Advances in Engineering Software 000 (2016) 1–10

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

A kernel-based mixed effect regression model for earthquake ground motions

J. Tezcan[a,*], Y. Dak Hazirbaba[a], Q. Cheng[b]

[a] Department of Civil and Environmental Engineering, Southern Illinois University Carbondale, IL, United States
[b] Computer Science Department, Southern Illinois University Carbondale, IL, United States

ARTICLE INFO

Article history:
Received 22 June 2015
Revised 2 May 2016
Accepted 4 June 2016
Available online xxx

Keywords:
Least squares kernel machine
Mixed-effect model
Semi-parametric regression
Residual maximum likelihood method
Ground motion analysis
Confidence intervals

ABSTRACT

This paper presents a semi-parametric mixed-effect regression approach for analyzing and modeling earthquake ground motions, taking into account the correlations between records. Using kernels, the proposed method extends the classical mixed model equations to complicated relationships. The predictive equation is composed of parametric and nonparametric parts. The parametric part incorporates known relationships into the model, while the nonparametric part captures the relationships which cannot be cast into a simple parametric form. A least squares kernel machine is used to infer the nonparametric part of the model. The resulting semi-parametric model combines the strengths of parametric and nonparametric approaches, allowing incorporation of prior, well-justified knowledge into the model while retaining flexibility with respect to the explanatory variables for which the functional form is uncertain. Equations for pointwise confidence and prediction intervals around the conditional mean are provided. The validity of the proposed method is demonstrated through numerical simulations and using recorded ground motions.

1. Introduction

The main goal in seismic design of structures is to ensure that the structure will perform as intended throughout its service life, when subjected to ground shaking. While the magnitude and location of future earthquakes cannot be known with certainty, the probability distribution of the intensity of ground shaking expected at a given site can be described using Probabilistic Seismic Hazard Analysis (PSHA). At its most basic form, PSHA can be thought of as a two-step process: 1) Characterization of the distribution of magnitude and source-to-site distances from each potential earthquake source, and 2) Characterization of the distribution of the intensity of the ground shaking that would result from each scenario. The second step is achieved using Ground Motion Prediction Equations (GMPEs) which describe the relationship between a ground motion intensity variable and a set of predictors containing, at a minimum, moment magnitude, and source-to-site distance. Both the conditional mean of the target variable and the prediction uncertainty have major implications in the development of seismic code provisions, insurance calculations, and public policy decisions.

A typical GMPE has the following form

$$y_{ij} = f(\mathbf{x_{ij}}) + u_i + e_{ij}, \tag{1}$$

where y_{ij} is the target variable, typically chosen to be the logarithm of an intensity measure corresponding to the jth recording from the ith earthquake, $f(\mathbf{x_{ij}})$ is a parametric function of predictive variables $\mathbf{x_{ij}}$, u_i is the inter-event residual and e_{ij} is the intra-event residual. Both u_i and ε_{ij} are assumed to be zero-mean Gaussian random variables.

Two methods are common in fitting Eq. (1) to ground motion data: one-stage and two-stage methods. One-stage methods [1–3] estimate the model parameters simultaneously by maximizing the likelihood function, while two-stage methods determine the magnitude and distance dependence on separate steps. The first stage estimates the distance dependence and assigns an amplitude factor to each earthquake. The second stage estimates the magnitude dependence by maximizing the likelihood of the amplitude factors determined in the first stage. Joyner and Boore [4] have shown that both methods are equivalent in terms of bias and variance.

While traditional GMPEs involved a small set of predictors, recent growth in seismic databases and advances in computer technology have led to the creation of sophisticated GMPEs involving tens of parameters. The increase in the number of parameters gives the model the flexibility to describe a large class of functions, however, this flexibility often comes at the expense of increased susceptibility to overfitting, especially when least squares criterion is used without any complexity penalty. As correct specification of

* Corresponding author.
E-mail address: jale@siu.edu (J. Tezcan).

http://dx.doi.org/10.1016/j.advengsoft.2016.06.002
0965-9978/

the functional form of a GMPE is critical to its predictive performance, whether the size of the available ground motion databases justifies this level of complexity is a question of growing concern in the earthquake engineering community.

A nonparametric approach, which does not require a fixed parametric form may alleviate the problem of model misspecification. However, it also does not allow incorporation of known relationships into the model. A semiparametric approach combines the strengths of parametric and nonparametric approaches, allowing the model to include known relationships, and retaining flexibility with respect to the predictors whose relation to the target variable cannot be specified with certainty. In conjunction with a complexity penalty, the overfitting susceptibility of the semiparametric model can be controlled.

In this study, we consider the semi-parametric model

$$y_{ij} = f(\mathbf{x_{ij}}) + h\left(\mathbf{s_{ij}}\right) + u_i + e_{ij},\ \ i = 1:N,\ \ j = 1:n_i, \tag{2}$$

where f is a parametric function, h is a nonparametric function, $u_i \sim N(0, \sigma_u^2)$ is the random effect term associated with event i, and $e_{ij} \sim N(0, \sigma_r^2)$ is the error term associated with y_{ij}. The vectors $\mathbf{x_{ij}}$ and $\mathbf{s_{ij}}$ contain the predictors used in f and h, respectively. When f is a linear combination of the fixed effects, and the nonparametric function h is assumed to lie in a Reproducing Kernel Hilbert Space (RKHS), the solution can be obtained by solving a linear system of equations.

The main advantage of the semi-parametric model over parametric models is the flexibility it provides in modeling complicated relationships when a mathematical form cannot be specified with reasonable confidence, while allowing incorporation of prior knowledge regarding the functional form with respect to a subset of the explanatory variables. In addition, unlike existing kernel-based ground motion models (e.g. [5,6]), the proposed model has the ability to take into account the correlations between records.

This paper is based upon Tezcan et al. [7], but the current paper includes the following extensions. First, recognizing that the proposed estimator is a linear smoother, we derive equations for conditional bias and variance based on the associated smoother matrix. Next, using these two quantities, we construct bias-corrected pointwise confidence and prediction intervals around the conditional expectation. With this extension, the uncertainty in the predictions is quantified.

The remainder of this paper is organized as follows. Section 2 is devoted to a brief overview of the Least Squares Kernel Machine (LSKM) and the linear Mixed Model Equations (MME), and examines the connection between them. Section 3 presents the semiparametric mixed-effect model. Following a derivation of the regression function, equations for bias-corrected confidence and prediction intervals around the conditional expectation are presented. Section 4 shows a demonstrative example, using a dataset containing 182 records from 23 shallow earthquakes in western North America. Section 5 presents simulation studies comparing the accuracy and computational efficiency of the proposed approach to the conventional, parametric approach. Finally, Section 6 presents the conclusions of this study.

2. Relationship between least squares kernel machines and linear mixed models

This section is intended to demonstrate the mathematical connection between least squares kernel machines and linear mixed models. Following a brief overview of the two approaches, the least squares kernel machine model is presented as a special case of the linear mixed model. This connection allows constructing a nonlinear model using Henderson's mixed model equations, which is a system of linear equations.

2.1. Least squares kernel machine

Kernels offer an efficient way to formulate nonlinear generalizations of linear algorithms that are based on inner products. Introduced in 1960s [8], kernel-based learning has been increasingly used in a multitude of data analysis applications in various disciplines.

The Support Vector Machine (SVM) [9] is a supervised, kernel-based algorithm that can be used in classification and regression applications. The SVM finds the solution of a constrained optimization problem, which, in the primal form, is expressed in terms of a nonlinear feature mapping function, $\boldsymbol{\phi}(\mathbf{x})$. Through introduction of Lagrange multipliers and a kernel function $K(\mathbf{x_i},\ \mathbf{x_j}) = \boldsymbol{\phi}(\mathbf{x_i})^T\boldsymbol{\phi}(\mathbf{x_j})$ satisfying the Mercer's condition [10], the primal form is cast into the dual form which is often easier to solve.

Least Squares Kernel Machine (LSKM) [11] was recently introduced as a variant of the classical SVM classifier, where the inequality constraints are replaced by equality constraints. This reformulation allows solving the SVM problem by reducing the corresponding convex quadratic programming problem to a linear system of equations. In addition to offering simpler software implementation and increased numerical stability, this approach allows extension of the classical SVM to a wider range of problems in data analysis and pattern recognition [12].

The use of kernels allows solving regression problems by performing a ridge regression in the feature space [13,14]. Given an input matrix $\mathbf{X} \in \mathbb{R}^{N\times d}$ representing the coordinates of N points, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, in a d-dimensional space and the corresponding output vector $\mathbf{y} \in \mathbb{R}^{N\times 1}$, and assuming that each observation is related to the corresponding input as

$$y_i = \mathbf{w}^T\ \boldsymbol{\phi}(\mathbf{x_i}) + b + \ e_i,\ \ i = 1, 2, .., N, \tag{3}$$

the predictive model $f(\mathbf{x}) = \mathbf{w}^T\ \boldsymbol{\phi}(\mathbf{x}) + b$ is obtained by minimizing the objective function

$$\frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{\gamma}{2}\sum_{i=1}^{N} e_i^2 \tag{4}$$

under the set of N constraints given in Eq. (3), using a penalty parameter γ. Introducing Lagrange multipliers $\alpha_i,\ i = 1, \ldots, N$, the Lagrangian function is written as

$$L(\mathbf{w}, b, \mathbf{e},\ \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{\gamma}{2}\mathbf{e}^T\mathbf{e} - \ \sum_{i=1}^{N}\alpha_i\left[\mathbf{w}^T\boldsymbol{\phi}(\mathbf{x_i}) + b + \ e_i - y_i\right]. \tag{5}$$

where $\mathbf{e} = [e_1, e_2, \ldots, e_N]^{\mathrm{T}}$ and $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^{\mathrm{T}}$. Optimal values of $\mathbf{w}, b,\ e_i$ and α_i are found by locating the stationary points of the Lagrangian function given in Eq. (5). Using $\mathbf{S} = [\boldsymbol{\phi}(\mathbf{x}_1), \boldsymbol{\phi}(\mathbf{x}_2), \ldots, \boldsymbol{\phi}(\mathbf{x}_N)]^{\mathrm{T}}$, the optimality condition can be summarized as

$$\begin{bmatrix} \mathbf{I}_k & \mathbf{0_k} & \mathbf{0_{k\times N}} & -\mathbf{S}^T \\ \mathbf{0_k^T} & 0 & \mathbf{0_N^T} & 1_N^T \\ \mathbf{0_{N\times k}} & \mathbf{0_N} & \gamma\mathbf{I}_N & -\mathbf{I}_N \\ \mathbf{S} & \mathbf{1_N} & \mathbf{I}_N & \mathbf{0_N} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ b \\ \mathbf{e} \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{0_k} \\ 0 \\ \mathbf{0_N} \\ \mathbf{y} \end{bmatrix} \tag{6}$$

where $\mathbf{I}_p$ is an identity matrix of size p, $0_{p\times q}$ is the $p \times q$ matrix of zeros, $\mathbf{1_p}$ and $\mathbf{O}_p$ are the $p \times 1$ vectors of ones and zeros respectively, and k is the dimension of the feature space. Elimination of vectors $\mathbf{w}$ and $\mathbf{e}$ leads to

$$\begin{bmatrix} 0 & \mathbf{1_N^T} \\ \mathbf{1_N} & \boldsymbol{\Omega} + (1/\gamma)\mathbf{I}_N \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{7}$$

where $\boldsymbol{\Omega} = \mathbf{S}\mathbf{S}^T$ is the kernel matrix of size $N \times N$ where $\Omega_{ij} = K(\mathbf{x}_i,\ \mathbf{x}_j)$. If the square matrix in Eq. (7) is invertible, b and $\boldsymbol{\alpha}$ can

Please cite this article as: J. Tezcan et al., A kernel-based mixed effect regression model for earthquake ground motions, Advances in Engineering Software (2016), http://dx.doi.org/10.1016/j.advengsoft.2016.06.002