Bulletin of the Seismological Society of America, Vol. 100, No. 6, pp. 3295–3303, December 2010, doi: 10.1785/0120090366

Short Note

Considering Spatial Correlation in Mixed-Effects Regression and the Impact on Ground-Motion Models

by Nirmal Jayaram and Jack W. Baker

Abstract Ground-motion models are commonly used in earthquake engineering to predict the probability distribution of the ground-motion intensity at a given site due to a particular earthquake event. These models are often built using regression on observed ground-motion intensities and are fitted using either the one-stage mixed-effects regression algorithm proposed by Abrahamson and Youngs (1992) or the two-stage algorithm of Joyner and Boore (1993). In their current forms, these algorithms ignore the spatial correlation between intraevent residuals. This paper emphasizes the theoretical importance of considering spatial correlation while fitting ground-motion models and proposes an extension to the Abrahamson and Youngs (1992) algorithm that allows the consideration of spatial correlation.

By refitting the Campbell and Bozorgnia (2008) ground-motion model using the mixed-effects regression algorithm considering spatial correlation, it is apparent that the variance of the total residuals and the ground-motion model coefficients used for predicting the median ground-motion intensity are not significantly different from the published values even after the incorporation of spatial correlation. However, there is an increase in the variance of the intraevent residual and a significant decrease in the variance of the interevent residual. These changes have implications for risk assessments of spatially-distributed systems because a smaller interevent residual variance implies lesser likelihood of observing large ground-motion intensities at all sites in a region.

Introduction

Ground-motion models are commonly used in earthquake engineering to predict the probability distribution of the ground-motion intensity at a given site due to a particular earthquake event. Typically, a ground-motion model takes the following form:

$$\ln(Y_{ij}) = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij} + \eta_i, \qquad (1)$$

where Y_{ij} denotes the ground-motion intensity parameter of interest (e.g., $S_a(T)$, the spectral acceleration at period T) at site j during earthquake i; $f(\boldsymbol{P}_{ij}, \boldsymbol{\theta})$ denotes the ground-motion prediction function with predictive parameters $\boldsymbol{P}_{ij}$ (e.g., magnitude, distance of source from site, site condition) and coefficient set $\boldsymbol{\theta}$; ε_{ij} denotes the intraevent residual, which is a zero-mean random variable with standard deviation σ_{ij}; η_i denotes the interevent residual, which is a random variable with zero mean and standard deviation τ_{ij}. The rest of this paper assumes for simplicity that the residuals have a constant σ (i.e., $\sigma_{ij} = \sigma$) and τ (i.e., $\tau_{ij} = \tau$) for any given ground-motion intensity parameter; that is, the residuals are homoscedastic. This assumption is not true in some modern models (e.g., Abrahamson and Silva, 2008), in which case, the concepts remain the same, but some of the equations are no longer directly applicable.

Ground-motion models are primarily fitted using two approaches: the two-stage regression algorithm of Joyner and Boore (1993) (e.g., Boore and Atkinson, 2008) and the one-stage mixed-effects model regression algorithm of Abrahamson and Youngs (1992) (e.g., Abrahamson and Silva, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008). Joyner and Boore (1993) provide a detailed comparison of these two algorithms. Both these algorithms, in their current forms, assume that the intraevent residuals are independent of each other. The intraevent residuals, however, are known to be spatially correlated (Boore *et al.*, 2003; Wang and Takada, 2005; Goda and Hong, 2008; Jayaram and Baker, 2009). Recently, Hong *et al.* (2009) investigated the influence of including spatial correlation in the regression analysis on the ground-motion models fitted using the two-stage regression algorithm and a one-stage algorithm of Joyner and Boore (1993). They concluded that the influence of considering spatial correlation on the estimated ground-motion models is negligible based on insignificant changes to the coefficient set $\boldsymbol{\theta}$. Fitting ground-motion models considering correlation does, however, change the variances of the interevent and the intraevent residuals (observed by Hong *et al.*, 2009). This short note provides a theoretical basis for such changes to the variance terms and also discusses the impact of these changes on the estimated seismic risk of spatially-distributed systems. Further, a modified algorithm based on that of Abrahamson and Youngs (1992) is developed that accounts for the spatial correlation in the mixed-effects regression. This modified algorithm is used to refit the Campbell and Bozorgnia (2008) ground-motion model in order to illustrate the impact of incorporating spatial correlation.

Current Regression Algorithm

Ground-motion models were originally treated as fixed-effects models that take the form

$$\ln(Y_{ij}) = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij}^{(t)}, \tag{2}$$

where $\varepsilon_{ij}^{(t)}$ denotes the total residual term at site j during earthquake i (e.g., Joyner and Boore, 1981; Bolt and Abrahamson, 1982).

Brillinger and Preisler (1984a, 1984b) first proposed regressing a ground-motion model as a mixed-effects model. The mixed-effects model differs from the fixed-effects model in its consideration of the error term as being the sum of an intraevent error term and an interevent error term (equation 1). The interevent term helps partially account for the correlation between the ground-motion intensities recorded during any particular earthquake. Brillinger and Preisler (1984b) employed the expectation-maximization algorithm (Dempster *et al.*, 1977) to estimate the ground-motion model parameters. Abrahamson and Youngs (1992) (henceforth referred to as AY92) found that the algorithm can yield incorrect results for bad initial estimates of the model parameters, and they subsequently modified the algorithm to obtain a more stable version that is less sensitive to the initial model parameter estimates. The AY92 algorithm uses a combination of a fixed-effects regression algorithm and a likelihood maximization approach and is described in more detail later in this short note.

In the first step of the algorithm, it is assumed that the random-effects terms $\eta_1, \eta_2, \ldots, \eta_M$ equal zero, in which case equation (1) simplifies to $\ln(Y_{ij}) = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij}$. The coefficient set $\boldsymbol{\theta}$ is then estimated using a fixed-effects regression algorithm for the observed Y_{ij}. In the next step, the standard deviations σ (for the intraevent residuals) and τ (for the interevent residuals) are computed using the likelihood maximization approach described subsequently here.

The total residuals (i.e., the sum of the interevent and the intraevent residuals), denoted $\varepsilon_{ij}^{(t)}$, can be computed using the $\boldsymbol{\theta}$ estimated in the previous step as

$$\varepsilon_{ij}^{(t)} = \varepsilon_{ij} + \eta_i = \ln(Y_{ij}) - f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}). \tag{3}$$

Jayaram and Baker (2008) observed that the total residuals recorded at multiple sites during any particular earthquake can be assumed to jointly follow a multivariate normal distribution. Therefore, the likelihood (L_1) of having observed the set of total residuals $\boldsymbol{\varepsilon}^{(t)} = (\varepsilon_{ij}^{(t)})$ can be estimated as

$$\ln(L_1) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{C}| - \frac{1}{2}(\boldsymbol{\varepsilon}^{(t)})'\boldsymbol{C}^{-1}(\boldsymbol{\varepsilon}^{(t)}), \tag{4}$$

where N is the total number of data points, $\boldsymbol{C}$ is the covariance matrix of the total residuals and $(\boldsymbol{\varepsilon}^{(t)})'$ denotes the transpose of $\boldsymbol{\varepsilon}^{(t)}$. While estimating the model coefficients, AY92 assume that the intraevent residuals are independent of each other and of the interevent residuals. Hence, the covariance matrix $\boldsymbol{C}$ can be written as

$$\boldsymbol{C} = \sigma^2\boldsymbol{I}_N + \tau^2 \sum_{i=1}^{M}{}^{+} \boldsymbol{1}_{n_i,n_i}, \tag{5}$$

where $\boldsymbol{I}_N$ is the identity matrix of size $N \times N$, $\boldsymbol{1}_{n_i,n_i}$ is a matrix of ones of size $n_i \times n_i$, $\sum^+$ indicates a direct sum operation (using the notation of AY92), M is the number of earthquake events, and n_i is the number of recordings for the ith event. The matrix $\boldsymbol{C}$ can be expanded as

$$\boldsymbol{C} = \begin{bmatrix} \sigma^2\boldsymbol{I}_{n_1} + \tau^2\boldsymbol{1}_{n_1,n_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\boldsymbol{I}_{n_2} + \tau^2\boldsymbol{1}_{n_2,n_2} & \cdots & \boldsymbol{0} \\ . & . & \cdots & \boldsymbol{0} \\ . & . & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \sigma^2\boldsymbol{I}_{n_M} + \tau^2\boldsymbol{1}_{n_M,n_M} \end{bmatrix}. \tag{6}$$

The maximum likelihood estimates of σ and τ are those that maximize the likelihood function L_1 and are obtained using numerical optimization. Now, for given $\boldsymbol{\theta}$ and the maximum likelihood estimates of σ and τ, the random-effects term η_i is estimated using the maximum likelihood approach as well. The maximum likelihood estimate of η_i is obtained as follows (Abrahamson and Youngs, 1992):

$$\eta_i = \frac{\tau^2 \sum_{j=1}^{n_i} \varepsilon^{(t)}}{n_i\tau^2 + \sigma^2}. \tag{7}$$

Finally, using the estimated value of η_i, a new set of coefficients $\boldsymbol{\theta}$ is obtained using a fixed-effects algorithm for $\ln(Y_{ij}) - \eta_i$ (i.e., considering $\ln(Y_{ij}) - \eta_i = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij}$). The new set $\boldsymbol{\theta}$ is then used to reestimate σ, τ, and $\boldsymbol{\eta}$, and this iterative algorithm is continued until the coefficient estimates converge.

In summary, the steps of the mixed-effects algorithm used by AY92 are as follows:

1. Estimate the model coefficients $\boldsymbol{\theta}$ using a fixed-effects regression algorithm assuming $\boldsymbol{\eta}$ equals 0.
2. Using $\boldsymbol{\theta}$, solve for the variances of the residuals, σ^2 and τ^2, by maximizing the likelihood function described in equation (4).
3. Given $\boldsymbol{\theta}$, σ^2, and τ^2, estimate η_i using equation (7).
4. Given η_i, estimate new coefficients ($\boldsymbol{\theta}$) using a fixed-effects regression algorithm for $\ln(Y_{ij}) - \eta_i$.
5. Repeat steps 2, 3, and 4 until the likelihood in step 2 is maximized and the estimates for the coefficient set converge.

One drawback of this algorithm is the assumption in equation (5) that the intraevent residuals are independent of each other. It is known that the intraevent residuals are spatially correlated, with the correlation decreasing with increasing separation distance (e.g., Jayaram and Baker, 2009). Before addressing that issue, the need to account for the spatial correlation in the regression algorithm is illustrated in the next section.

Should Spatial Correlation Be Considered in the Regression Algorithm?

Consider the hypothetical case where the correlation between the intraevent residuals at any two different sites is a constant equal to ρ. In this case, the covariance matrix ($\boldsymbol{C}$) for the total residuals ($\varepsilon_{ij}^{(t)}$) is defined by the following equations:

$$\boldsymbol{C}(\varepsilon_{ij}^{(t)}, \varepsilon_{ij'}^{(t)}) = \rho\sigma^2 + \tau^2 \quad \forall\ i, j \neq j', \tag{8a}$$

$$\boldsymbol{C}(\varepsilon_{ij}^{(t)}, \varepsilon_{ij}^{(t)}) = \sigma^2 + \tau^2 \quad \forall\ i, j, \tag{8b}$$

and

$$\boldsymbol{C}(\varepsilon_{ij}^{(t)}, \varepsilon_{i'j'}^{(t)}) = 0 \quad \forall\ j, j', i \neq i'. \tag{8c}$$

In summary, the covariance matrix for the total residuals can be expressed as

$$\boldsymbol{C} = (1-\rho)\sigma^2\boldsymbol{I}_N + (\tau^2 + \rho\sigma^2)\sum_{i=1}^{M}{}^{+}\boldsymbol{1}_{n_i,n_i}. \tag{9}$$

Denoting $\sqrt{1-\rho}\sigma$ by σ' and $\sqrt{\tau^2 + \rho\sigma^2}$ by τ', equation (9) can be rewritten as

$$\boldsymbol{C} = \sigma'^2\boldsymbol{I}_N + \tau'^2\sum_{i=1}^{M}{}^{+}\boldsymbol{1}_{n_i,n_i}. \tag{10}$$

Comparing the forms of equations (5) and (10), it can be seen that the algorithm of AY92 actually provides the estimates of σ' and τ' rather than σ and τ. (If spatial correlations are absent, this is correct because $\sigma' = \sigma$ and $\tau' = \tau$.)

Assume for simplicity that the set of coefficients $\boldsymbol{\theta}$ is not affected by the spatial correlation (this assumption is subsequently relaxed). Hence, the correct estimates of σ and τ can be estimated from the σ' and τ' provided by AY92 as follows:

$$\sigma = \frac{\sigma'}{\sqrt{1-\rho}} \tag{11a}$$

and

$$\tau = \sqrt{\tau'^2 - \rho\sigma^2}. \tag{11b}$$

It is to be noted from the preceding discussion and equation (11) that assuming independent intraevent residuals will underestimate σ and overestimate τ. This has implications for lifeline risk assessments because a larger τ implies a higher likelihood of observing large ground-motion intensities throughout the region of interest. Thus, it is important to determine whether fitting the ground-motion equations while considering correlated intraevent residuals significantly changes the estimates of σ and τ.

Regression Algorithm for Mixed-Effects Models Considering Spatial Correlation

This section describes an algorithm for fitting the mixed-effects model while accounting for spatial correlation between intraevent residuals. The algorithm described here differs from that of AY92 in the estimation of the likelihood function L_1 (used in step 2) and in the computation of the interevent residual η_i (step 4). Both these changes are necessary to account for the spatial correlation between intraevent residuals in the regression algorithm.

Covariance Matrix for the Total Residuals

The covariance matrix for the total residuals shown in equation (5) is based on the assumption of independence between spatially-distributed intraevent residuals. The covariance matrix in the presence of spatial correlation is described subsequently here.

Let $\rho(d_{jj'})$ denote the spatial correlation between intraevent residuals at two sites j and j' as a function of $d_{jj'}$, the separation distance between j and j'. Then,

$$\begin{aligned} \boldsymbol{C}(\varepsilon_{ij}^{(t)}, \varepsilon_{ij'}^{(t)}) &= \boldsymbol{C}(\varepsilon_{ij} + \eta_i, \varepsilon_{ij'} + \eta_i) \\ &= \rho(d_{jj'})\sigma^2 + \tau^2 \quad \forall\ i, j, j' \end{aligned} \tag{12a}$$

and

$$\boldsymbol{C}(\varepsilon_{ij}^{(t)}, \varepsilon_{i'j'}^{(t)}) = 0 \quad \forall\ j, j', i \neq i'. \tag{12b}$$

It is to be noted that $\rho(d_{jj'})$, which is required for computing the covariance matrix, is typically unknown. In concept, the parameters that define $\rho(d_{jj'})$ can be treated as unknown and estimated as part of the regression algorithm. There are, however, advantages (discussed later in this manuscript) in estimating the spatial correlation outside of the ground-motion model fitting algorithm. The Estimates of Spatial Correlation section outlines a procedure that can be used to estimate $\rho(d_{jj'})$ and compute the above-described covariance matrix.

Obtaining Interevent Residuals from Total Residuals

The maximum likelihood approach is typically used to estimate a constant but unknown parameter from observed data. The parameter η_i that is of interest here, however, is a random variable in itself, and hence we use a Bayesian framework rather than the method of maximum likelihood to estimate η_i.

The prior distribution of η_i is $N(0, \tau^2)$. Conditional on the knowledge of η_i, the values for $\varepsilon_{ij}^{(t)}$ marginally follow a normal distribution with mean η_i and variance σ^2 (because $\varepsilon_{ij}^{(t)} = \varepsilon_{ij} + \eta_i$). Also, the correlation coefficient between $\varepsilon_{ij}^{(t)}$ and $\varepsilon_{ij'}^{(t)}$ conditional on η_i is given by $\rho(d_{jj'})$. In other words, the conditional covariance matrix ($\boldsymbol{C_c}$) for the total residuals can be expressed as

$$\boldsymbol{C_c}(\varepsilon_{ij}^{(t)}, \varepsilon_{ij'}^{(t)}) = \rho(d_{jj'})\sigma^2 \quad \forall\ i, j, j' \tag{13a}$$

and

$$\boldsymbol{C_c}(\varepsilon_{ij}^{(t)}, \varepsilon_{i'j'}^{(t)}) = 0 \quad \forall\ j, j', i \neq i'. \tag{13b}$$

Hence the joint density of $\boldsymbol{\varepsilon}_i^{(t)} = [\varepsilon_{i1}^{(t)}, \varepsilon_{i2}^{(t)}, \cdots, \varepsilon_{\mathrm{in}_i}^{(t)}]$ and η_i is expressed as

$$\begin{aligned} f(\boldsymbol{\varepsilon}_i^{(t)}, \eta_i) &= f(\boldsymbol{\varepsilon}_i^{(t)}|\eta_i) f(\eta_i) \\ &\propto \exp\left[-\frac{1}{2}(\varepsilon_i^{(t)} - \eta_i \mathbf{1}_{n_i,1})' \boldsymbol{C_c}^{-1} (\varepsilon_i^{(t)} - \eta_i \mathbf{1}_{n_i,1})\right] \\ &\quad \times \exp\left[-\frac{1}{2\tau^2}\eta_i^2\right], \end{aligned} \tag{14}$$

where $\boldsymbol{\varepsilon}_i^{(t)} = [\varepsilon_{i1}^{(t)}, \varepsilon_{i2}^{(t)}, \cdots, \varepsilon_{\mathrm{in}_i}^{(t)}]$ is the collection of total residuals at all the sites during earthquake i, $f(.)$ denotes the probability density function, $(\varepsilon_i^{(t)} - \eta_i \mathbf{1}_{n_i,1})'$ denotes the transpose of $(\varepsilon_i^{(t)} - \eta_i \mathbf{1}_{n_i,1})$, and $\mathbf{1}_{n_i,1}$ denotes a column matrix of ones of length n_i. It is to be noted that equation (14) is valid only if the interevent residual follows a normal distribution and the intraevent residuals at multiple sites during a given earthquake jointly follow a multivariate normal distribution. These assumptions have been verified using recorded ground motions by Jayaram and Baker (2008).

Noting that $f(\boldsymbol{\varepsilon}_i^{(t)}, \eta_i) = f(\boldsymbol{\varepsilon}_i^{(t)}) f(\eta_i|\boldsymbol{\varepsilon}_i^{(t)})$, one possible approach to identify the posterior distribution of η_i given $\varepsilon_i^{(t)}$ is to divide the joint density into a function of just $\boldsymbol{\varepsilon}_i^{(t)}$ and a function that also contains η_i. Let $Q(\boldsymbol{\varepsilon}_i^{(t)})$ denote any generic function of only $\boldsymbol{\varepsilon}_i^{(t)}$ not containing η_i. Hence,

$$\begin{aligned} f(\boldsymbol{\varepsilon}_i^{(t)}, \eta_i) &\propto \exp\Bigg[-\frac{1}{2}(\varepsilon_i^{(t)} - \eta_i \mathbf{1}_{n_i,1})' \\ &\qquad \times \boldsymbol{C_c}^{-1}(\varepsilon_i^{(t)} - \eta_i \mathbf{1}_{n_i,1}) - \frac{1}{2\tau^2}\eta_i^2\Bigg] \\ &= Q(\boldsymbol{\varepsilon}_i^{(t)}) \exp\Bigg[\frac{1}{2}\varepsilon_i^{(t)\prime} \boldsymbol{C_c}^{-1} \eta_i \mathbf{1}_{n_i,1} + \eta_i \mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \varepsilon_i^{(t)} \\ &\qquad - \frac{1}{2}\eta_i^2 \mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1} - \frac{1}{2\tau^2}\eta_i^2\Bigg] \\ &= Q(\boldsymbol{\varepsilon}_i^{(t)}) \exp\Bigg[-\frac{1}{2}\left(\frac{1}{\tau^2} + \mathbf{1}'_{n_i} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1}\right) \\ &\qquad \times \left(\eta_i - \frac{\mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \varepsilon_i^{(t)}}{\frac{1}{\tau^2} + \mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1}}\right)^2\Bigg]. \end{aligned} \tag{15}$$

From equation (15), it can be seen that $f(\eta_i|\varepsilon_i^{(t)})$ has a normal distribution with mean $\frac{\mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \varepsilon_i^{(t)}}{\frac{1}{\tau^2} + \mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1}}$ and variance $\frac{1}{\frac{1}{\tau^2} + \mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1}}$. If the best estimator for η_i is to be obtained under the squared-error loss criterion, then the Bayesian estimator of η_i equals the posterior mean (Lehmann and Casella, 2003)

$$\hat{\eta}_i = \frac{\mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \boldsymbol{\varepsilon}_i^{(t)}}{\frac{1}{\tau^2} + \mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1}}. \tag{16}$$

If the spatial correlation is absent, $\boldsymbol{C_c}$ is simply σ^2 times an identity matrix of size $n_i \times n_i$, in which case, $\mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \mathbf{1}_{n_i,1}$ equals n_i/σ^2 and $\mathbf{1}'_{n_i,1} \boldsymbol{C_c}^{-1} \boldsymbol{\varepsilon}_i^{(t)}$ equals $\sum_{j=1}^{n_i} \varepsilon_{ij}^{(t)}/\sigma^2$, and equation 16 becomes identical to equation 7.

Algorithm Summary

In summary, the steps of the modified mixed-effects algorithm are as follows:

1. Estimate the model coefficients $\boldsymbol{\theta}$ using a fixed-effects regression algorithm assuming $\boldsymbol{\eta}$ equals 0.
2. Using $\boldsymbol{\theta}$, solve for the variances of the residuals, σ^2 and τ^2, by maximizing the likelihood function described in equation (4). The covariance $\boldsymbol{C}$ in equation (4) is

estimated using equation (12). Detailed discussion about estimating $\rho(d_{jj'})$, which defined $\boldsymbol{C}$, is provided in the section Estimates of Spatial Correlation.

3. Given $\boldsymbol{\theta}$, σ^2, and τ^2, estimate η_i using equation (16).
4. Given η_i, estimate new coefficients ($\boldsymbol{\theta}$) using a fixed-effects regression algorithm for $\ln(Y_{ij}) - \eta_i$.
5. Repeat steps 2, 3, and 4 until the likelihood in step 2 is maximized and the estimates for the coefficient set converge.

Large Sample Standard Errors of σ and τ

If desired, the standard errors of the interevent and intraevent residual variances can be approximately estimated as the corresponding large sample values, which are calculated based on the following results from Searle (1971):

$$\mathrm{var}(\sigma^2) = 2\left[\mathrm{tr}\left(\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial(\sigma^2)}\right)^2\right]^{-1} \tag{17a}$$

and

$$\mathrm{var}(\tau^2) = 2\left[\mathrm{tr}\left(\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial(\tau^2)}\right)^2\right]^{-1}, \tag{17b}$$

where $\boldsymbol{C}$ is the covariance matrix defined in equation (12), $\frac{\partial \boldsymbol{C}}{\partial(\sigma^2)}$ denotes the partial derivative of $\boldsymbol{C}$ with respect to σ^2, $\frac{\partial \boldsymbol{C}}{\partial(\tau^2)}$ denotes the partial derivative of $\boldsymbol{C}$ with respect to τ^2, tr denotes the trace of a matrix, and var denotes variance. The partial derivatives, $\frac{\partial \boldsymbol{C}}{\partial(\sigma^2)}$ and $\frac{\partial \boldsymbol{C}}{\partial(\tau^2)}$, can be evaluated using numerical differentiation.

Alternately, the standard errors can also be evaluated using statistical techniques such as bootstrap (Efron and Tibshirani, 1998).

Mixed-Effects Regression Procedure in R

While mixed-effects regression procedures that consider spatial correlation (referred to as within-group correlation in statistical literature) are available in statistical programming languages such as R (e.g., the nlme function of Pinheiro *et al.*, 2009), it is potentially more convenient for current users of the Abrahamson and Youngs (1992) algorithm to switch to the modified algorithm described in this short note. Moreover, we experienced numerical instabilities with nlme while fitting the ground-motion model with consideration of within-group correlation. It is also to be noted that the development of the nlme function has stalled in favor of the revised version nlmer, which, however, does not yet provide the option of fitting nonlinear mixed-effects models considering within-group correlation.

Results and Discussion

In the current study, the algorithm described in the previous section is used to refit the Campbell and Bozorgnia (2008) ground-motion prediction model (henceforth referred to as the CB08 model) for illustration. First, in order to provide a baseline model for comparison, the coefficients of the CB08 model are reestimated while ignoring spatial correlation. For consistency, only records of CB08 are used by us for estimating the coefficients. Table 1 shows the regression coefficients estimated in this study for predicting spectral accelerations at 1 s (denoted $S_a(1\ \mathrm{s})$) in the uncorrelated case. Also shown in the table for comparison are the corresponding published CB08 model coefficients. Documentation of how these coefficients are used to make predictions is provided by CB08. The estimates of the standard deviations of the intraevent residual and the interevent residual (σ and τ, respectively) are shown in Table 2. The value of the published intraevent residual standard deviation reported here corresponds to that at large values for V_{S30}. (The V_{S30} is set above a threshold value beyond which the ground-motion model no longer considers soil nonlinearity effects, wherein the intraevent residuals have a constant variance at any given period. This allows a direct comparison of the published values to those estimated in this study, assuming homoscedastic residuals.) The refitted coefficients and variance estimates obtained in this work are similar, but not identical, to those reported by CB08. These small discrepancies are likely due to the manual coefficient smoothing carried out by us on the CB08 model (K. Campbell, personal comm, 2009). For consistency, the refitted model coefficients are treated as the benchmark values for comparison to model coefficients obtained considering spatial correlation. It is to be noted that the functional form of the CB08 model required knowledge about the A1100 value (median estimate of peak ground acceleration [PGA] on a reference rock outcrop

Table 1
Regression Coefficients for Estimating Median $S_a(1\ \mathrm{s})$

	Case 1*	Case 2*	Case 3*
c_0	−6.406	−6.487	−6.942
c_1	1.196	1.181	1.297
c_2	−0.772	−0.878	−1.073
c_3	−0.314	−0.379	−0.182
c_4	−2.000	−2.064	−2.112
c_5	0.170	0.195	0.198
c_6	4.00	3.884	4.440
c_7	0.255	0.264	0.324
c_8	0.000	−0.110	−0.093
c_9	0.490	0.897	0.796
c_{10}	1.571	1.577	1.565
c_{11}	0.150	0.122	0.093
c_{12}	1.000	0.871	0.865
k_1	400.0	400.0	400.0
k_2	−1.955	−1.955	−1.955
k_3	1.929	1.929	1.929

*Case 1, published CB08 results (Campbell and Bozorgnia, 2008); Case 2, estimated in this study without considering spatial correlation; Case 3, estimated in this study considering spatial correlation.

Table 2
Standard Deviations of Residuals Corresponding to $S_a(1\ \mathrm{s})$

Case*	Standard Deviation of Intraevent Residual (σ)	Standard Deviation of the Interevent Residual (τ)	Standard Deviation of the Total Residual ($\sqrt{\sigma^2+\tau^2}$)
1	0.568	0.255	0.623
2	0.578	0.223	0.620
3	0.654	0.157	0.673

*Case 1, published CB08 results (Campbell and Bozorgnia, 2008); Case 2, estimated in this study without considering spatial correlation; Case 3, estimated in this study considering spatial correlation.

with $V_{S30} = 1100$ m/s) for the median prediction. This is obtained directly using the coefficients of the CB08 model corresponding to PGA (as against fitting a separate model for the PGAs) for simplicity. This is reasonable because the model coefficients used for predicting median values do not change significantly after incorporating spatial correlation as shown subsequently in this paper.

The model coefficients are then reestimated considering spatial correlation. The spatial correlation model is obtained from Jayaram and Baker (2009) and is shown as

$$\rho(h) = e^{-3h/b}, \tag{18}$$

where h (km) denotes the separation distance between the sites of interest, and b denotes the range parameter that determines the rate of decay of correlation. This range is a function of the spectral period and equals 26 km when $S_a(1\ \mathrm{s})$ is considered. (The section Estimates of Spatial Correlation provides additional discussion about the choice of spatial correlation model.) The coefficient estimates (i.e., $\boldsymbol{\theta}$) obtained in this case are shown in Table 1. As can be seen from this table, the coefficients obtained by considering spatial correlation are similar to those obtained by ignoring spatial correlation. This is reinforced by a plot of the predicted medians at all the data sites using these two approaches (Figure 1). This matches with the observation of Hong *et al.* (2009) that the ground-motion model coefficients do not change significantly when considering spatial correlation.

While the coefficients for the median predictions are found to be relatively insensitive to the incorporation of spatial correlation, significant changes are seen in the estimates of the variance of the residuals (Table 2). In particular, the value of σ increases from 0.578 to 0.654 and the value of τ decreases from 0.223 to 0.157 after incorporating the spatial correlation. This trend is to be expected based on the illustrative example shown in the Introduction.

Standard Deviation of Residuals as a Function of Period

The results presented in the previous section support the use of the published coefficients (i.e., $\boldsymbol{\theta}$) for predicting the median intensities. The values of σ and τ, however, must be obtained considering spatial correlation. This implies that the iterative mixed-effects algorithm described earlier in the paper can be simplified to a computation of only the residual variances σ^2 and τ^2 (step 3) using the published values of $\boldsymbol{\theta}$; that is, the mixed-effects regression is now simply a random-effects regression procedure.

Hence, in this work, the CB08 model coefficients are assumed to be the fixed-effects model coefficients, and the total residuals are computed using the records in the Pacific

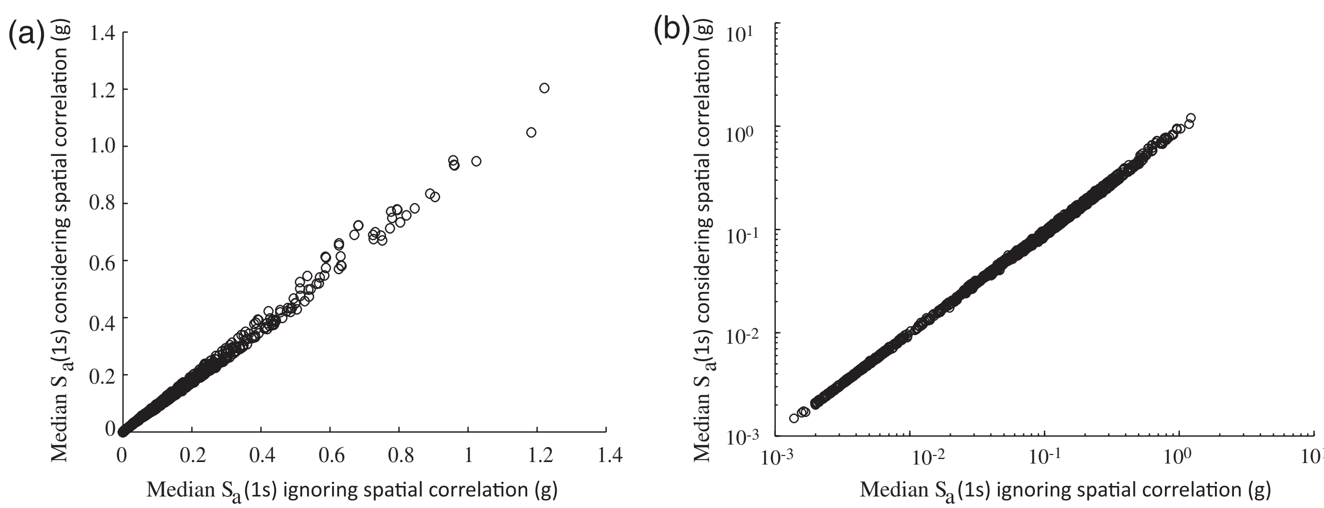

Figure 1. Comparison of predicted median Sa(1 s) values obtained using the CB08 model fitted with and without the consideration of spatial correlation: (a) linear scale and (b) log scale.

Earthquake Engineering Research–Next Generation Attenuation of Ground Motions (PEER NGA) database (only those records of the CB08 model used by us are considered for compatibility) (Chiou *et al.*, 2008). The maximum likelihood estimates of σ and τ are then obtained at different spectral acceleration periods from the total residuals using the procedures described earlier. Figure 2a compares the estimates of σ obtained in this study to those reported by CB08. It can be seen that the values of σ obtained considering spatial correlation are mostly larger than the published σ values (which have been estimated ignoring spatial correlations). Figure 2b shows that the values of τ, on the other hand, are considerably smaller when spatial correlations are considered. The values of σ and τ are then used to compute the standard deviations of the total residuals (computed as $\sqrt{\sigma^2 + \tau^2}$) and plotted in Figure 2c. It can be seen from this figure that considering spatial correlation does not significantly alter the total residual standard deviation. (Hong *et al.*, 2009, noticed a small reduction in the total residual standard deviation when the spatial correlation was considered. The alteration in the total residual standard deviation could depend on the data set and the spatial correlation model used.)

Though the current work only refits the CB08 model, the trends in the values of σ and τ are the same for the other recent NGA ground-motion models (e.g., Boore and Atkinson, 2008; Chiou and Youngs, 2008). This can be seen from Figure 2d, which shows typical ratios of the interevent residual standard deviation to the total residual standard deviation reported by these ground-motion models. It is seen that the ratios reported by the ground-motion modelers are generally much larger than those estimated in this work, considering spatial correlation.

Estimates of Spatial Correlation

As discussed earlier, step 2 of the proposed algorithm (see Algorithm Summary section) requires the computation of the covariance matrix shown in equation (12). The covariance matrix is defined by the spatial correlation between the intraevent residuals denoted $\rho(d_{jj'})$; $\rho(d_{jj'})$ is unknown, however, and in concept can be estimated as part of the regression algorithm. Alternately, $\rho(d_{jj'})$ can also be precomputed using ground-motion models that are fitted without consideration of spatial correlation and used while

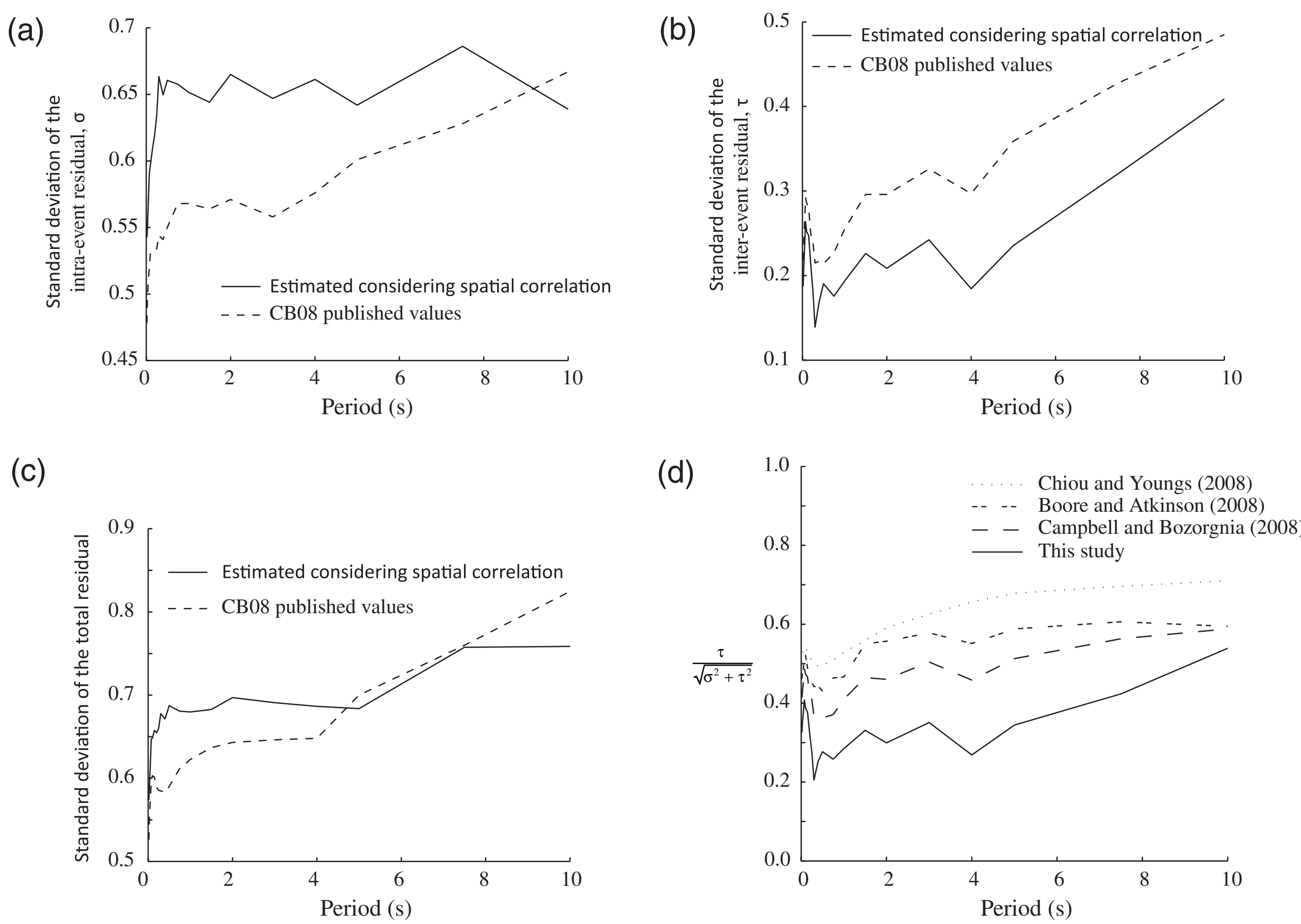

Figure 2. Effect of spatial correlation on: (a) estimated intraevent residual standard deviation (σ), (b) estimated interevent residual standard deviation (τ), (c) estimated total residual standard deviation. (d) Ratio of interevent residual standard deviation to total residual standard deviation.

developing ground-motion models with consideration of spatial correlation. As discussed earlier, the consideration of spatial correlation while fitting the models does not change the median predictions and, therefore, the total residuals (equation 1). Jayaram and Baker (2009) also showed that the spatial correlation between intraevent residuals can be estimated directly from total residuals (exactly when the intraevent residuals are homoscedastic and approximately otherwise). Therefore, accurate spatial correlation estimates can be obtained using ground-motion models fitted without consideration of spatial correlation. In other words, it is still appropriate to use the correlation models previously developed using the published ground-motion models. The advantages of precomputing a spatial correlation model as suggested previously in this short note rather than estimating the spatial correlation while fitting the ground-motion model are:

1. Separating out the development of the spatial correlation model and the ground-motion model allows the use of different ground-motion data sets for these two purposes. This is advantageous because the NGA database used by the ground-motion modelers has numerous events with very few recordings that cannot be used for estimating reliable spatial correlation estimates.
2. Jayaram and Baker (2009) argued that the spatial correlation model should provide accurate estimates of the correlation at short separation distances even if that means slightly inaccurate estimates at longer separation distances. This is hard to implement if the spatial correlation model parameters are estimated as part of the ground-motion model regression procedure, particularly if existing software packages such as nlme are used.
3. The extent of spatial correlation sometimes depends on site-related parameters such as V_{S30} (e.g., Jayaram and Baker, 2009). It is hard to incorporate such dependencies while developing spatial correlation models as part of the ground-motion model regression procedure.
4. Developing the spatial correlation model separately can potentially improve the numerical stability of the ground-motion model regression procedure.

The current study estimates $\rho(d_{jj'})$ using the spatial correlation model provided by Jayaram and Baker (2009) (equation 18), which is based on residuals computed using the published ground-motion models (fitted without consideration of spatial correlation). This model has been fitted using seven well-recorded earthquakes. Future studies could also explore the option of using event-specified spatial correlation models for well-recorded events and generic correlation models for poorly-recorded events (for which reliable spatial correlation estimates can not be obtained).

Implications for Risk Assessment

Because ignoring spatial correlation while fitting the ground-motion model does not significantly affect the estimates of the ground-motion medians ($f(\boldsymbol{\theta})$) or the standard deviation of the total residuals (Figure 2c), hazard and loss analyses for single structures will produce accurate results if the existing ground-motion models are used. Risk assessments for spatially-distributed systems, however, are influenced by the standard deviation of the interevent and the intraevent residuals and not just by the medians and the standard deviation of the total residuals.

A large value of τ increases the likelihood of observing large positive interevent residuals, which will simultaneously increase the ground-motion intensity at all the sites in the region. If spatial correlations are large, a large value of σ will have a similar effect and can result in large ground-motion intensities at multiple sites. In such a case, the effect of underestimating σ is compensated by the effect of overestimating τ. If the spatial correlations are small, however, underestimating σ and overestimating τ will have the net effect of jointly producing more extreme ground-motion intensities at multiple sites than is probable in reality. It can be inferred from equation (18) that the spatial correlation will be small if h is large or if b is small. Therefore, when the components of a spatially distributed system are well separated (large h) or if the correlation range is small, the ground-motion models fitted without considering spatial correlation will overestimate the likelihood of jointly observing extreme ground-motion intensities at multiple sites. It is difficult to make general conclusions about the size of this effect, but it is clear that this will have some impact on the estimated seismic risk of spatially distributed systems.

Conclusions

This work illustrated the impact of considering spatial correlation between intraevent residuals while developing ground-motion models. The mixed-effects algorithm of Abrahamson and Youngs (1992), which assumes independence between intraevent residuals, was modified to account for the spatial correlation between the intraevent residuals. This was done by changing the likelihood function used for estimating the interevent and the intraevent residual variances given other model coefficients and changing the estimate of the interevent residual given the total residuals at multiple sites. The modified algorithm was used to refit the Campbell and Bozorgnia (2008) ground-motion model to illustrate the effect of this refinement. The variance of the total residuals and the model coefficients used for predicting the median ground-motion intensity were not significantly affected by the proposed refinement. Significant changes, however, were seen in the variance of the intraevent and the interevent residuals. Incorporating spatial correlation was seen to increase the intraevent residual variance and to decrease the interevent residual variance. These changes have implications for risk assessments of spatially-distributed systems because a smaller interevent residual variance implies a lesser likelihood of simultaneously observing larger-than-median ground-motion intensities at all sites in a region.

Data and Resources

The data for all the ground motions studied here came from the Pacific Earthquake Engineering Research–Next Generation Attenuation of Ground Motions (PEER NGA) database, available at http://peer.berkeley.edu/nga (last accessed 29 April 2010).

Acknowledgments

The authors thank Peter Stafford and an anonymous reviewer for their helpful reviews of the manuscript. Thanks to Ken Campbell for helpful explanations regarding the development of the Campbell and Bozorgnia (2008) ground-motion model. Thanks to Shrey Shahi for useful discussions on this research topic. This work was partially supported by the U.S. Geological Survey (USGS) via External Research Program award 07HQGR0031. Any opinions, findings, and conclusions or recommendations expressed in this material are ours and do not necessarily reflect those of the USGS.

References

Abrahamson, N. A., and W. J. Silva (2008). Summary of the Abrahamson & Silva NGA ground-motion relations, *Earthq. Spectra* **24,** no. 1, 99–138.

Abrahamson, N. A., and R. R. Youngs (1992). A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am.* **82,** no. 1, 505–510.

Bolt, B., and N. Abrahamson (1982). New attenuation relations for peak and expected accelerations of strong ground motion, *Bull. Seismol. Soc. Am.* **72,** 2307–2321.

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 0.01 s and 10.0 s, *Earthq. Spectra* **24,** no. 1, 99–138.

Boore, D. M., J. F. Gibbs, W. B. Joyner, J. C. Tinsley, and D. J. Ponti (2003). Estimated ground motion from the 1994 Northridge, California, earthquake at the site of the Interstate 10 and La Cienega Boulevard bridge collapse, West Los Angeles, California, *Bull. Seismol. Soc. Am.* **93,** no. 6, 2737–2751.

Brillinger, D. R., and H. K. Preisler (1984a). An exploratory analysis of the Joyner–Boore attenuation data, *Bull. Seismol. Soc. Am.* **74,** 1441–1450.

Brillinger, D. R., and H. K. Preisler (1984b). Further analysis of the Joyner–Boore attenuation data, *Bull. Seismol. Soc. Am.* **75,** 611–614.

Campbell, K. W., and Y. Bozorgnia (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s, *Earthq. Spectra* **24,** no. 1, 139–171.

Chiou, B. S.-J., and R. R. Youngs (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **24,** no. 1, 173–215.

Chiou, B., R. Darragh, N. Gregor, and W. Silva (2008). NGA project strong-motion database, *Earthq. Spectra* **24,** no. 1, 23–44.

Dempster, A., N. Laird, and H. Preisler (1977). An exploratory analysis of the Joyner–Boore attenuation data, *J. Roy. Stat. Soc.* **39,** 1–38.

Efron, B., and R. Tibshirani (1998). *An Introduction to the Bootstrap*, CRC Press LLC, Boca Raton, Florida, 456 pp.

Goda, K., and H. P. Hong (2008). Spatial correlation of peak ground motions and response spectra, *Bull. Seismol. Soc. Am.* **98,** no. 1, 354–365.

Hong, H. P., Y. Zhang, and K. Goda (2009). Effect of spatial correlation on estimated ground-motion prediction equations, *Bull. Seismol. Soc. Am.* **99,** no. 2A, 928–934.

Jayaram, N., and J. W. Baker (2008). Statistical tests of the joint distribution of spectral acceleration values, *Bull. Seismol. Soc. Am.* **98,** no. 5, 2231–2243.

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

Joyner, W., and D. Boore (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seismol. Soc. Am.* **71,** 2011–2038.

Joyner, W., and D. Boore (1993). Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am.* **83,** no. 2, 469–487.

Lehmann, E., and G. Casella (2003). *Theory of Point Estimation*, 2nd Ed., Springer, New York, 590 pp.

Pinheiro, J., D. Bates, S. DebRoy, and D. Sarkarthe R Core team (2009). *nlme: Linear and Nonlinear Mixed Effects Models*, R package version 3.1-96, available at http://cran.r-project.org/web/packages/nlme/index.html (last accessed August 2010)

Searle, S. (1971). *Linear Models*, John Wiley and Sons, Inc., New York, 560 pp

Wang, M., and T. Takada (2005). Macrospatial correlation model of seismic ground motions, *Earthq. Spectra* **21,** no. 4, 1137–1156.

Dept. of Civil and Environmental Engineering
Stanford University
Stanford, California 94305-4020
nirmal.jayaram@rms.com
bakerjw@stanford.edu

Manuscript received 26 November 2009