is not clear how they extrapolate to higher magnitude earthquakes. A study currently underway by the PI is addressing this issue (USGS Award G10AP00029).

The computation of the standard deviation at the soil surface (Eq. 1.4) also needs an estimate of the standard deviation of the amplification factor (σ_{lnAF}). A lower bound of this variability can be obtained from ratios of surface to downhole ground motion in the KiK-net array. These lower bounds are only approximations to the actual σ_{lnAF} because they do not include epistemic uncertainty on the soil profile and because they are obtained from surface to downhole records rather than surface to outcrop record. Estimates of this uncertainty from the KiK-net array are shown in Figure 4.2 as a function of Vs30 for KiK-net sites with 15 or more recordings and at two selected periods.

The computation of standard deviation for the soil, given by Eq. 1.4 assumes that the residuals of the spectral acceleration at rock and the residuals of the amplification factor are uncorrelated. If the assumption that lnS_a^s is a linear function of lnS_a^r holds (Equation 1.3), then Equation 1.4 can be rewritten as:

$$\sigma_{lnS_a^s(f)} \approx \sqrt{(c_1+1)^2\left(\sigma_{lnS_a^r(f)}\right)^2 + \left(\sigma_{lnAF(f)}\right)^2 + 2(c_1+1)\rho(f)\sigma_{lnS_a^r(f)}\sigma_{lnAF(f)}} \qquad (4.1)$$

where $\rho(f)$ is the correlation coefficient between ε_{lnAF} and $\varepsilon_{lnS_a^r(f)}$ (see Equation 1.2). Note that $\sigma_{lnS_a^r(f)}$ here is the single station standard deviation. Hence, the correlation factor $\rho(f)$ must be estimated using single station records. Figure 4.3 shows the estimates of the correlation coefficients for the 44 stations that recorded 15 or more records in the KiK-net database. The estimates of $\rho(f)$ have large variability (at each period, the estimates of $\rho(f)$ is averaged from the $\rho(f)$ for each of the 44 stations). However, both for short and for long period the value is distinct form zero, and at all periods the median value is negatively correlated as suggested by McGuire (see Bazurro and Cornell 2004b). Note, however, that the value of the correlation coefficient can be very different from zero and neglecting it in Equation 1.4 can lead to an overestimation of the standard deviation of soil ground motion. The correlation coefficients are also plotted for selected periods versus the site condition parameterized by Vs30 (Figure 4.4). No clear correlation is shown with Vs30, indicating again that estimates of $\rho(f)$ from one station will not have systematic biases due to site conditions and may be used for other stations.

The KiK-net array has the benefit of having both surface and downhole records. In general practice, the amplification factors and their standard deviation can be estimated using site response analyses. Past studies have indicated that this variability is around 0.3 for short periods and as high as 0.6 for long periods (Bazurro and Cornell 2004a). Abrahamson and Silva (2008) proposed a constant value of 0.3 based on simulation results. A comparison of measured and predicted amplification factors using equivalent linear analysis is shown in Figures 4.5 and 4.6 for two selected stations in the KiK-net database. Observe that the 1-D assumption appears to hold well for these stations and the amplitude of the amplification factors is predicted well. Also note that both the standard deviation of the amplification factor and the correlation coefficient between the Amplification factor and the input spectral accelerations are well predicted. Note that the comparisons are made between surface and borehole records to match the empirical estimates from the KiK-net stations (e.g., the amplification factor is between the surface and the "within" motion).

The results shown in Figures 4.5 and 4.6 assume no uncertainty in the site profile. When shear-wave velocity is measured at a given profile, measurement uncertainty must also be included. Various estimates of measurement uncertainty for shear-wave velocity profiles are shown in Table 4.1.

Figure 4.7 shows the standard deviations at the surface computed using equivalent linear analyses for a site that recorded the 2001 Southern Peru Earthquake (Rodriguez-Marek et al. 2010). The results are presented for two cases: a case where the ground motions are varied, and a case when the ground motion is a constant but the site profile is varied. While this is a particular case, is shows how the larger uncertainty results from variations in the ground motion and not from uncertainty in the site profile. Bazurro and Cornell (2004a) indicate that site variability appears to be comparable in magnitude to the record-to-rcord variability in the amplification factor only for long periods. For short periods (periods shorter than the natural site period), the inclusion of uncertainty in soil parameters does not increase the uncertainty in amplification factors by more than 20%.

It is important to note that an approximation to site-specific hazard can also be obtained by using Ground Motion Prediction Equations (GMPE) that include a site term based on a simplified parameterization (e.g., Vs30, Depth to Vs = 1.0 km/s at the site in the NGA relationships). When additional information at a site is known (such as the site specific profile), the estimates of ground motion at the site can be improved in some cases by site specific analyses (Baturay and Stewart 2003). In these cases, the proposed methodology would be an improvement over the baseline approach because it leads to a better estimate of the site term. In cases where the site-specific analysis results in a site term that is similar to that obtained using the baseline case, the proposed methodology can still lead to better estimates of ground motion if the resulting uncertainty as computed by Equation 4.1 using single-station estimates for the reference site $\sigma_{lnS_a^r(f)}$ and the site-specific estimate of the amplification factor uncertainty ($\sigma_{\ln AF}$) is lower than the uncertainty predicted by the GMPE.

4.2 Site factor approach

The approach described in Section 4.1 is applicable when a measured site velocity profile at a station exists and site response analyses are possible. When site-specific shear-wave velocity is not available, the UHS at a site can be computed by performing a PSHA using GMPEs with an assumed value of Vs30 and a logic tree to account for epistemic uncertainty on the Vs30. Alternatively, the methodology outlined in Section 4.1 can be used. In this case, code-prescribed amplification factors can be used as the estimate of the mean of the amplification factor. In this

study we explore estimates of the uncertainty in the amplification factor for code-type factors using IBC site categories.

The two alternatives presented in the previous paragraph involve a full PSHA analyses. Depending on project constrains, that may not be viable. Alternatively, the hazard curve for soil can be computed using the convolution approach proposed by Bazurro and Cornell (2004b). This methodology is embedded in the equation below:

$$G_{S_a^s}\left(S_a^s = z\right) = \int_0^\infty P\left[AF \geq \frac{z}{x} \middle| S_a^r = x\right] f_{S_a^r}(x)dx \tag{4.2}$$

where $G_{S_a^s}(z)$ if the complementary cumulative distribution function of S_a^s (CCDF), $f_{S_a^r}(x)$ is the probability density function of S_a^r. Both x and z are dummy variables representing a test value for S_a^s and an integration variable for the integration over values of S_a^r. The dependency on frequency was dropped from Equation 4.2 for simplicity. Note that under the assumptions of low rate of occurrence (e.g., the rare event assumption), the CCDF is numerically equivalent to the mean annual rate of exceedance, hence $G_{S_a^s}(z)$ is the sought hazard curve for soil. Moreover, the probability density function for rock is the derivative of the CCDF and hence can be obtained as the value of the slope of the hazard curve for rock at each S_a^r value.

In the Bazurro and Cornell methodology, $f_{S_a^r}(x)$ is the mean rate density for ground motion at rock, which is obtained from a hazard curve at a rock site (a reference site). These curves were used by the USGS to develop the national hazard maps and should be available. However, as already mentioned in Section 4.1, we suggest that the reference site should not include site-to-site variability; hence the hazard curve at the reference site should be computed using only single station variability. This is because the uncertainty in the amplification factor accounts for the uncertainty in the site term. The application of the integral in Equation 4.2 with a standard hazard curve for rock would imply that the site-to-site variability is double-counted.

In addition to the hazard curves at a reference site using single station standard deviation, the application of the proposed methodology also needs an estimate of the uncertainty in the site factor. The following section explores this uncertainty using the Monte Carlo simulation approach described in Section 3.

Amplification factors and their uncertainties

Amplification factors were estimated using equivalent linear site response analyses for a set of artificially generated profiles representative of the recorded shear-wave velocities at the KiK-net stations. The profiles are generated using the model presented in Section 3.1 (Modified EPRI model), and the Monte Carlo simulation approach follows the approach described in Section 3.2. The results are presented in terms of the site categories of the IBC for Site Category C (360 m/s < Vs30 < 760 m/s) and Site Category D (180 m/s < Vs30 < 360 m/s).

Figure 4.8 compares the amplification factors obtained in the Monte Carlo approach with those from the IBC. The simulation is conducted by an artificial generation of 600 profiles, site response are conducted for 12 ground motions selected according to the methodology discussed in Section 4.2. Figure 4.8a shows the short period factors from the code compared with the amplification factor at a period of 0.3s. Note that the code factors are in general lower than the factors computed in this study. A similar observation was made by Rodriguez-Marek et al. (2001) using data from the Northridge and Loma Prieta earthquakes, and on computing amplification factors for 4 stations that recorded the 2001 Southern Peru Earthquake (Rodriguez-Marek et al. 2010). Figure 4.8b compares the long period factors from the code with factors at 1 s from the simulations. In this case, the code factors are higher than those obtained from the simulation. The same trend of over-prediction of long period factors by the code was observed when comparing code factors with factors obtained from the Northridge and Loma Prieta ground motion data (Rodriguez-Marek et al. 2001) and for computed site factors for stations that recorded the 2001 Southern Peru (Rodriguez-Marek et al. 2010).

Factors computed for Site D are compared with site factors in Figure 4.8c and d, for the short- and long-period factors, respectively. For the short period factor, the computed factors overpredict those from the code, while for the long-period factor, the predictions and the code are approximately equal. The trends are similar to the trends observed in Northridge and Loma Prieta data (Rodriguez-Marek et al. 2001).

A trend of nearly constant AF with PGA is observed in the simulations results for both Site C and Site D for both short and long periods for Site C, and for long period factors for Site D. This trend is unexpected and does not match well with results of past studies. However, as suggested previously by Bazurro and Cornell (2004a), the amplification factor at a given frequency f is

better predicted by the spectral acceleration at the same frequency. These plots are shown in Figure 4.9 for Site C and Figure 4.10 for Site D. A linear relationship between the log of the amplification factor (lnAF) and the spectral acceleration is generally observed, as implied in Equation 1.3.

The uncertainty in the amplification factors (σ_{lnAF}) is needed for the site-specific analysis methodology proposed in this work. The values estimated for the KiK-net stations using the Monte Carlo approach are shown in Figures 4.11 and 4.12, respectively. Observe that the uncertainty values for site C at short periods and low intensities (T=0.3 s), and site D at long periods (T = 1.0s) are around 0.3. As indicated previously, past studies have indicated that this variability is around 0.3 for short periods and as high as 0.6 for long periods (Bazurro and Cornell 2004a). Abrahamson and Silva (2008) proposed a constant value of 0.3 based on simulation results. For C sites at short periods (Figure 4.11 a), the value of σ_{lnAF} increases to about 0.4, while at long periods, the value is generally smaller than 0.2. This is consistent with the low amplitude of the long period amplification factors for Site C. For Site D, the value of σ_{lnAF} for short periods is significantly larger than 0.3 and, significantly, the uncertainty in the amplification factor increases with increasing ground motion intensity.

Figures 4.13 and 4.14 explore the variability in the amplification factors as a function of the variability in Vs30. These figures were generated by selecting 600 profiles out of a larger dataset of generated profiles such that the 600 profiles have a common mean Vs30 value, but their standard deviation has a desired value. The standard deviation in Vs30 value would reflect the uncertainty on the value of Vs30 that could result from either measurement errors or simply an absence of measurements. Surprisingly, there is no clear trend in the variability of σ_{lnAF} with uncertainty in Vs30. This seems to indicate that when dealing with large uncertainties in the ground motion profile (uncertainties both in Vs30 and in layering), the uncertainty in the amplification factor can not be reduced with an improvement on the knowledge of Vs30 alone.

Figures 4.13 and 4.14 also illustrate the variability in amplification factors due to differences in the input ground motion. The uncertainties shown already account for variability in the amplification factor with the input spectral acceleration (e.g., they are variabilities in the residuals of the amplification factor). The median line shown as a thick line in these figures corresponds to the median value for response at all the profiles in consideration (600 profiles for

each value of Vs30) and 12 ground motions with 5 different scaling values. The thick lines correspond to a one standard deviation band, with the uncertainty of σ_{lnAF} corresponding to a sample standard deviation obtained using the 12 ground motions. Note that the additional uncertainty due to the ground motion variability is relatively small when compared with median values of σ_{lnAF}. These estimates, however, are biased by the small sample size of earthquake ground motions.

5 Conclusions

This study presents a methodology for developing site specific hazard curves. The proposed methodology is based on the work of Bazurro and Cornell (2004b), with the significant difference that we propose that the standard deviation of the ground motion at the reference rock (an input parameter to the analyses) should be estimated using single-station estimates of standard deviation.

Single-station estimates of standard deviation were obtained using the KiK-net database. Two important observations can be made with respect to the estimated single-station standard deviations:

- The values of single-station standard deviation obtained in this study are similar to those obtained by other researchers using different ground motion databases.
- The values of single-station standard deviation for borehole records in the KiK-net database are remarkably similar to those at the surface.

The first observation implies that estimates of single-station standard deviation, such as those given in Table 2.2, can be used for regions outside of Japan, despite being obtained from exclusively Japanese records. The caveat is that single-station standard deviations may be magnitude dependent and all estimates of single-station standard deviation to date are made with small magnitude earthquakes, hence the value may be different for earthquake magnitudes that typically control hazard. The second observation implies that most of the station-to-station variability is controlled by a poor characterization of the site term (e.g., of site effects), likely as a result of the limited parameterization of site response (only Vs30 in this study).

The proposed methodology also needs an estimate of the amplification factor (spectral acceleration at the surface over spectral acceleration at a reference rock site) and its uncertainty. The median value of such factors has been the subject of numerous studies. Its variability, however, is not well reported. We used the KiK-net network to estimate the variability in amplification factors of surface to borehole records. This variability is a lower bound estimate of the variability of surface-to-surface ratios. This variability is found to be around 0.25, a value similar to that used by Abrahamson and Silva (2008) in the NGA relationships. In addition, we used the KiK-net database to estimate the correlation between the residuals of the amplification factor and the residuals of the reference-site (bedrock at borehole depth in this case). This correlation is negative (as suggested by Bazurro and Cornell 2004b), but for some periods, the absolute value of the correlation coefficient is sufficiently large so that it merits inclusion in the estimation of site-specific ground motions (Figure 4.3).

This study also presented a random field model for simulating shear-wave velocity profiles. The model is an update of an earlier model (EPRI 1993) and is calibrated using the KiK-net shear-wave velocity database. A new model based on the Markov-chains concept is also presented. The ability of the models to simulate the site-response behavior of real profiles was evaluated by comparing site response computed for recorded profiles, with the site response computed for artificially generated profiles. The proposed models give reasonable results, with a non-stationary model predicting site response closest to the predictions for the measured KiK-net sites.

The random field model was used within a Monte Carlo simulation approach to estimate the uncertainty in amplification factors. This uncertainty is needed as part of the proposed site-specific seismic hazard assessment. A comparison of an empirical estimate of amplification factors using surface and downhole KiK-net stations with predicted amplification factors using site response analysis indicates that the site response analyses are able to predict relatively well both median values of the amplification factors as well as their standard deviation (Figures 4.5 and 4.6). Standard deviation for the amplification factor (σ_{lnAF}) was computed as a function of IBC site category. The values of σ_{lnAF} were a function of the spectral acceleration at the reference site, but remarkably, no trend was seen when σ_{lnAF} were plotted with respect to the uncertainty in the Vs30 value used for site classification. The values of σ_{lnAF} are rather large for

general site categories when looking at periods where resonances could occur, implying that the proposed methodology is not likely to reduce amplification with respect to the direct use of GMPEs.

6 References

Abrahamson, N. A. and Youngs, R. R. (1992). “A Stable Algorithm for Regression Analyses using the Random Effects Model.” Bulletin of the Seismological Society of America 82(1): 505-510.

Abrahamson, N.A., and Silva, W. (2008). “Summary of the Abrahamson and Silva NGA Ground-Motion Relations”, Earthquake Spectra 24(1), 67-98.

Aoi, S., K. Obara, S. Hori, K. Kasahara, and Y. Okada (2000). New Japanese Uphole /Downhole, Strong-Motion Observation Network : KiK-net, Seismological Research Letters, 72, 239.

Asten, M. W., Boore, D. M. (2005). Comparison of shear-velocity profiles of unconsolidated sediments near the coyote borehole (CCOC) measured with fourteen invasive and non-invasive methods. U.S. Geol. Surv. Open-File Rept. 2005-1169.

Atkinson, G.M. (2006). Single-station sigma. Bulletin of the Seismological Society of America, 96 (2), 446-455.

Baturay, M. B., Stewart, J. P. (2003). Uncertainty and Bias in Ground-Motion Estimates from Ground Response Analyses. Bulletin of the Seismological Society of America 93 (5), 2025–2042.

Bazzurro, P., Cornell, C. A. (2004a). Ground-Motion Amplification in Nonlinear Soil Sites with Uncertain Properties. Bulletin of the Seismological Society of America 94 (6), 2090–2109.

Bazzurro, P., Cornell, C. A. (2004b). Nonlinear Soil-Site Effects in Probabilistic Seismic-Hazard Analysis. Bulletin of the Seismological Society of America 94 (6), 2110–2123.

Boore, D. and G. Atkinson (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped SA at spectral periods between 0.01s and 10.0 s. Earthquake Spectra 24(1), 99-138.

Borcherdt, R. D. (1994). Estimates of site-dependent response spectra for design (methodology and justification). Earthquake Spectra 10 (4), 617–653.

Brown, L. T., Boore, D. M., Stokoe, Kenneth H., I. (2002). Comparison of shear-wave slowness profiles at 10 strong-motion sites from noninvasive SASW measurements and measurements made in boreholes. Bulletin of the Seismological Society of America 92 (8), 3116–3133.

Chen, Y.-H., and C.-C. P. Tsai (2002). A new method for estimation of the attenuation relationship with variance components, Bull. Seismol. Soc. Am. 92, 1984–1991.

Chiou, B. S.-J., and Youngs (2008). An NGA model for the average horizontal component of peak ground motion and response spectra. Earthquake Spectra 24(1), 173-216.

Cotton, F., G. Pousse, F. Bonilla, and F. Scherbaum (2008). On the discrepancy of recent European ground-motion observations and predictions from empirical models: analysis of KiK-net accelerometric data and point-sources stochastic simulations. Bull. Seism. Soc. Am. , 98(5), 2244-2261.

Dobry, R., Oweis, I., Urzua, A. (1976). Simplified procedures for estimating the fundamental period of a soil profile. Bulletin of the Seismological Society of America 66 (4), 1293–1321.

Dobry, R., Borcherdt, R. D., Crouse, C. B., Idriss, I. M., Joyner, W. B., Martin, G. R., Power, M. S., Rinne, E. E., Seed, R. B. (2000). New site coefficients and site classification system used in recent building seismic code provisions. Earthquake Spectra 16 (1), 41–67.

EPRI, Electrical Power Research Institute (1993). Guidelines for Determining Design Basis Ground Motions. Report, TR-102293, Palo Alto: California.

Fujiwara, H., S. Aoi, T. Kunugi, and S. Adachi (2004). Strong-Motion Observation Networks of NIED : K-NET and KiK-net, www.cosmos-eq.org/Projects/Fujiwara_paper.pdf, abstract.

Fukushima, Y. (1996). Scaling relations for strong ground-motion prediction models with M2 terms, Bull. Seismol. Soc. Am. 86, 329–336.

Goldstein, P., D. Dodge, M. Firpo, and L. Minner (2003). SAC2000: signal processing and analysis tools for seismologists and engineers, in The IASPEI International Handbook of

Earthquake and Engineering Seismology, W. H. K. Lee, H. Kanamori, P. C. Jennings and C. Kisslinger (Editors), Academic Press, London.

Goulet, C. A., Stewart, J. P. (2009). Pitfalls of deterministic application of nonlinear site factors in probabilistic assessment of ground motions. Earthquake Spectra 25 (3), 541–555.

Idriss, I.M. (2008). An NGA empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. Earthquake Spectra 24(1), 217-242.

Idriss, I. M. and Sun, J. I. (1992). "User's Manual for SHAKE91, A computer Program for Conducting Equivalent Linear Seismic Response Analyses of Horizontally Layered Soil Deposits Program Modified based on the Original SHAKE Program Published in December 1972 by Schnabel, Lysmer and Seed."

International Code Council, ICC (2006). 2006 International Building Code. ICC, Washington DC.

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

Kinoshita, S. (1998). Kyoshin Net (K-NET), Seis. Res. Lett., 69(4) :309–332.

Marosi, K. T., Hiltunen, D. R. (2004). Characterization of spectral analysis of surface waves shear-wave velocity measurement uncertainty. Journal of Geotechnical and Geoenvironmental Engineering 130 (10), 1034–1041.

Marosi, K. T., Hiltunen, D. R. (2004). Characterization of spectral analysis of surface waves shear wave velocity measurement uncertainty. Journal of Geotechnical and Geoenvironmental Engineering 130 (10), 1034–1041.

Martin, A. J., Diehl, J. G. (2004). Practical experience using a simplified procedure to measure average shear- wave velocity to a depth of 30 meters (vs30). 13 World Conference on Earthquake Engineering, Vancouver, B.C., Canada.

Montalva, G. (2010). “Site-Specific Seismic Hazard Analyses.” Ph.D. Dissertation, Washington State University (to be published August 2010).

Moss, R. E. S. (2008). Quantifying Measurement Uncertainty of Thirty-Meter Shear-Wave Velocity. Bulletin of the Seismological Society of America 98 (3), 1399–1411.

Pousse, G., C. Berge-Thierry, F. Bonilla, and P. Y. Bard (2005). Eurocode 8 Design response spectra evaluation using the K-net Japanese database, J. Earthq. Eng. 9, 547–574.

Power, M., Chiou, B., Abrahamson, N., Bozorgnia, Y., Shantz, T., and Roblee, C. (2008). An overview of the NGA project. Earthquake Spectra, Volume 24, No. 1, pages 3–21.

Rodríguez-Marek, A, J. Bay, K.S. Park, G.A. Montalva, A. Cortez-Flores, J. Wartman, and R. Boroschek (2010). "Engineering analysis of ground motion records from the Mw 8.4 southern Peru earthquake." *Earthquake Spectra*, In Press.

Rodriguez-Marek, A., Bray, J. D., and Abrahamson, N. (2001) "An empirical geotechnical seismic site response procedure." *Earthquake Spectra*, 17(1), p. 68-88.

Schnabel, P., Lysmer, J., Seed, H. (1972). Shake - a computer program for equation response analysis of horizontally layered sites. Report, EERC, 72–12.

Toro, G. (1995). Probabilistic models of site velocity profiles for generic and site-specific groundmotion amplification studies. Department of Nuclear Energy Brookhaven National Laboratory, Upton, New York.

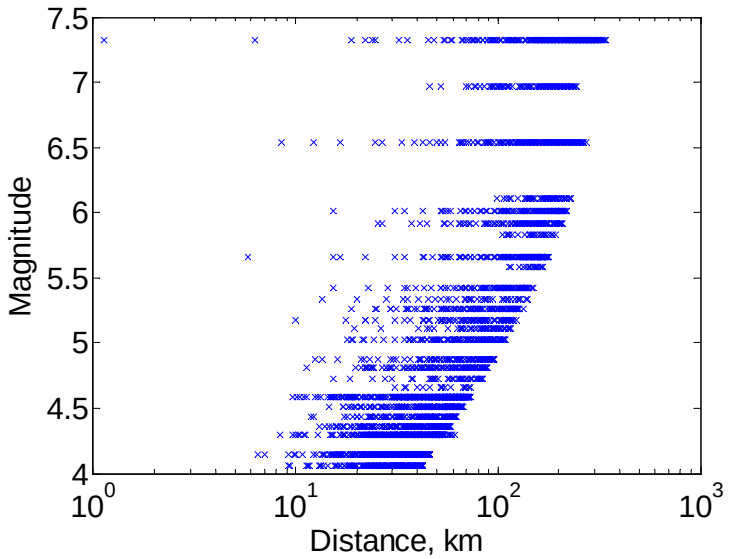


Figure 2.1: Magnitude-distance distribution of the KiKnet database (records up to October 2004).

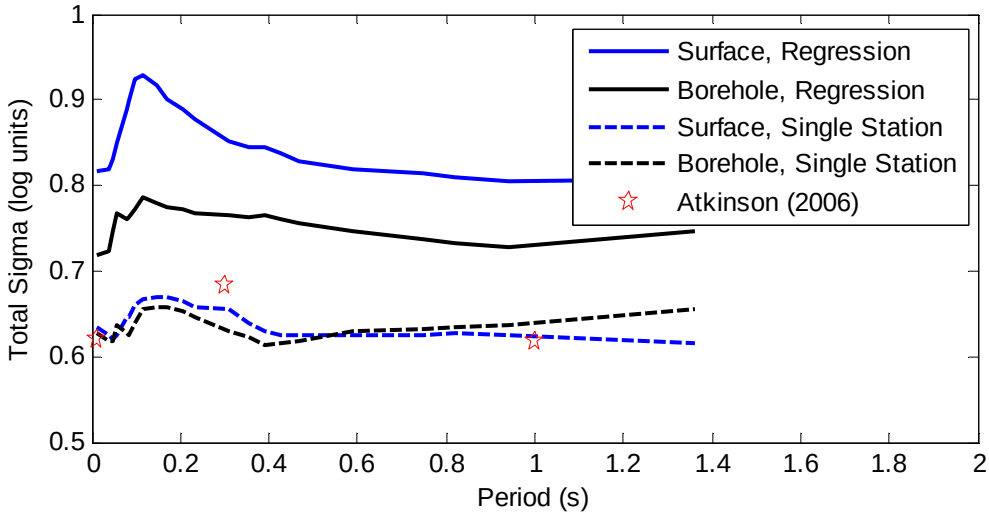
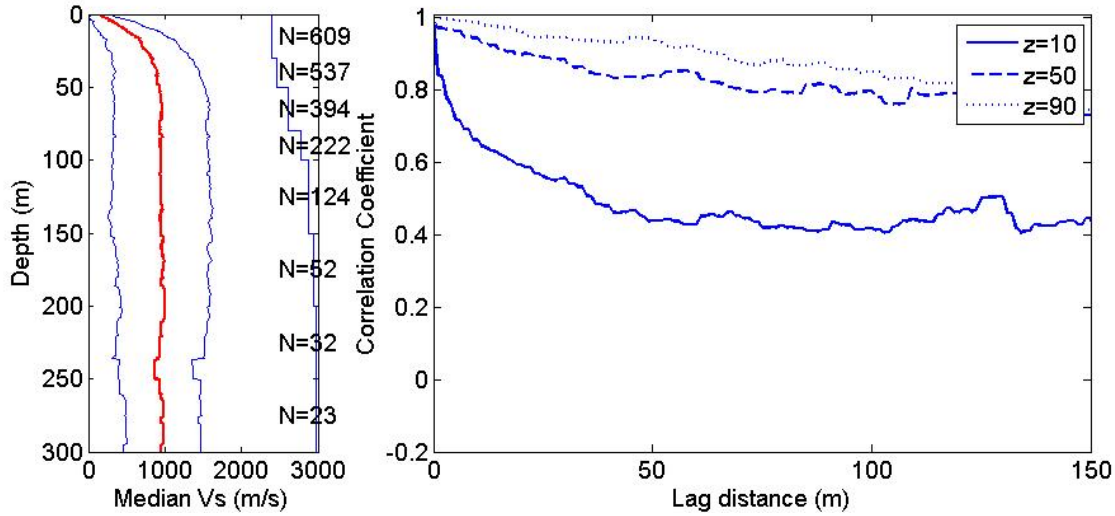
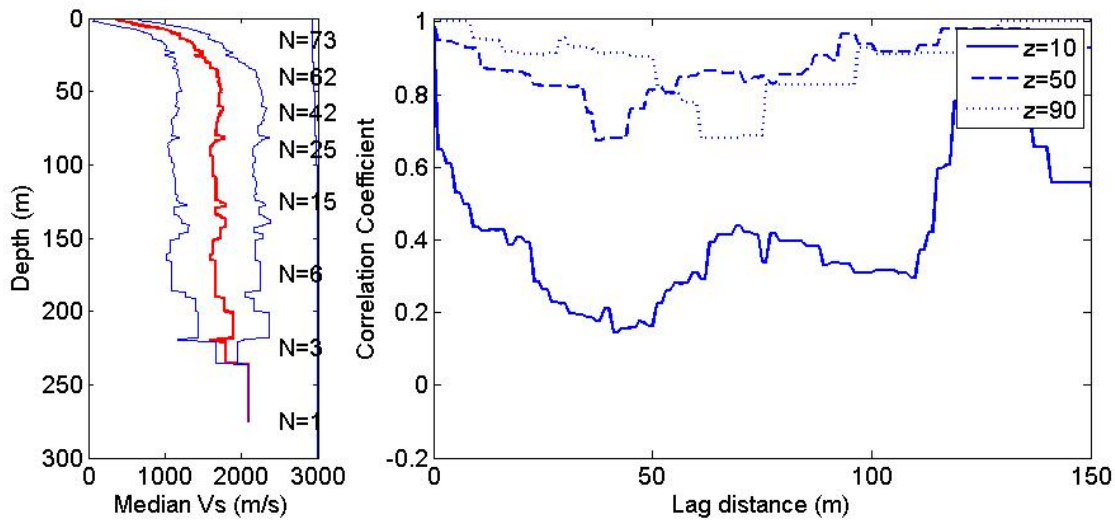


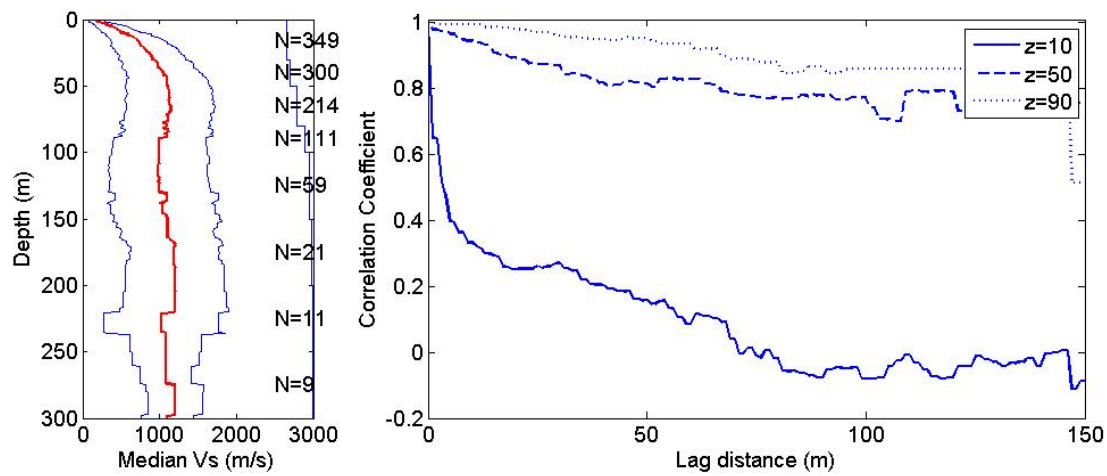
Figure 2.2. Preliminary results of an analysis of residuals using the KiKnet database with records up to October 2004. The solid lines are estimates of total standard deviation on the entire database. The dashed lines are similar estimates from single-stations that recorded more than 15 events.


N=609
N=537
N=394
N=222
N=124
N=52
N=32
N=23
Depth (m)
Median Vs (m/s)
Correlation Coefficient
Lag distance (m)
z=10
z=50
z=90

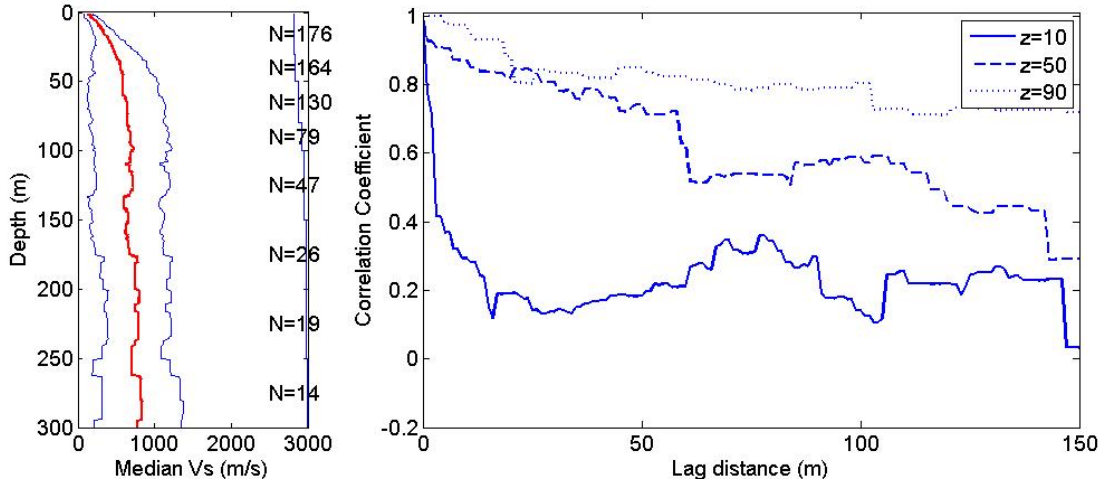
(a)


N=73
N=62
N=42
N=25
N=15
N=6
N=3
N=1
Depth (m)
Median Vs (m/s)
Correlation Coefficient
Lag distance (m)
z=10
z=50
z=90

(b)


N=349
N=300
N=214
N=111
N=59
N=21
N=11
N=9
Depth (m)
Median Vs (m/s)
Correlation Coefficient
Lag distance (m)
z=10
z=50
z=90

(c)



(d)

Figure 3.1 Shear-wave velocity statistics and correlation coefficients for entire data and subsets. (a) entire database, (b) site classes A and B, (c) site class C, and (d) site class D.

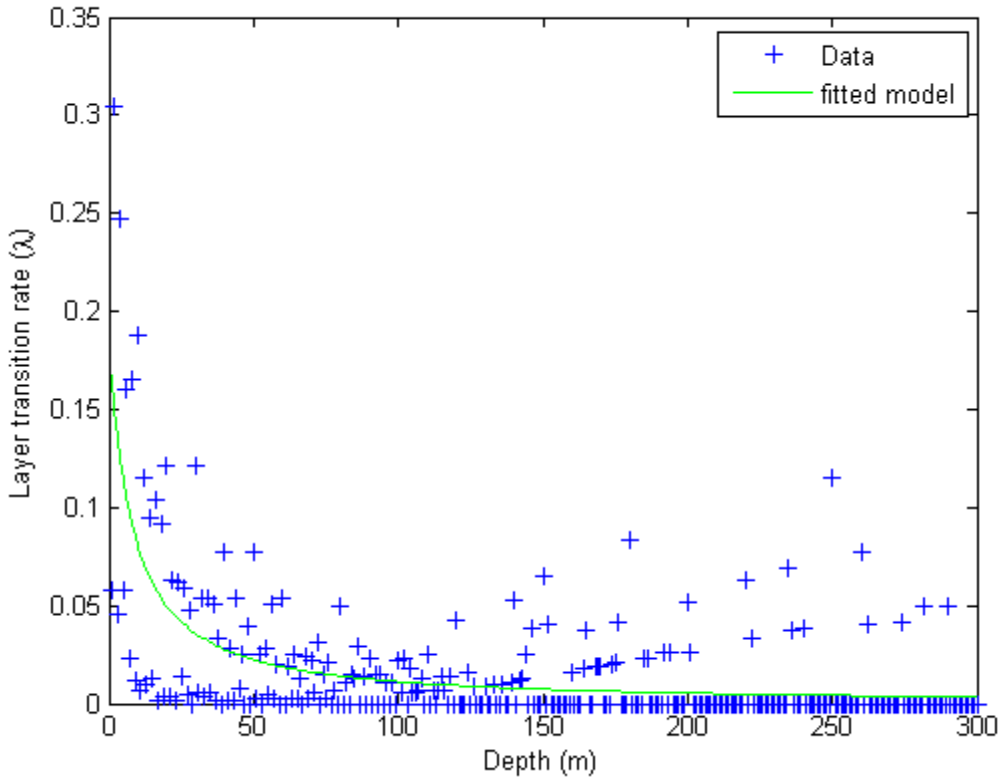


Figure 3.2. Rate of layer transitions versus depth for the generic case (λ). This parameter is the reciprocal of layer thickness for each depth.

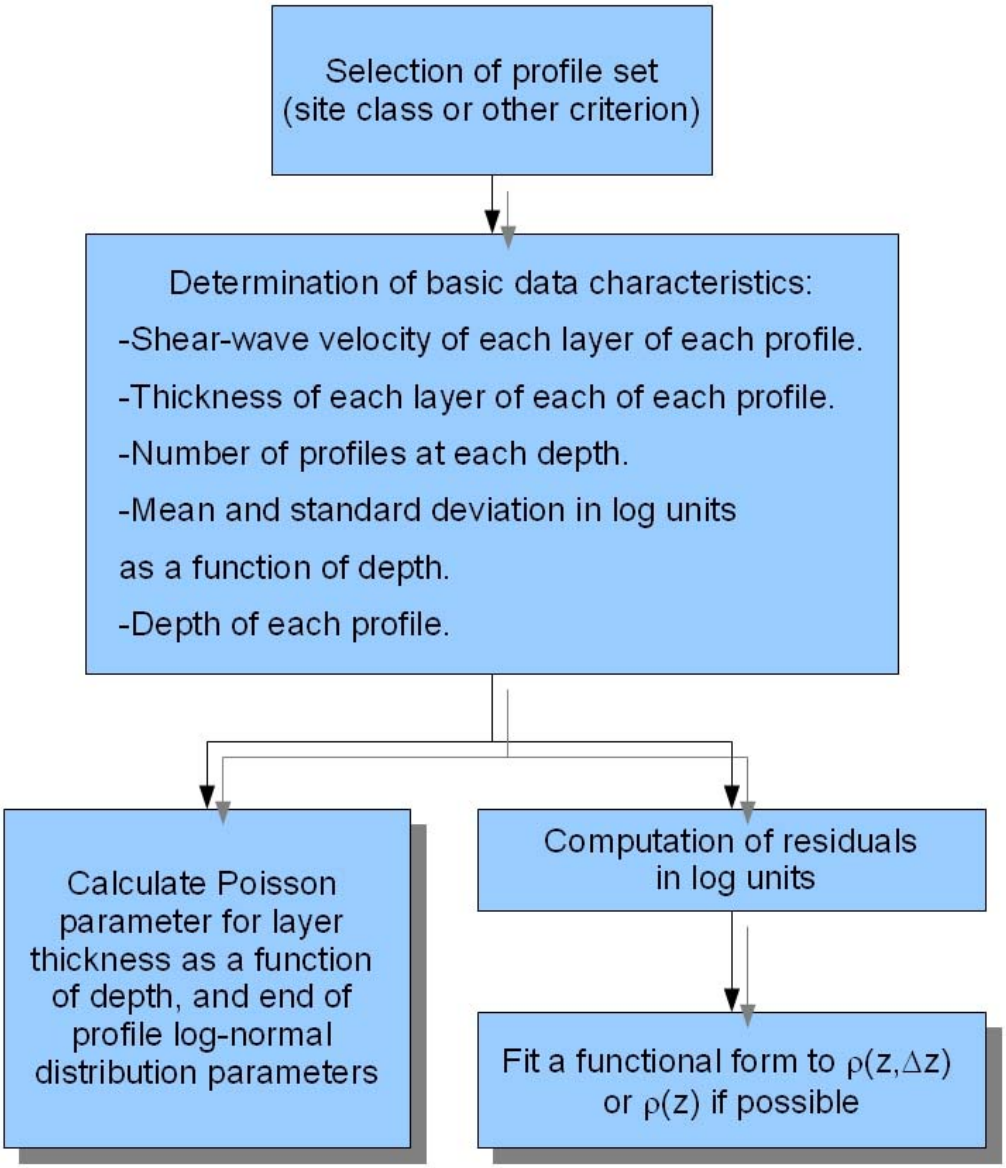


Figure 3.3: Flow chart for the Modified EPRI model.

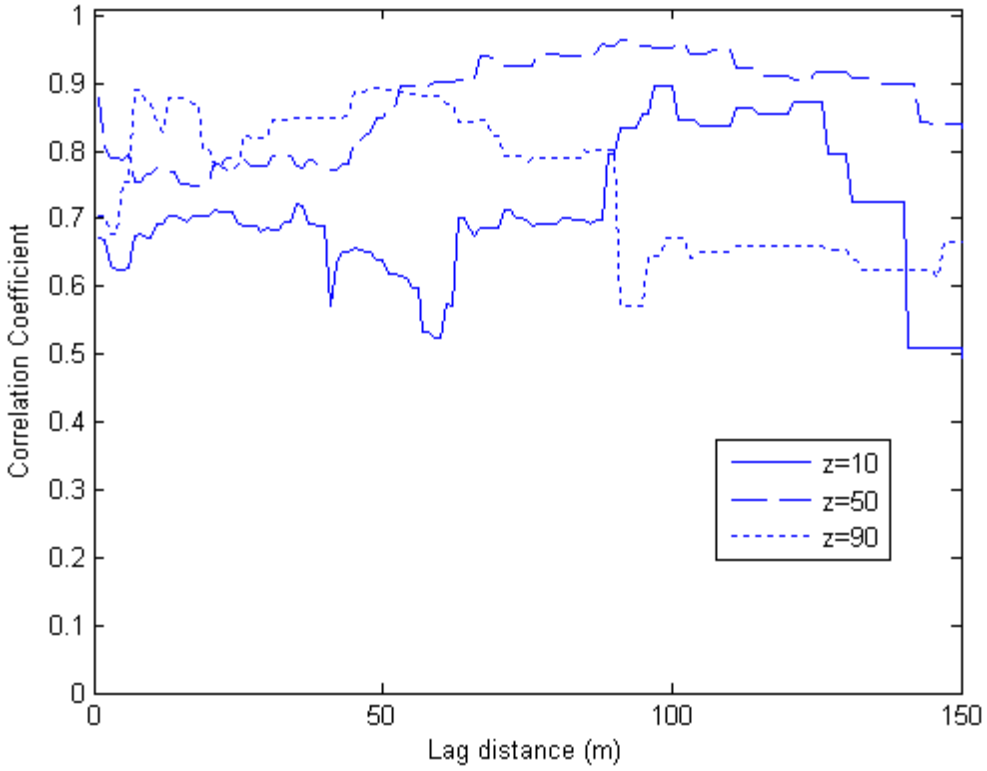


Figure 3.4 Shear-wave velocity correlation coefficients considering only lag distances that are one layer away from each other. Initial depths of 10, 50, and 90 m.

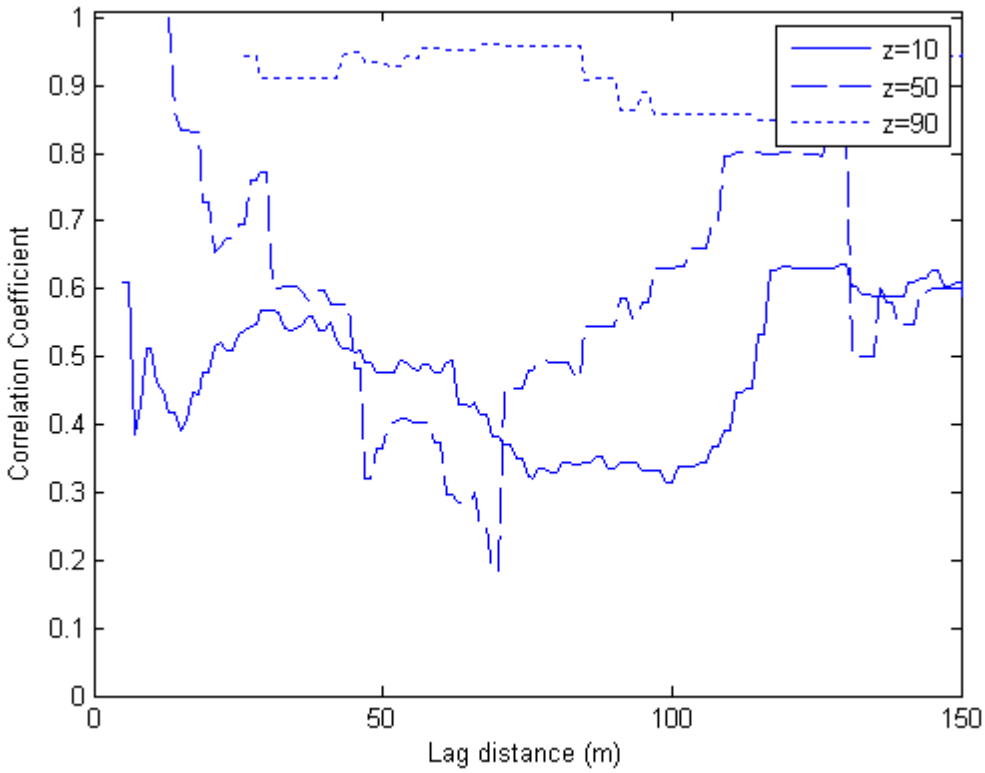


Figure 3.5 Shear-wave velocity correlation coefficients considering only lag distances that are two layers away from each other. Initial depths of 10, 50, and 90 m.

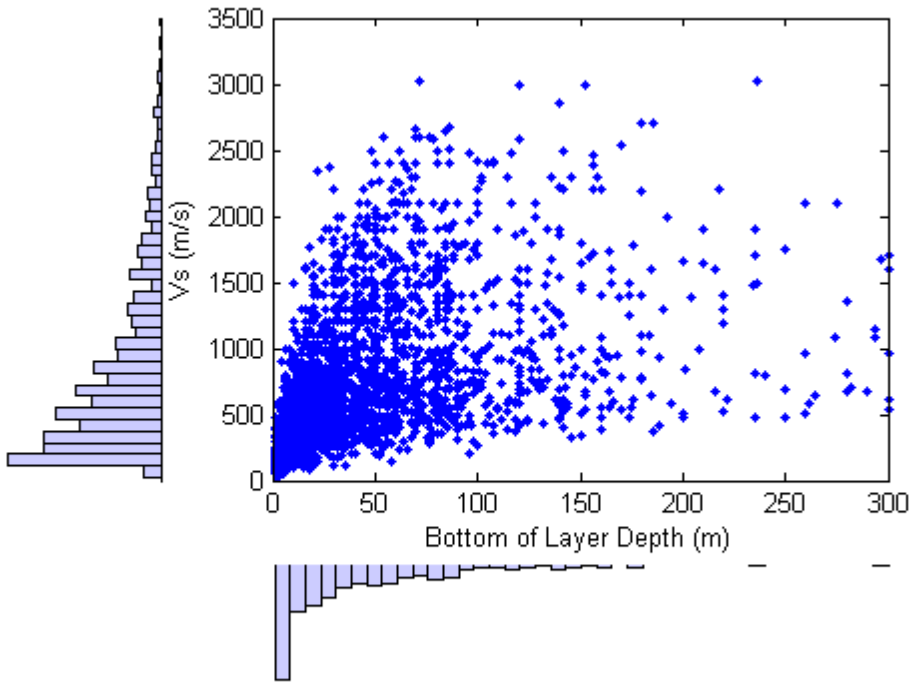


Figure 3.6. KiK-net database correlation between layer depth (defined as the depth to the bottom of the layer) and shear-wave velocity.

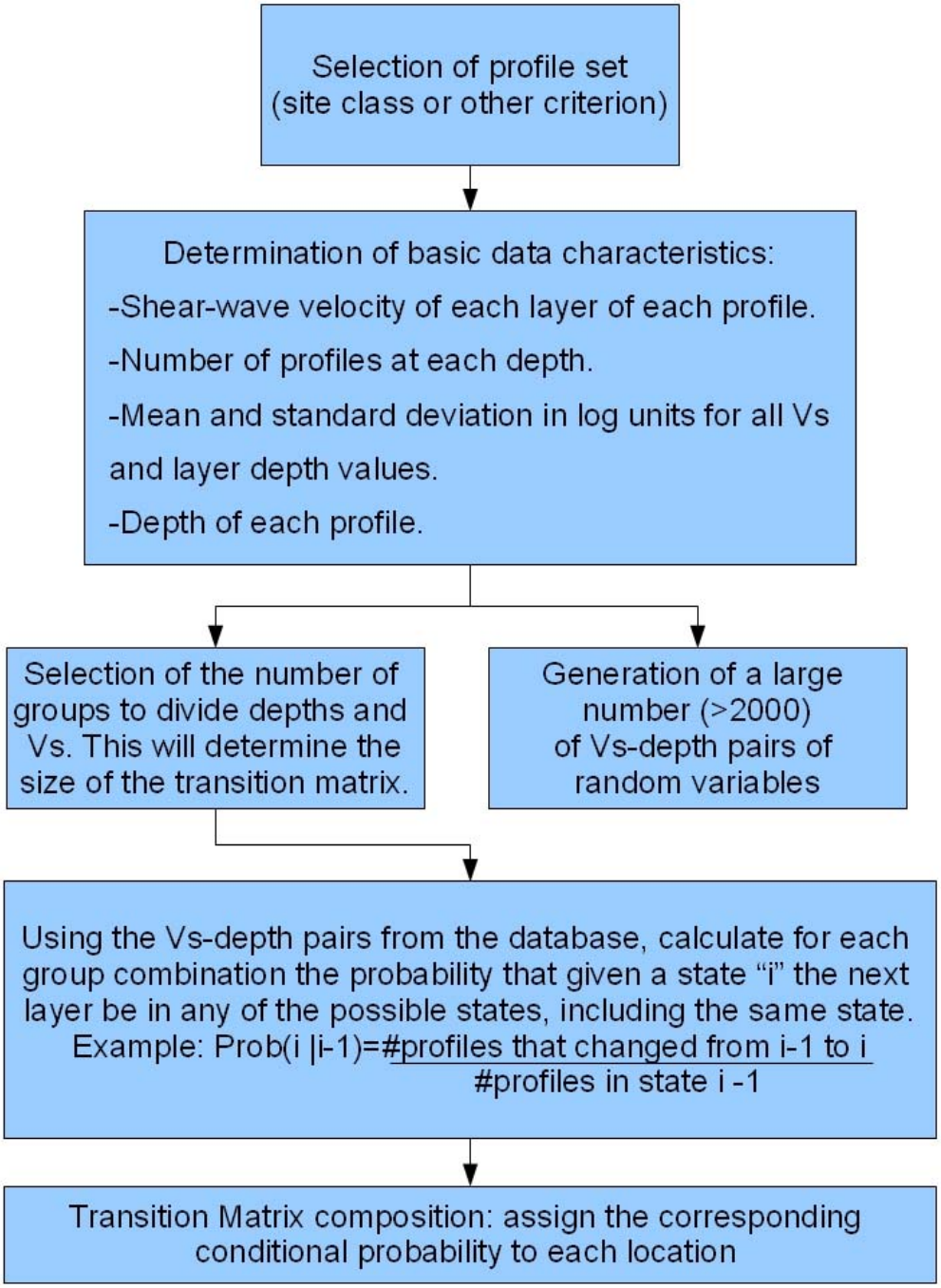
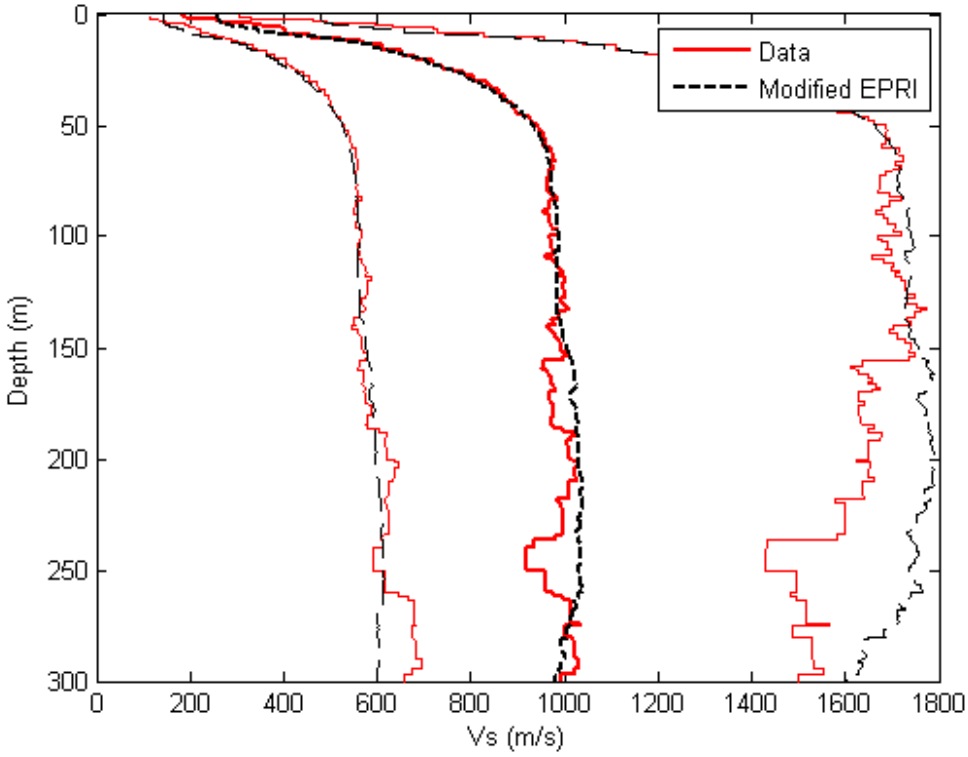
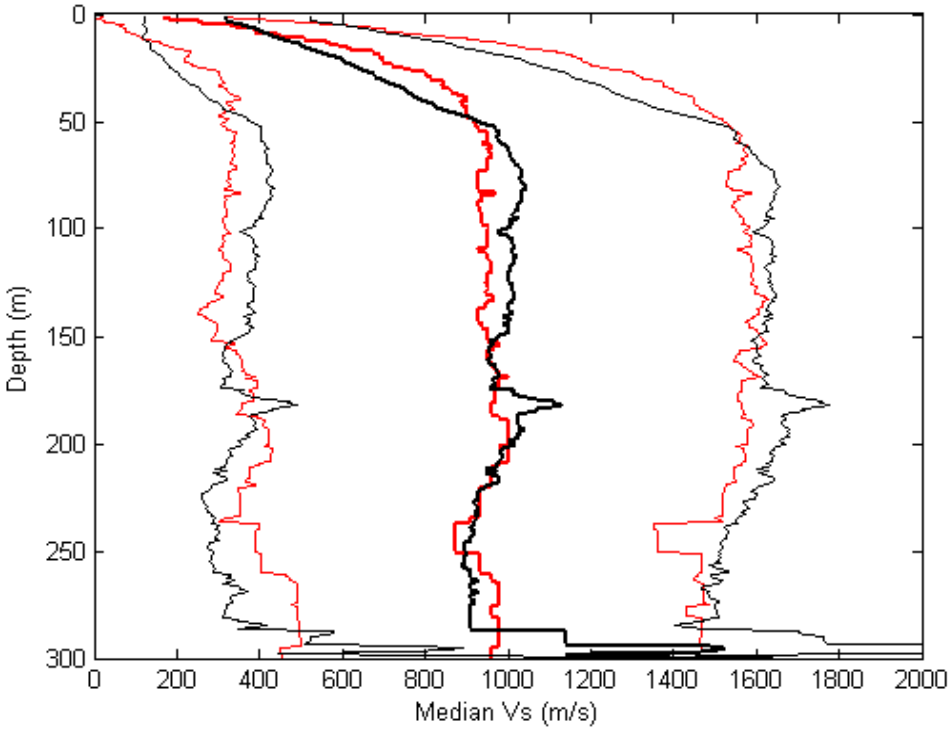


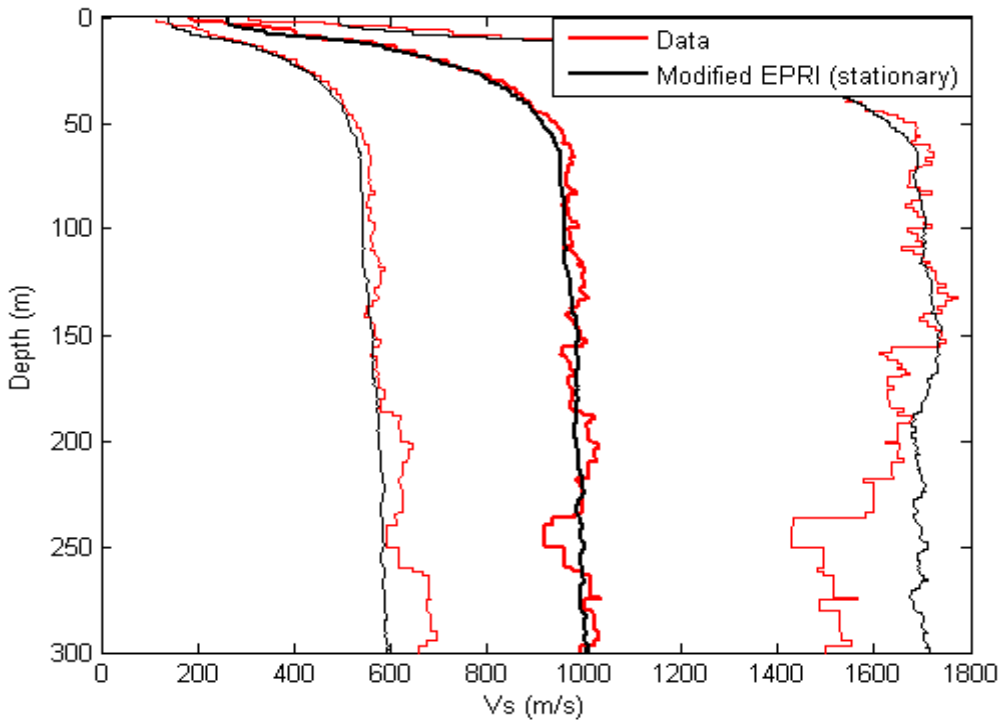
Figure 3.7: Markov chain flow chart.



a) Modified EPRI Model

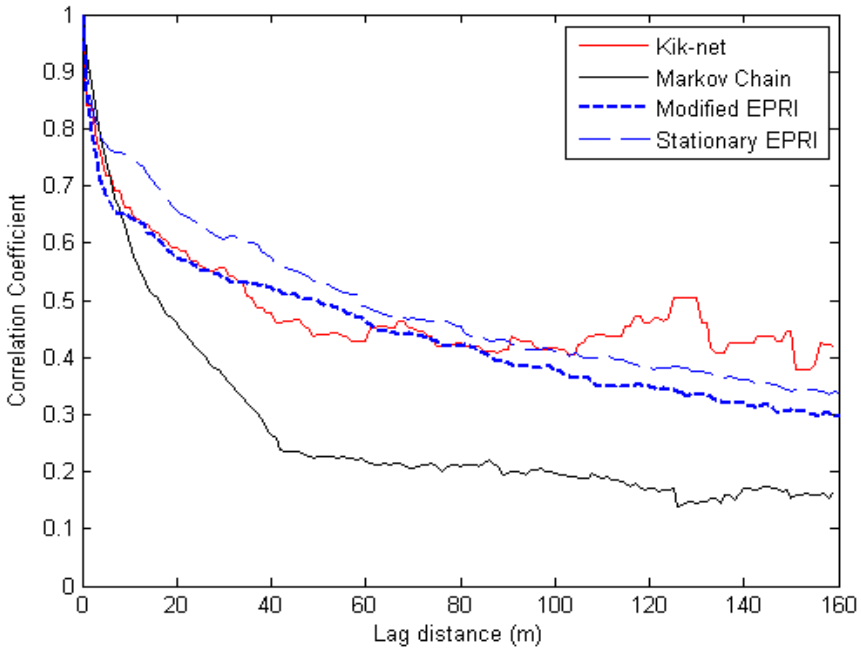


b) Markov Chain Model

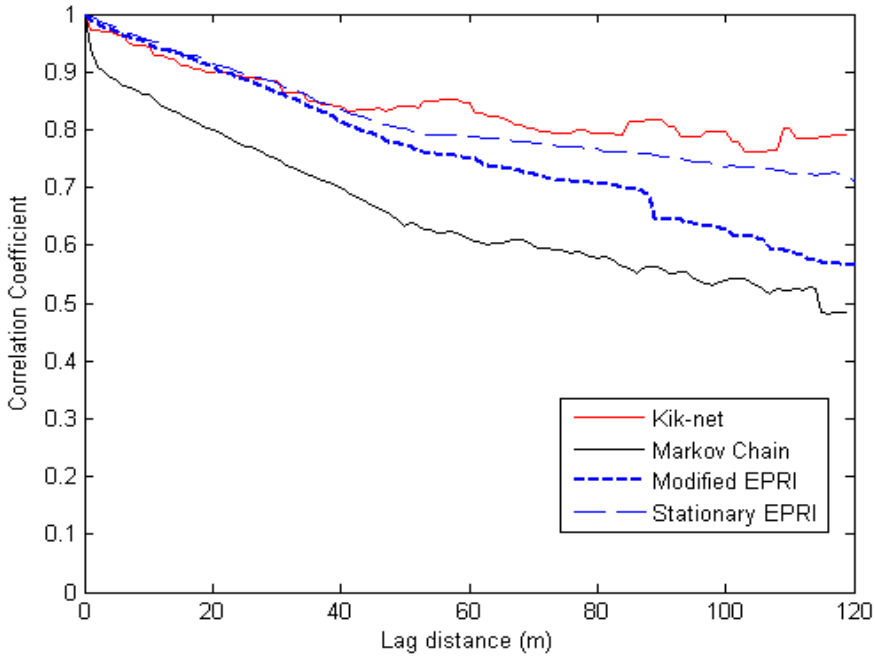


c) Stationary model

Figure 3.8. Measured values of shear wave velocity versus depth for the entire KiK-net database. Also shown are the measured values corresponding to an artificially generated set of profiles using the various proposed models: a) Modified EPRI, b) Markov Chain, and c) Stationary model.



a) Depth = 10m



b) Depth = 50 m.

Figure 3.9. Empirical correlation functions for a) depth =10 m and b) depth = 50 m for the KiK-net database. Also shown are empirical correlation functions obtained from artificially generated dataset using the various models proposed in this study.

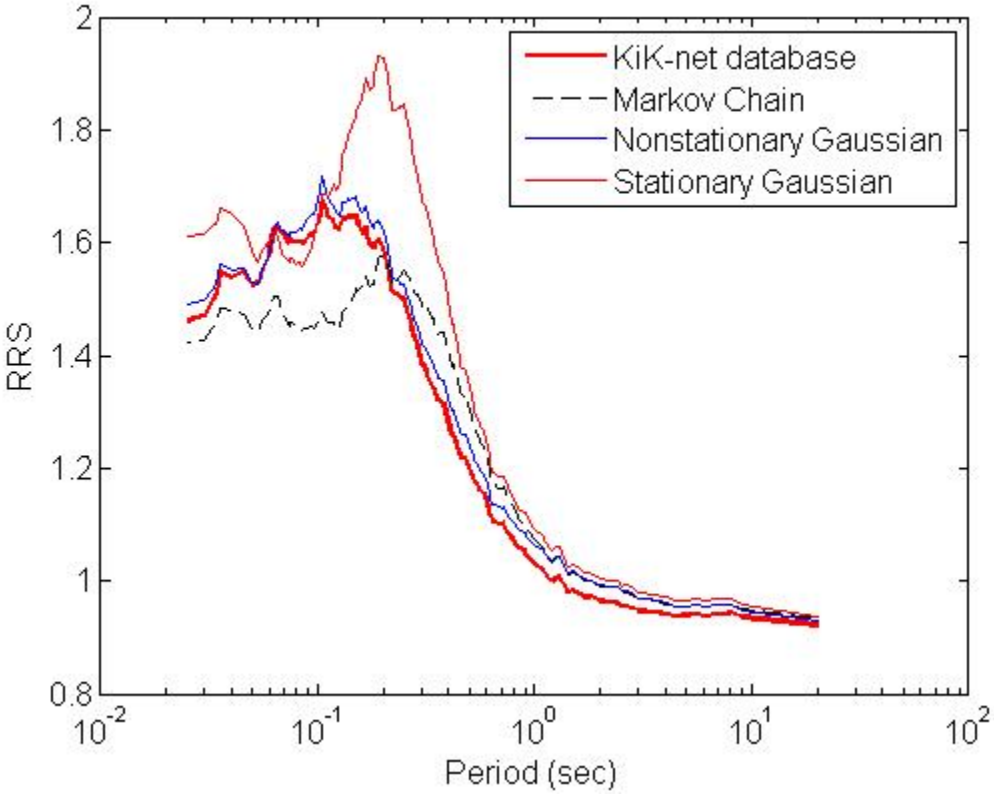


Figure 3.10. Estimated value of the amplification function (or Ratio of Response Spectra) from computed spectral acceleration to input (outcrop) spectral accelerations. The values shown correspond to the median value obtained for measured KiKnet velocity profiles, and for artificial set of profiles generated using the three indicated methodologies.

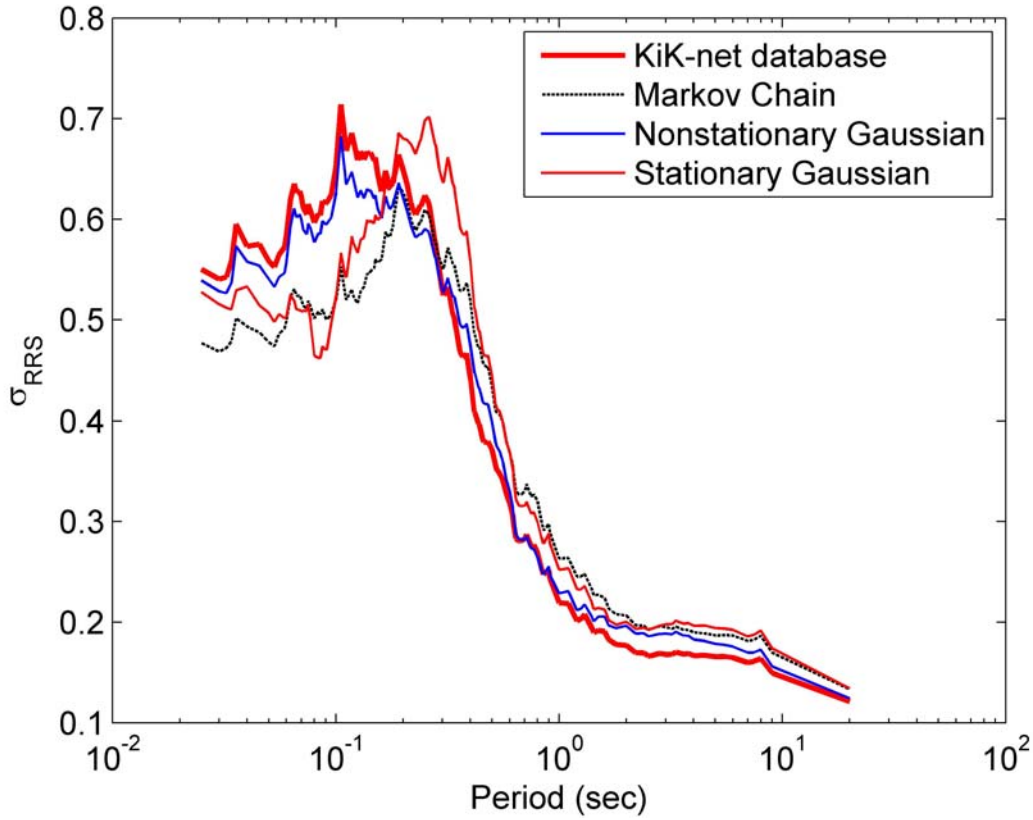
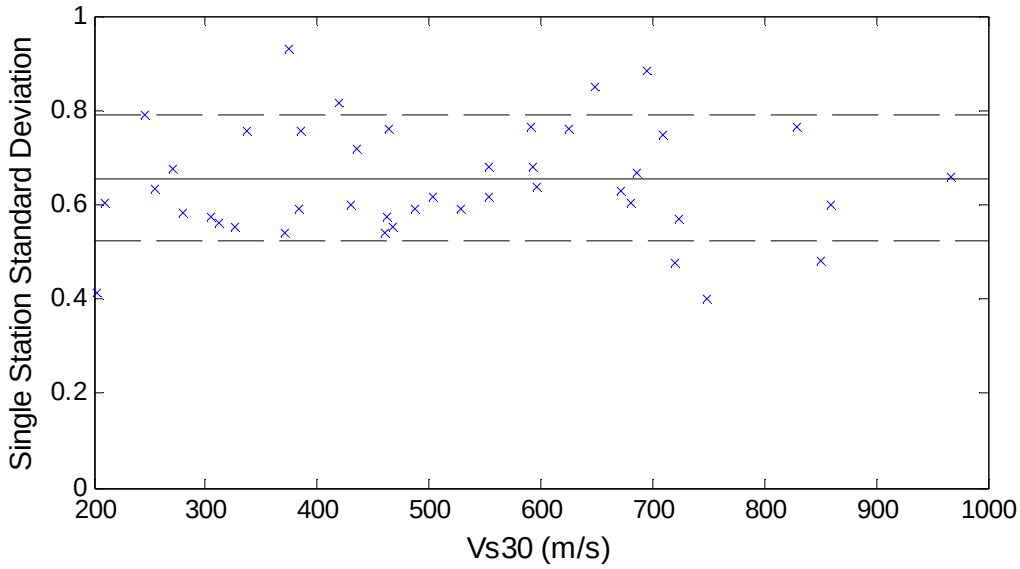
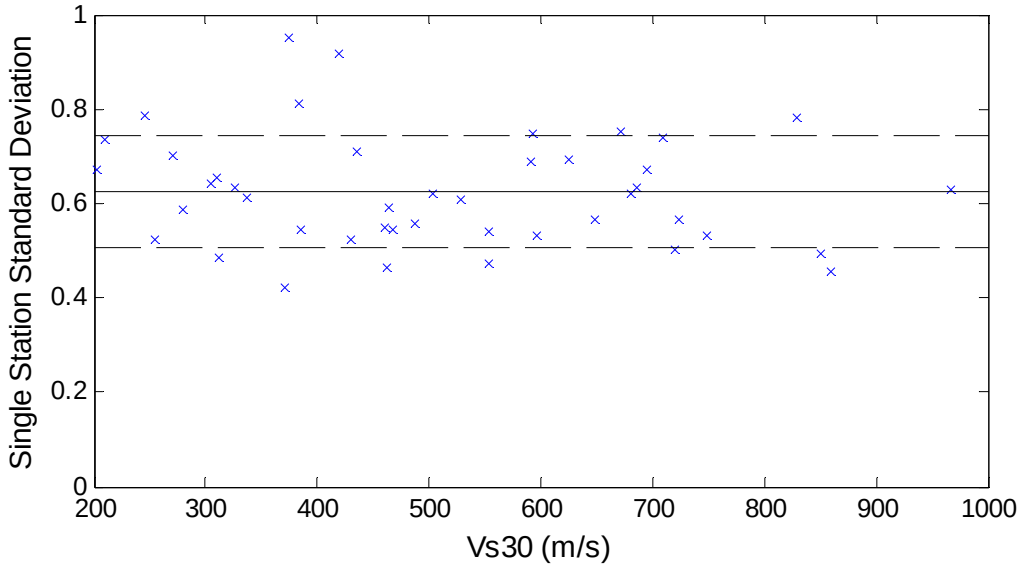


Figure 3.11. Estimated value of the standard deviation of the amplification function (or Ratio of Response Spectra) from computed spectral acceleration to input (outcrop) spectral accelerations.

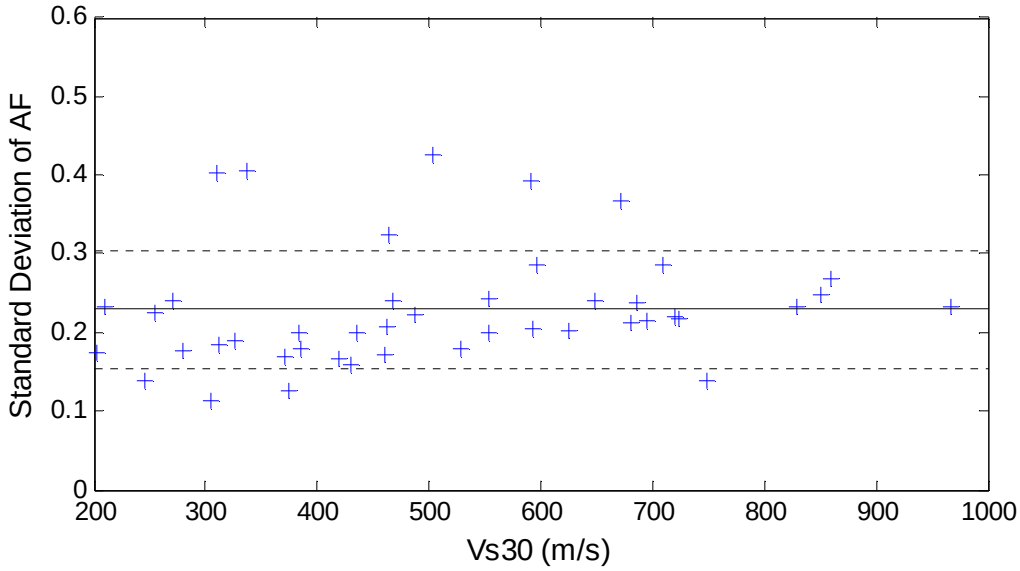


a) T = 0.3 s

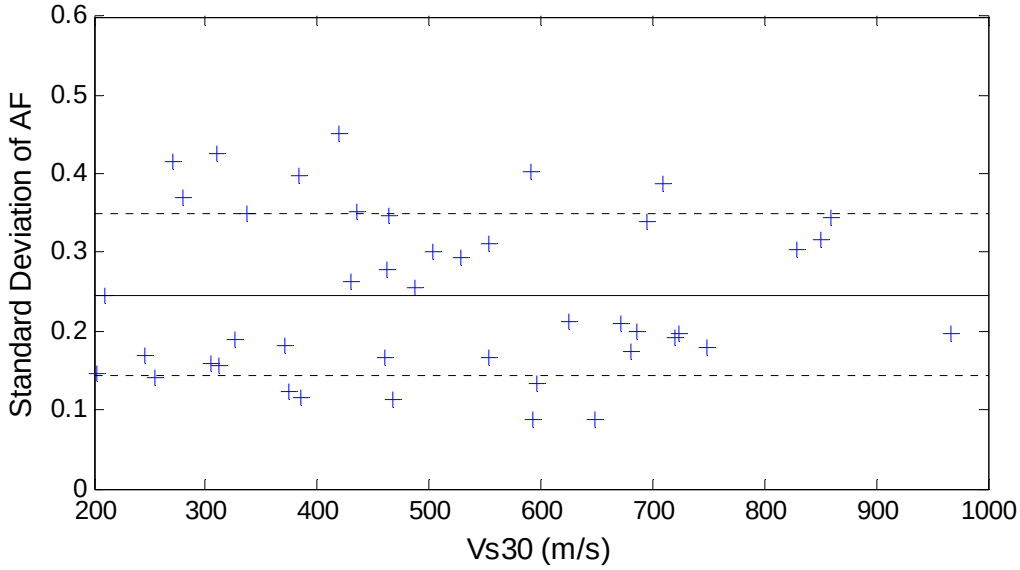


b) T = 1.0 s

Figure 4.1. Single station standard deviations at 44 stations from the KiK-net database (stations with 15 or more recordings), plotted as a function of the measured Vs30 value of each station. The solid line is the mean value of single-station standard deviation for all 44 stations, the dash lines are a one standard deviation band of the mean single-station standard deviation estimate.



a) T = 0.3 s



b) T = 1.0 s

Figure 4.2. Values of the standard deviation of the empirical amplification factor for 44 stations from the KiK-net database (stations with 15 or more recordings), plotted as a function of the measured Vs30 value of each station. The solid line is the mean estimate and the dashed lines are a one-standard deviation from the mean estiamate.

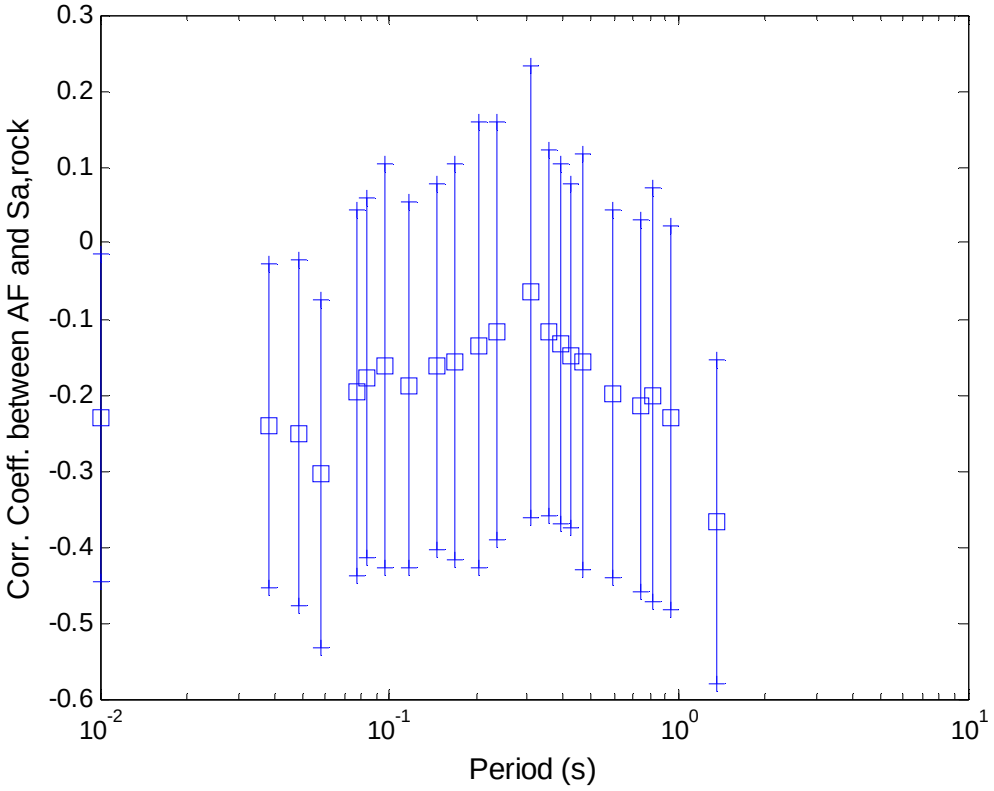
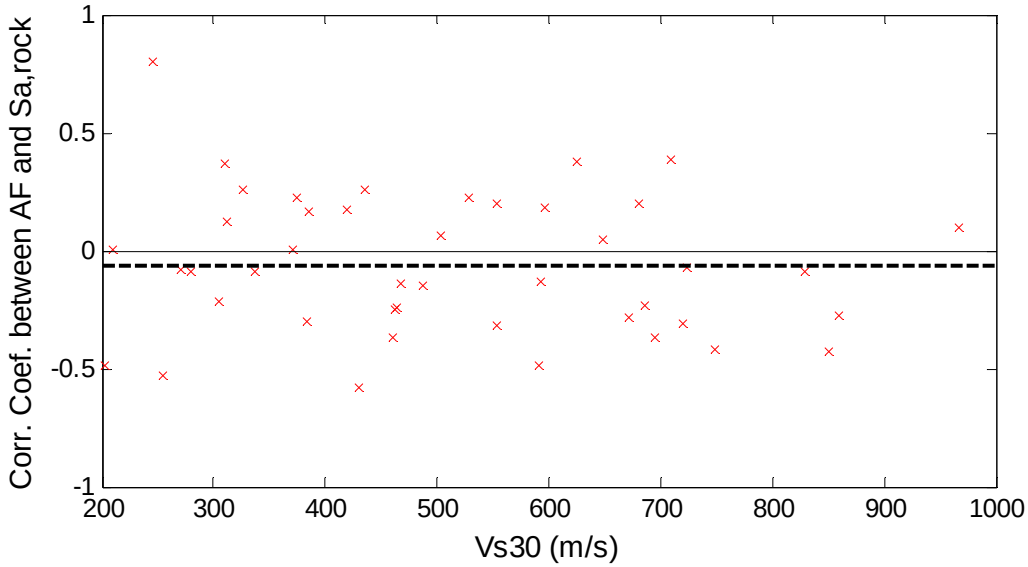
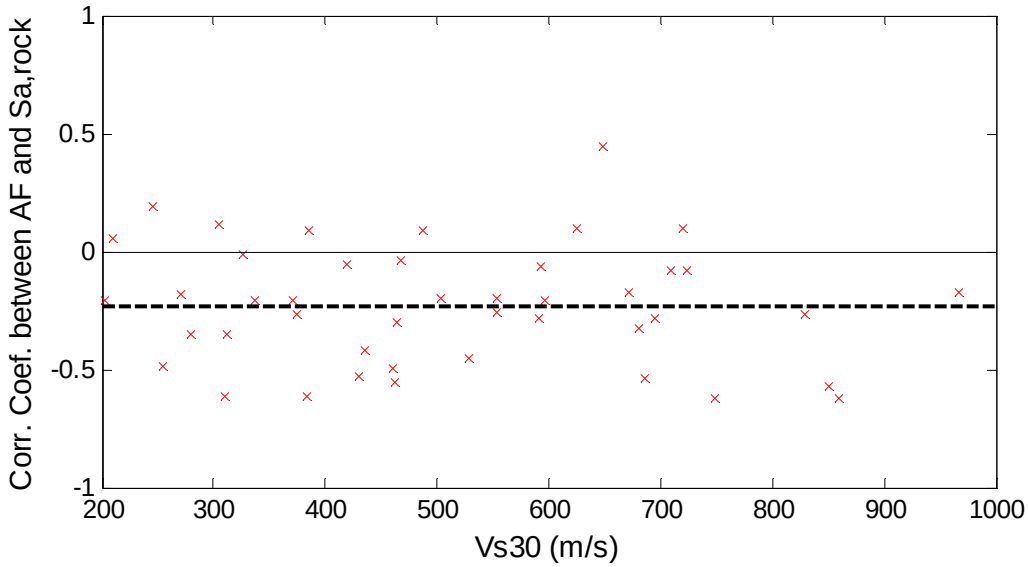


Figure 4.3. Average value of the correlation coefficient between the residuals of the amplification factor (AF) and the residuals of the standard amplification at the borehole. The average is obtained from 44 stations in the KiK-net network that recorded 15 or more earthquakes.



a) T = 0.3 s



b) T = 1.0 s

Figure 4.4. Correlation coefficient between the residuals of the amplification factor (AF) and the residuals of the borehole spectral acceleration at selected periods for the 44 KiKnet stations that recorded 15 or more earthquakes.

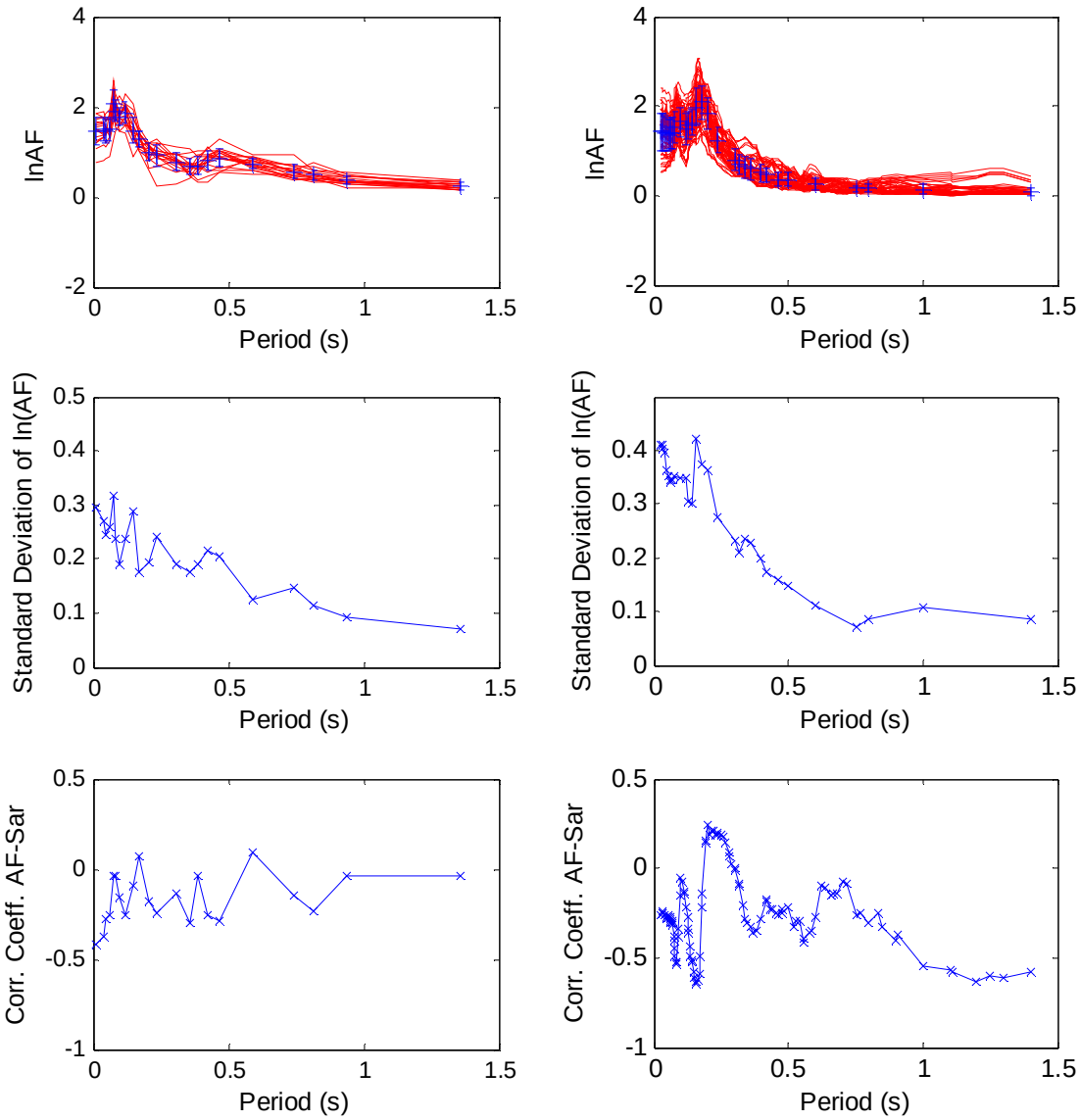


Figure 4.5. Comparison between data estimated from KiK-net recordings (left) and data estimated from site response analyses (right) for the TCGH09 station in the KiK-net array. The comparison shown is for the logarithm of the amplification factor (top), its standard deviation (middle), and the correlation coefficient between the amplification factor and the bedrock ground motion (bottom).

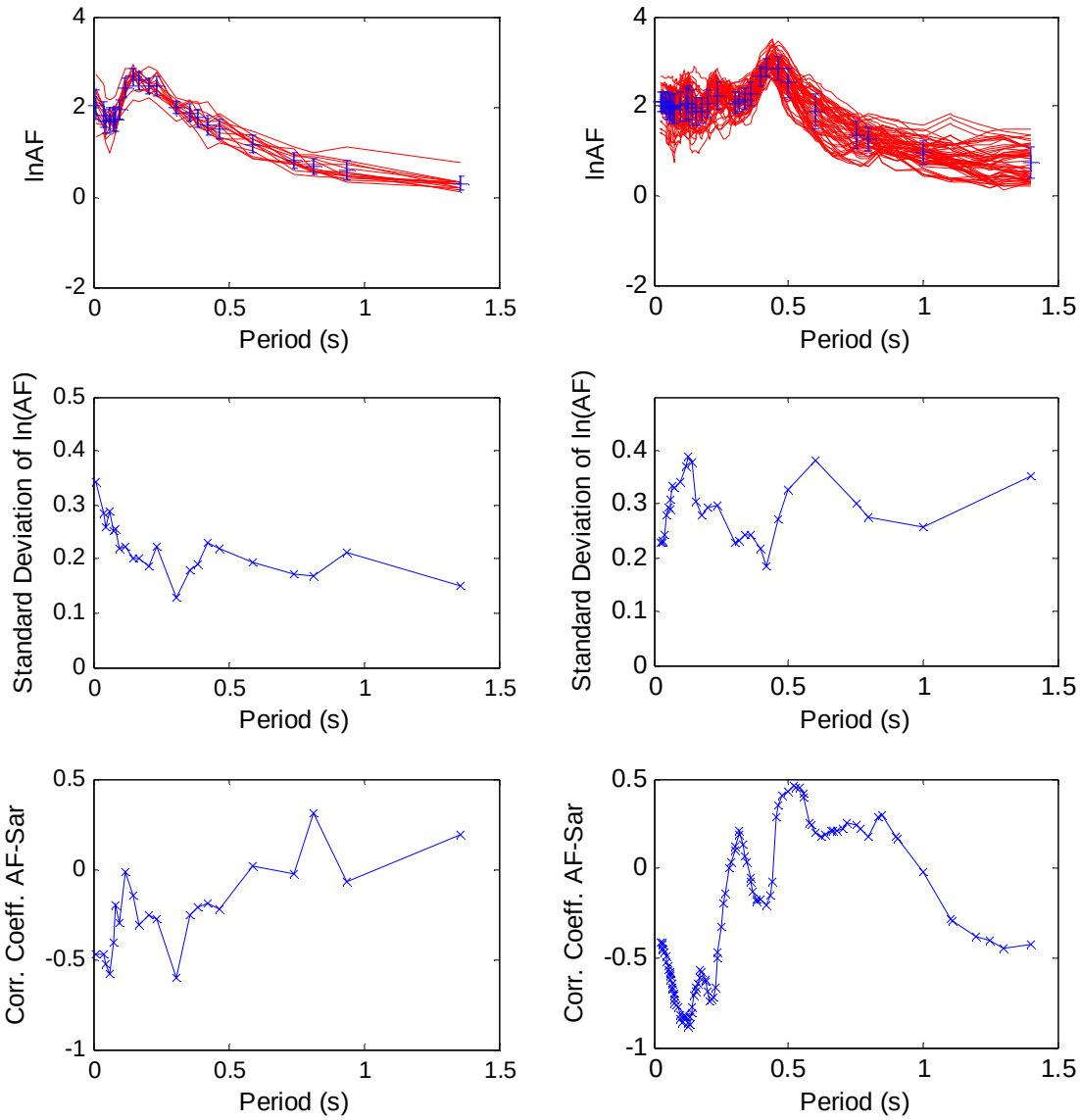


Figure 4.6. Comparison between data estimated from KiK-net recordings (left) and data estimated from site response analyses (right) for the TCGH11 station in the KiK-net array. The comparison shown is for the logarithm of the amplification factor (top), its standard deviation (middle), and the correlation coefficient between the amplification factor and the bedrock ground motion (bottom).

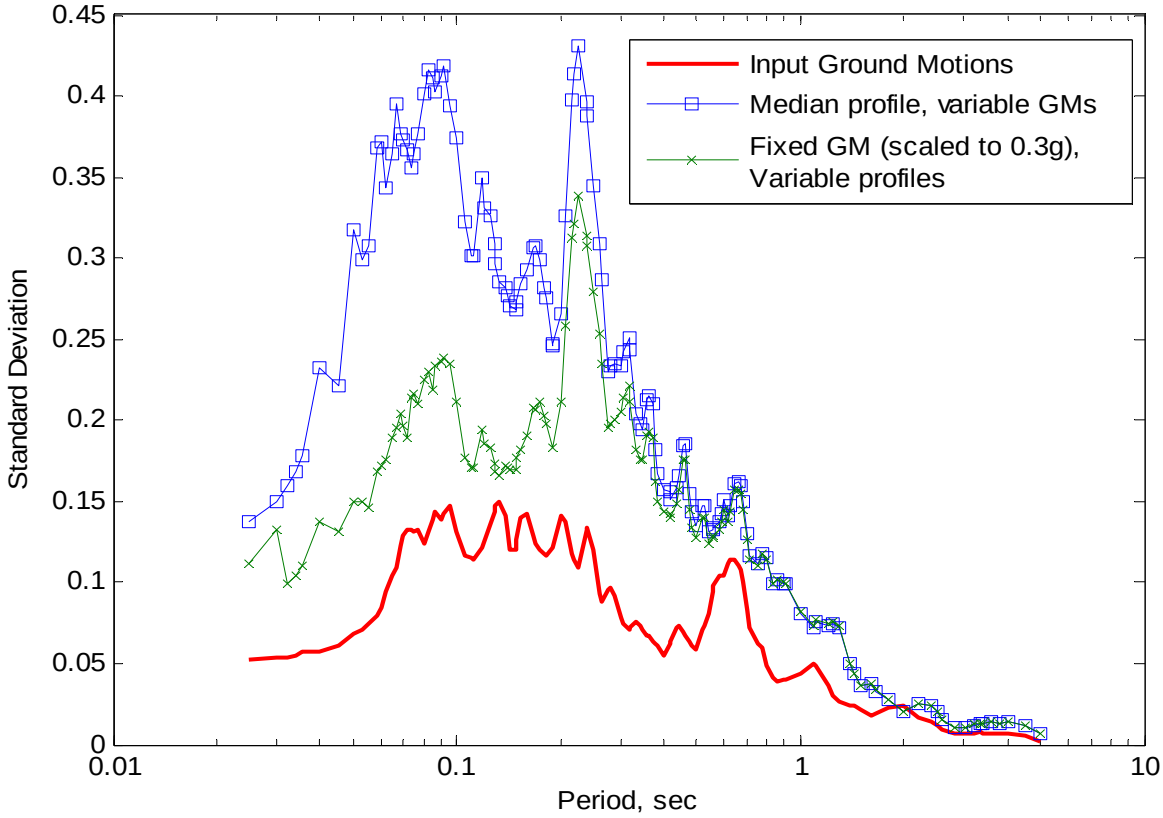
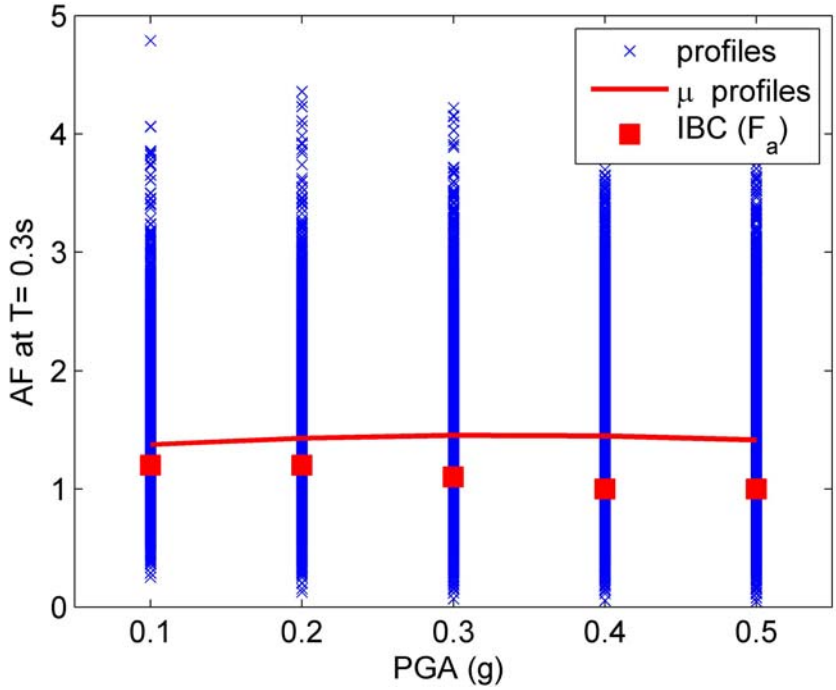
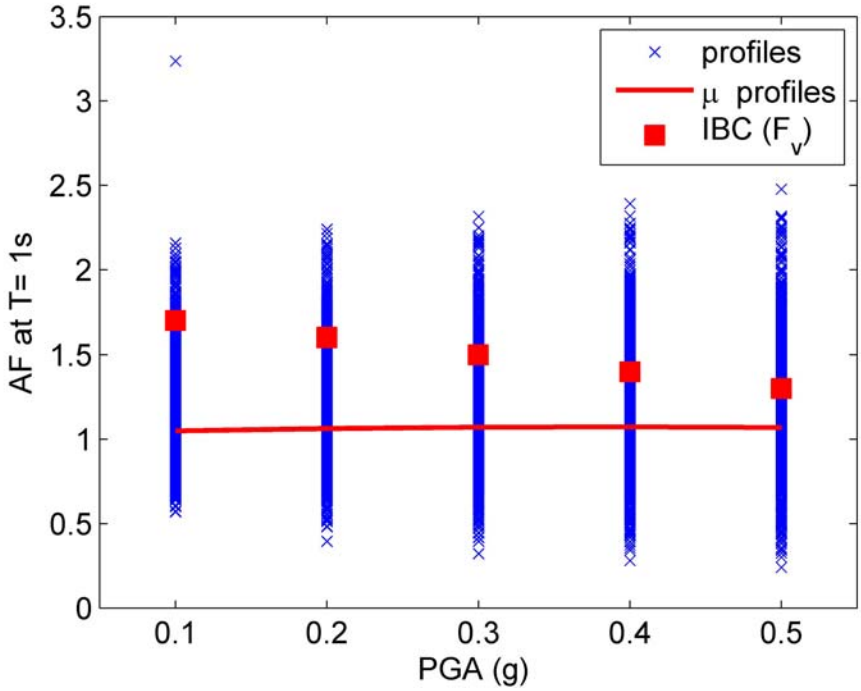


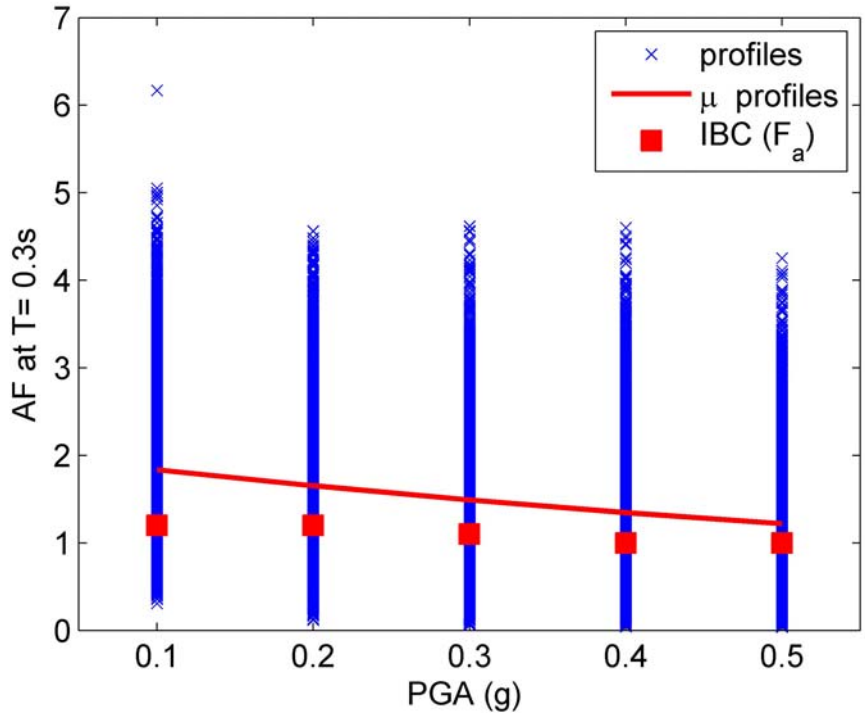
Figure 4.7. Standard deviation of output ground motions computed for a site that recorded the 2001 Southern Peru earthquake. The standard deviation corresponds to the sample standard deviation at each period using 30 simulated ground motions and 500 randomly generated profiles (from Rodriguez-Marek et al. 2010).



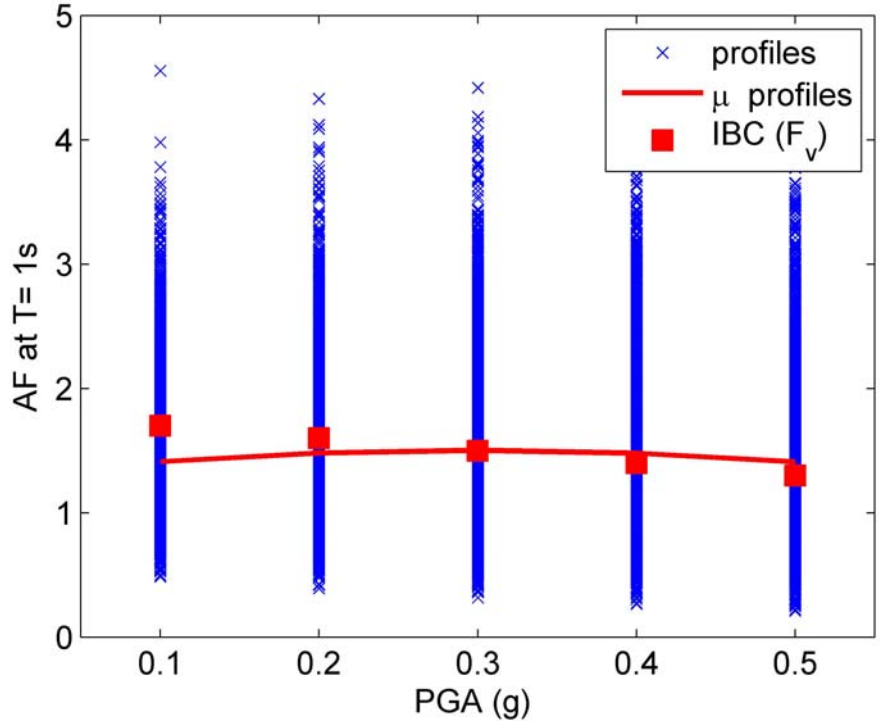
a) Site C, T = 0.3 s



b) Site C, T = 1.0 s

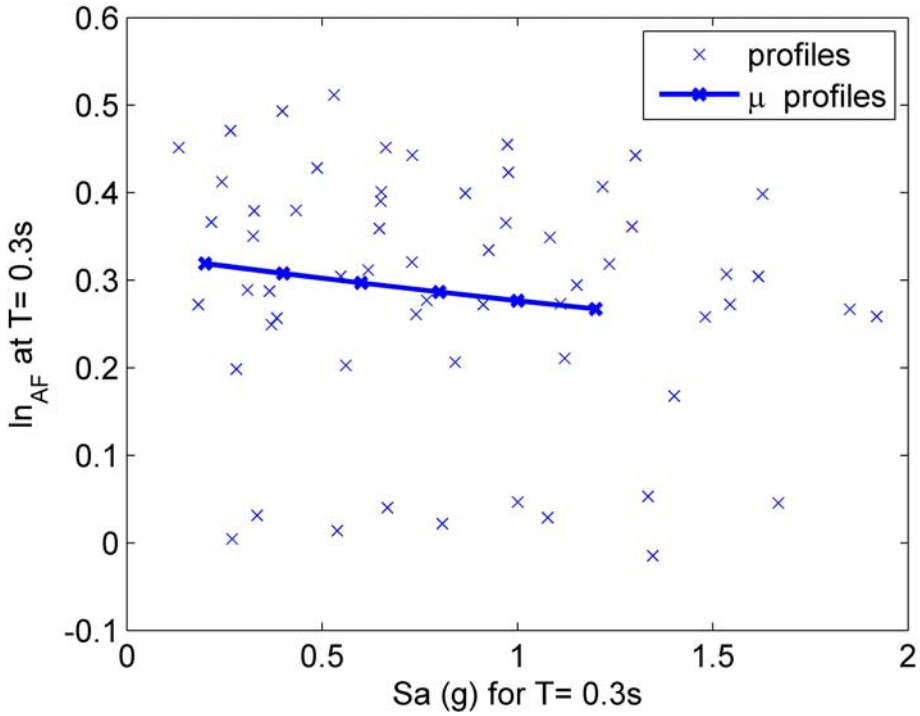


c) T=0.3, Site D

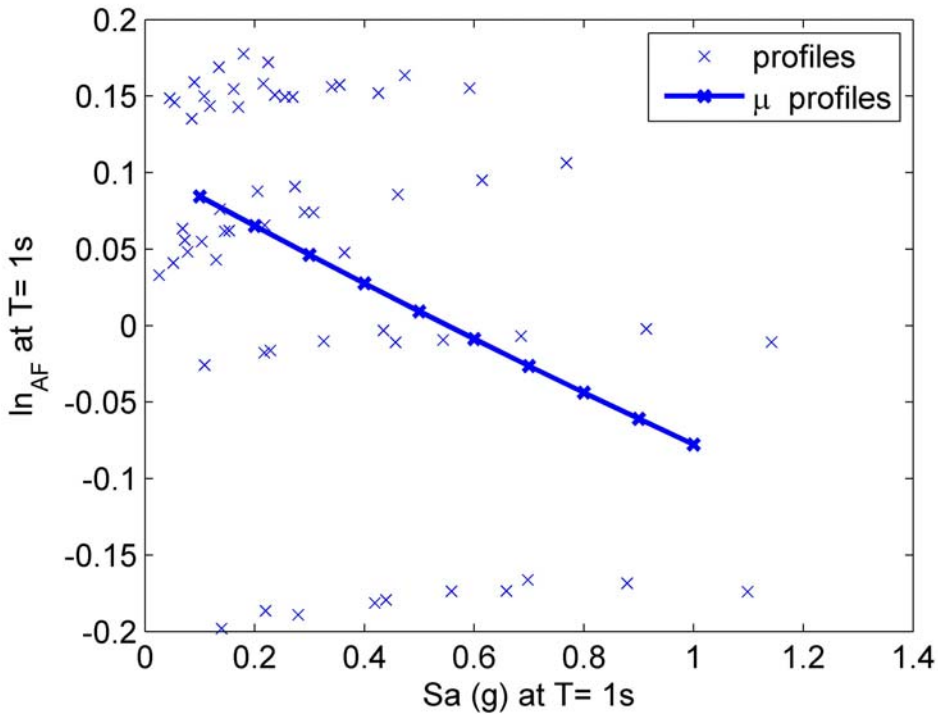


d) T = 1.0 s, Site D

Figure 4.8. Comparison of IBC code site factors for IBC sites C and D with amplification factors estimated in this study. Factors are estimated using 12 ground motion and 600 sites for each of the two site categories.



a) T = 0.3 s



b) T = 1.0 s

Figure 4.9. Amplification factor for simulated C sites plotted as a function of input spectral acceleration.

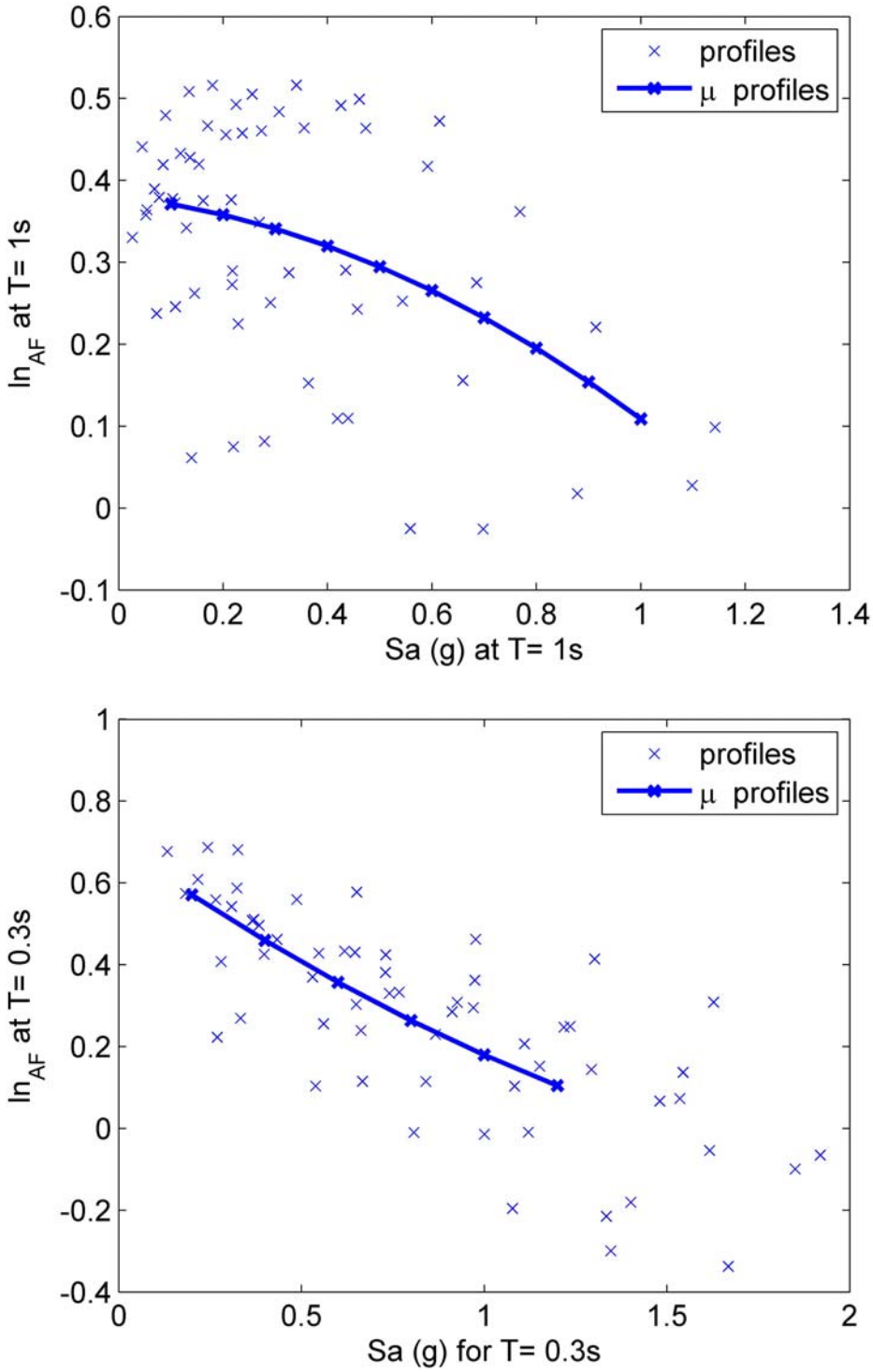
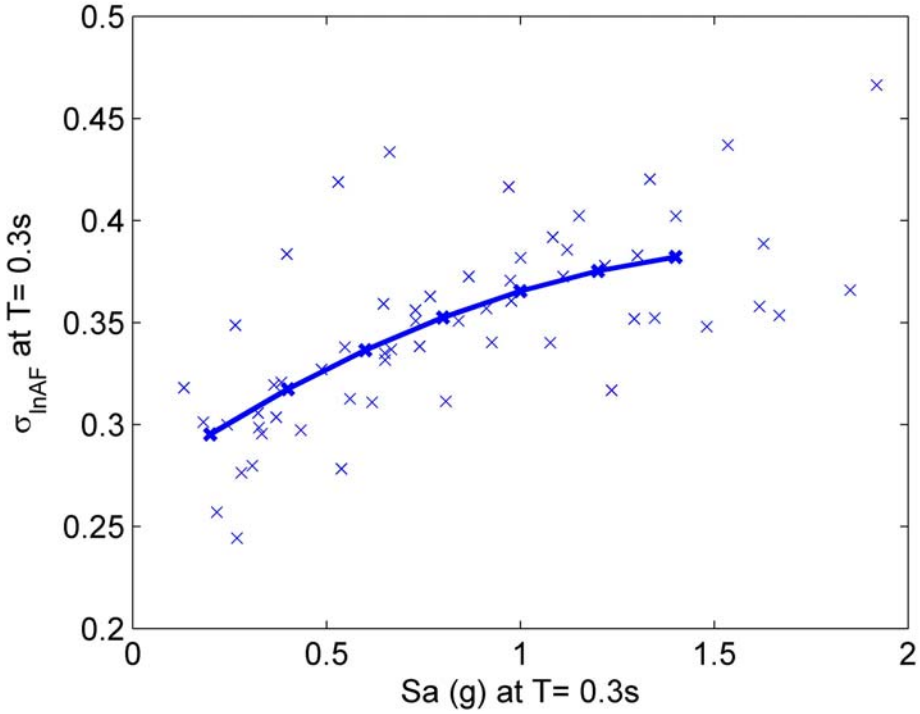
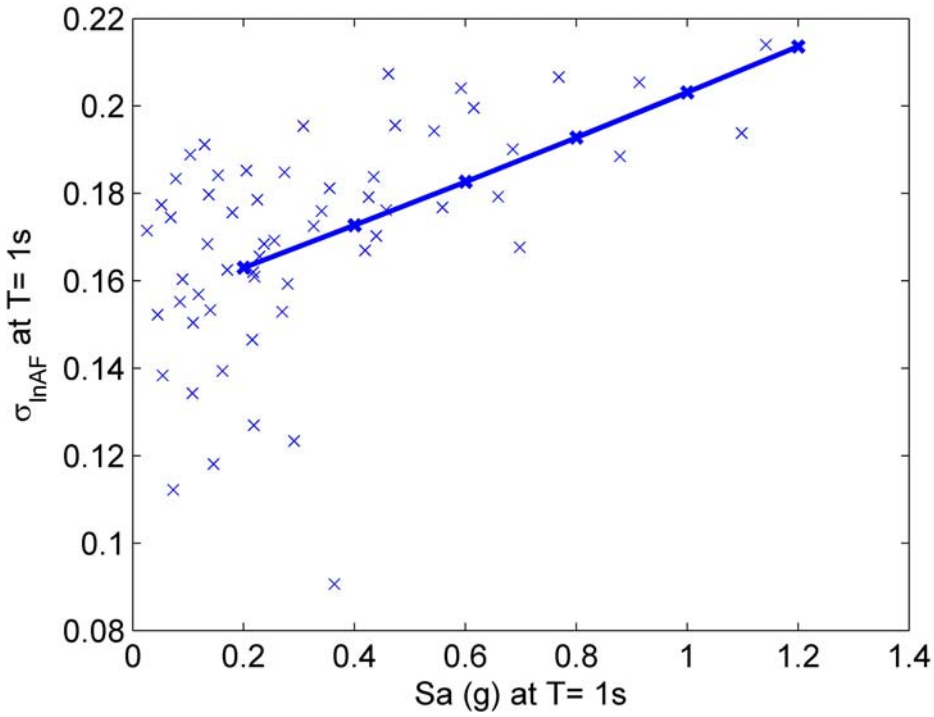


Figure 4.10. Amplification factor for simulated D sites plotted as a function of input spectral acceleration.

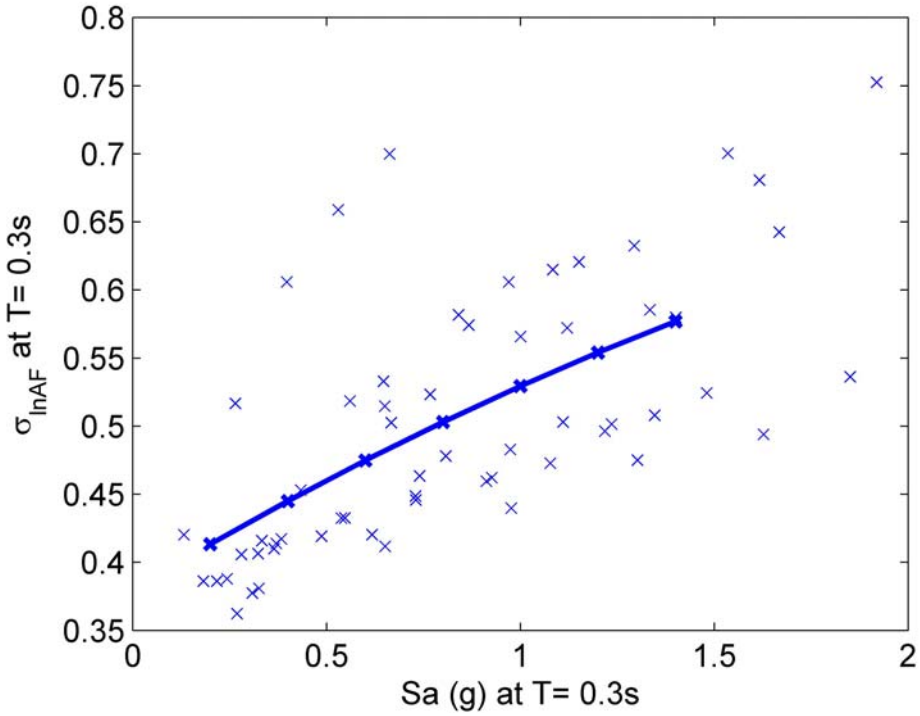


a) T -= 0.3 s

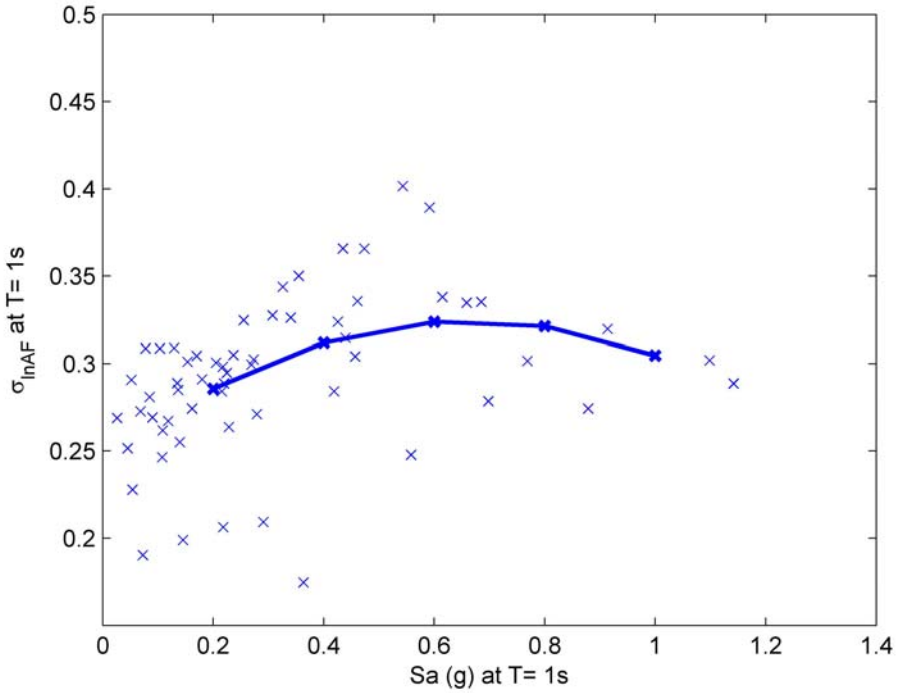


b) T = 1 s

Figure 4.11. Standard deviations of the amplification factor for Site C each. Each point is the standard deviation of the amplification factor for 600 artificially generated profiles. The 60 points correspond to 12 ground motions with 5 different linear scaling factors.

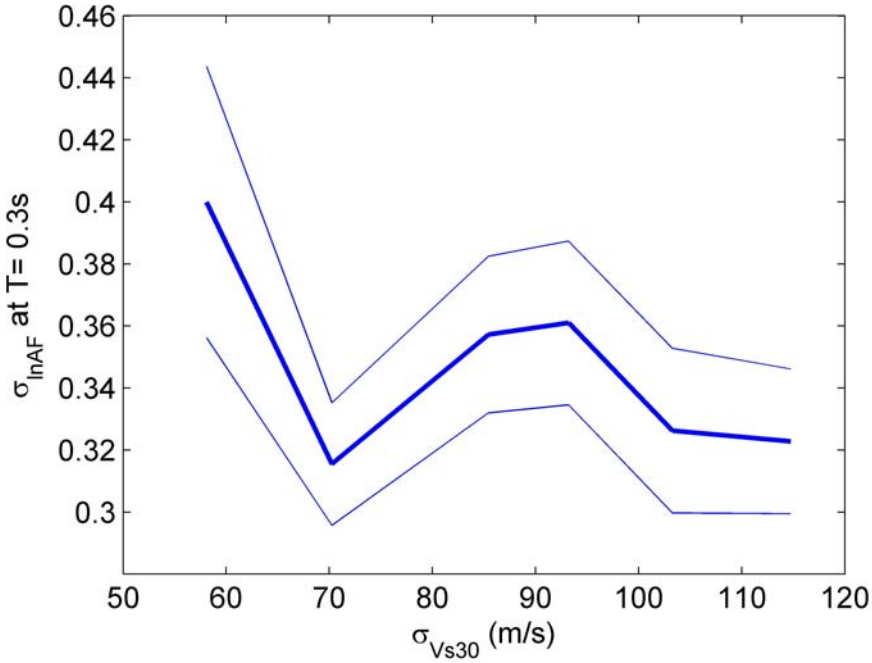


a) T = 0.3 s.

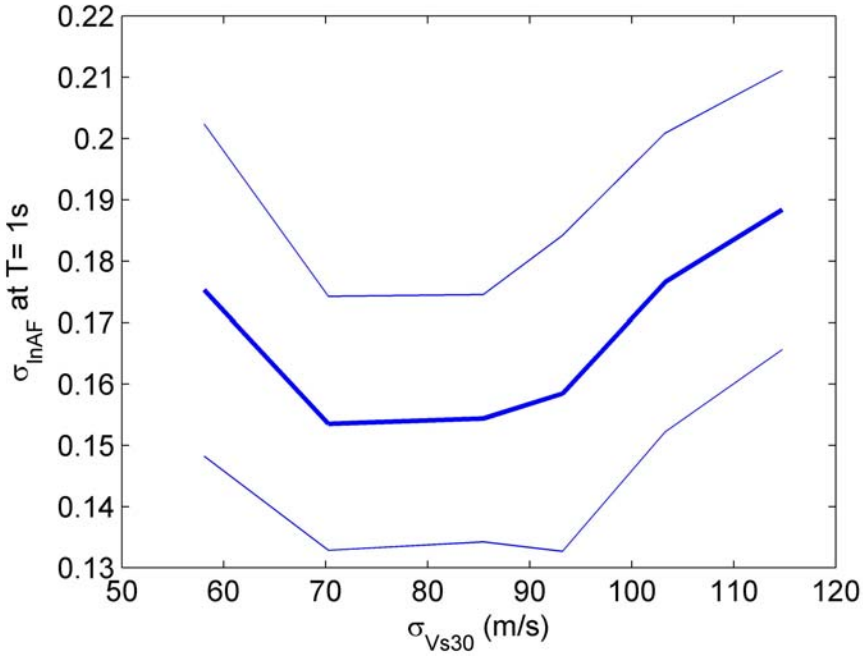


b) T = 1.0 s.

Figure 4.12. Standard deviations of the amplification factor for Site D each. Each point is the standard deviation of the amplification factor for 600 artificially generated profiles. The 60 points correspond to 12 ground motions with 5 different linear scaling factors.

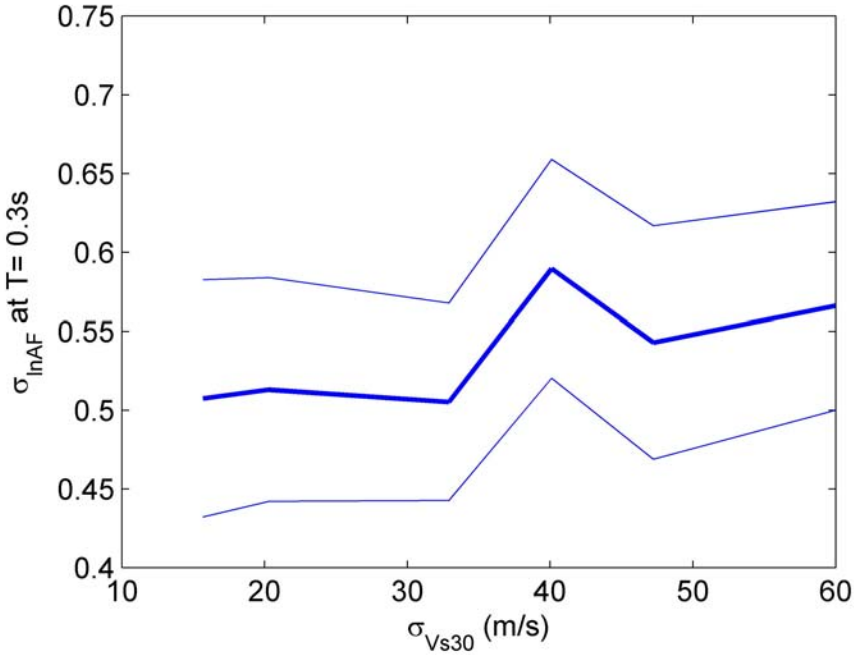


a) T = 0.3s

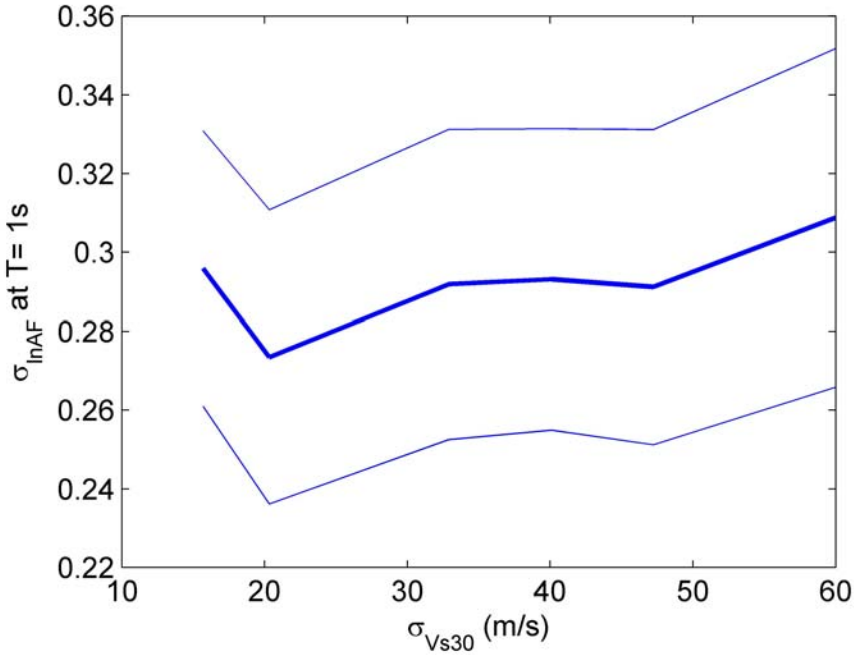


b) T = 1 s

Figure 4.13. Standard deviations of the amplification factor for Site C as a function of the standard deviation of Vs30. 600 profiles were used for each Vs30 set. The thick line corresponds to the median estimate for 60 ground motions (12 ground motions scaled to 5 different PGAs). The thin lines correspond to the sample standard deviation for the 60 ground motions.



a) T=0.3



b) T =- 1.0

Figure 4.14. Standard deviations of the amplification factor for Site D as a function of the standard deviation of Vs30. 600 profiles were used for each Vs30 set. The thick line corresponds to the median estimate for 60 ground motions (12 ground motions scaled to 5 different PGAs). The thin lines correspond to the sample standard deviation for the 60 ground motions.

Table 2.1a. Parameters of the GMPE (Equation 2.5) for selected periods. In addition to the paramters listed below, the following parameters are constant for all periods: Rref = 1.0, Mref = 4.5, Vref = 760, Vsholeref = 3000, and e7 = 0.

Period (s)	c_1	c_2	c_3	e_1	e_5	e_6	b_{lin}	a_{100}	a_{200}	b_{100}	b_{200}	c_{100}	c_{200}	M_h	h
0.01	-1.434	0.301	-0.009	1.325	0.455	-0.231	-0.292	-1.560	-1.756	0.001	-0.137	-0.314	-0.410	5.6	1.36
0.097	-1.675	0.041	-0.003	4.566	2.473	0.274	-0.198	-1.765	-1.957	0.070	-0.161	-0.328	-0.414	5.6	1.41
0.309	-1.624	0.124	-0.002	2.785	2.184	-0.028	-0.998	-1.075	-1.324	-0.053	-0.003	-0.446	-0.731	5.6	0.57
0.469	-1.535	0.157	-0.002	1.875	1.091	-0.425	-1.027	-0.748	-1.130	-0.136	-0.028	-0.328	-0.756	6	0.13
0.7456	-1.438	0.159	-0.002	1.018	1.347	-0.414	-0.974	-0.553	-0.798	-0.105	0.043	-0.375	-0.650	6	0.13
0.9401	-1.414	0.169	-0.002	0.559	1.364	-0.441	-0.971	-0.499	-0.666	0.003	0.052	-0.502	-0.601	6	0.66
1.3622	-1.389	0.180	-0.002	-0.058	1.634	-0.344	-0.908	-0.573	-0.596	-0.134	0.010	-0.606	-0.663	6	3.97

Table 2.1b. Standard deviations for the GMPE in Equation 2.5.

Period (s)	σ_{tot} **(Surf)**	σ_{tot} **(BH)**
0.01	0.816	0.719
0.097	0.924	0.773
0.309	0.851	0.766
0.469	0.829	0.756
0.7456	0.814	0.737
0.9401	0.804	0.728
1.3622	0.808	0.746

.

Table 2.2. Single-Station Standard deviations for the KiK-net ground motion data set.

	Surface		Borehole	
Period	**σ_{ss} (Mean Value)**	**σ_{ss} (Standard Deviation)**	**σ_{ss} (Mean Value)**	**σ_{ss} (Standard Deviation)**
0.01	0.6341	0.1165	0.6268	0.126
0.097	0.6599	0.1657	0.6389	0.1407
0.309	0.6562	0.1341	0.629	0.1237
0.469	0.6239	0.1248	0.6183	0.1212
0.7456	0.6256	0.1164	0.6328	0.1276
0.9401	0.6245	0.1195	0.6366	0.1246
1.3622	0.615	0.1117	0.6561	0.1156

Table 3.1: Parameters for non-stationary correlation function for the Modified EPRI model.

	ρ_o	z_o	b	κ_o	κ_{max}	z_{max}
Generic	0.5532	-5.0031	5.4041	-19.0000	83.9813	35.0000
Sites A+B	0.8860	1.6252	9.5663	3.1600	5.2920	43.0000
Site C	0.5756	7.3486	4.2589	-8.1000	77.5863	50.0000
SiteD	0.8143	38.0000	12.3571	-105.0000	58.4465	16.6

Table 3.2: Statistics for a lognormal distribution for the depth to the bottom of the profile.

	Mean (log-depth)	**Standard Deviation**
Generic	4.34	0.66
Sites A+B	4.2	0.67
Site C	4.27	0.62
Site D	4.53	0.68

Table 3.3: Coefficients for layering model, 95% confidence interval in parenthesis.

	a	***b***	***c***
Generic	2.2 (-2.2; 6.6)	8.8 (-0.9; 18.6)	1.1 (0.67; 1.6)
Sites A+B	0.86 (-1; 2.8)	3.6 (-4.4; 11.5)	0.94 (0.38; 1.5)
Site C	4.0 (-7.7; 15.7)	10.43 (-2.6; 23.5)	1.3 (0.62; 1.9)
SiteD	2.6 (-4.4; 9.7)	12.95 (-4; 29.9)	1.1(0.55; 1.7)

Table 3.4: Correlation Coefficient for Stationary Model.

Category	ρ **(**ΔL**=1)**
Generic	0.82
Sites A+B	0.72
Site C	0.85
Site D	0.66

Table 4.1. Intra-Method Variability for Vs30 Measurement

Method	**Coefficient of Variation**
MASW	1-4 % [1]
SASW	5-10 %[2]; 6 % [3]
P-S logging and SCPT	1-3 % [1;4]
Geologic Estimates	20-35 % [5]

[1] Asten and Boore (2005)

[2] Marosi and Hiltunen (2004)

[3] Martin and Diehl (2004)

[4] Brown et al. (2002)

[5] Moss (2008)