STATISTICS IN MEDICINE
Statist. Med. 2009; **28**:1896–1912
Published online 28 April 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3600

Hierarchical additive modeling of nonlinear association with spatial correlations—An application to relate alcohol outlet density and neighborhood assault rates

Qingzhao Yu[1,*,†], Bin Li[2] and Richard Allen Scribner[3]

[1]*Biostatistics Program, School of Public Health, LSU Health Science Center, New Orleans, LA, U.S.A.*
[2]*Department of Experimental Statistics, Louisiana State University, New Orleans, LA, U.S.A.*
[3]*Epidemiology Program, School of Public Health, LSU Health Science Center, New Orleans, LA, U.S.A.*

SUMMARY

Previous studies have suggested a link between alcohol outlets and assaults. In this paper, we explore the effects of alcohol availability on assaults at the census tract level over time. In addition, we use a natural experiment to check whether a sudden loss of alcohol outlets is associated with deeper decreasing in assault violence. Several features of the data raise statistical challenges: (1) the association between covariates (for example, the alcohol outlet density of each census tract) and the assault rates may be complex and therefore cannot be described using a linear model without covariates transformation, (2) the covariates may be highly correlated with each other, (3) there are a number of observations that have missing inputs, and (4) there is spatial association in assault rates at the census tract level. We propose a hierarchical additive model, where the nonlinear correlations and the complex interaction effects are modeled using the multiple additive regression trees and the residual spatial association in the assault rates that cannot be explained in the model are smoothed using a conditional autoregressive (CAR) method. We develop a two-stage algorithm that connects the nonparametric trees with CAR to look for important covariates associated with the assault rates, while taking into account the spatial association of assault rates in adjacent census tracts. The proposed method is applied to the Los Angeles assault data (1990–1999). To assess the efficiency of the method, the results are compared with those obtained from a hierarchical linear model. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: alcohol-related crimes; backfitting; conditionally autoregressive (CAR) model; multiple additive regression trees (MART); nonparametric regression

*Correspondence to: Qingzhao Yu, School of Public Health, LSUHSC, Suite 1400, 1615 Poydras Street, New Orleans, LA 70112, U.S.A.
†E-mail: qyu@lsuhsc.edu

Contract/grant sponsor: National Institute of Alcohol Abuse and Alcoholism; contract/grant number: R01 AA013810_03

Received 6 November 2008
Accepted 13 March 2009

1. INTRODUCTION

Research on the relationship between alcohol outlets and assault is important for informing policy related to the competing interests of the alcohol industry, e.g. to increase the distribution and consumption of alcohol, and public safety, e.g. to minimize risks associated with increasing the alcohol availability. There are growing literature that demonstrate relations between alcohol outlets and measures of interpersonal violence. Since the effect of alcohol outlets on violence is believed to be contextual, these analysis involve areal data that require more sophisticated techniques to account for their spatial structure. Initial studies on the role of alcohol outlets in the neighborhood environment and assault were ecological in design and conducted at the city level [1–3]. As geographical information systems (GIS) software became widely available, more local units of analysis (e.g. census tracts) were used in modeling the relation between alcohol outlets and assault [4, 5]. Accounting for spatial autocorrelation, i.e. the possible spillover of assault into contiguous neighborhoods, was a subsequent advancement [6–9]. Recently, longitudinal models are being incorporated into the analysis of the data, introducing an additional level of complexity, i.e. temporal structure [10].

In this study we apply a hierarchical additive model to explore possible coefficients that are related to changes in assault rates among census tracts affected by the 1992 Civil Unrest in Los Angeles, which resulted in the immediate loss of over 250 alcohol outlets and the permanent loss of roughly 150 alcohol outlets. We use this natural experiment to check whether a sudden loss of alcohol outlets is associated with deeper decrease in assault rates. We want also to check in general, whether alcohol outlet density is positively related to assault rates.

Several features of the data present substantial challenges in this study of contextual effects of alcohol outlets. First, the association between the assault violence rates and other covariate may be nonlinear. There are two common approaches to deal with nonlinearity: (1) transformation and (2) basis expansion (such as spline basis). However, the former is not readily applicable when there are a large number of covariates, while the latter requires basis specification (e.g. specify the number and locations of knots). Second, complicated interactions might exist among covariates. We could use model selection methods such as Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) to choose appropriate interaction terms. But then we need to create all possible interactions. In addition the test of all combination of variables results in big calculation burden. Third, more than 7 per cent of the observations are missing one or more values for some covariates. Finally, in the analysis, we should take into account the spatial correlations among observations in adjacent tracts. A handy method to deal with the correlation is to use a hierarchical model where the spatial correlations are modeled through correlated spatial random effects. We are consequently challenged by combining the hierarchical structure of spatial errors with nonlinear models. In this analysis, we apply both multiple additive regression trees (MART) and conditional autoregressive model (CAR) to both address these challenges and provide a better description of the relations between covariates and the assault.

MART, originally proposed by Friedman [11], is an ensemble technique that aims to improve the performance of a single model by fitting many models and combining them for prediction. MART employs two algorithms: 'regression tree' from classification and regression trees (CART) [12] and 'boosting', which builds and combines a collection of models, i.e. trees.

CART is a binary recursive partitioning algorithm that provides an alternative to traditional parametric models for regression and classification problems. The term 'binary' implies that at each step, CART splits a multidimensional covariate space into two regions, and models the response

DOI: 10.1002/sim

as a constant for each region. Then an optimal variable and split-point are chosen to achieve the best fit again on one or both of these regions. Thus, each node can be split into two child nodes, in which case the original node is called a parent node. The term 'recursive' refers to the fact that the binary partitioning process can be applied over and over again. Thus, each parent node can give rise to two child nodes and, in turn, each of these child nodes may themselves be split to generate additional children. Although CART represents information in a way that is intuitive and easy to be visualized, it is not usually as accurate as its competitors.

Boosting is one of the recent enhancements to tree-based methods that have met with considerable success in prediction accuracy. In boosting, models such as regression trees are fitted iteratively to the training data, using appropriate methods to gradually increase emphasis on observations modeled poorly by the existing collection of trees.

Empirical results have shown that MART achieves accurate prediction performance compared with its competitors. Moreover, compared with the classical parametric regression methods, MART has the following advantages: (1) MART is able to capitalize on the nonlinear relationships between the dependent and independent variables with no need for specifying the basis functions. Unlike many automated learning procedures, which lack interpretability and operate as a 'black box', MART provides results that represent valuable tools for interpreting the nature and magnitude of covariate relations with the outcome (see, for example, relative variable importance and partial dependence plot in Section 3.4). (2) Owing to the hierarchical splitting scheme in regression trees, MART is able to capture complex and/or high-order interaction effects. (3) As a tree-based method, MART can handle mixed-type predictors (i.e. quantitative and qualitative covariates) and missing values in covariates. Therefore MART addresses the first three challenges in analyzing the data. Further description of MART is given in Section 3.1.

To tackle the last challenge related to the spatial structure of data, we propose a two-stage iterative algorithm to build hierarchical additive models. At the first stage, MART is built to explore the associations between the smoothed assault rates and the covariates. Here the *smoothed assault rates* refer to the original assault rates minus the estimated spatial terms obtained from the second stage. At the second stage, the spatial correlations in assault rate that could not be explained by covariates are 'smoothed' through the CAR. The two stages iterate until convergence. Our algorithm is an extension of the backfitting process [13] to more complicated nonparametric settings. In this paper, we apply the hierarchical additive modeling strategy to evaluate the association between the alcohol availability and assault rates in census tracts of Los Angeles affected by the 1992 civil unrest for the years 1990–1999.

The rest of the paper is organized as follows. We describe in Section 2 the data and environment. Next, the two-stage hierarchical additive model is presented in Section 3. In Section 4, we apply the hierarchical additive model to analyze the data and the results are compared with those from a hierarchical linear regression model. Concluding remarks and implications of the research are provided in Section 5.

2. THE 1992 CIVIL UNREST AND DATA

2.1. The 1992 civil unrest

Our study is designed to capitalize on a natural experiment. The experiment was made possible by the civil unrest in Los Angeles that followed the verdicts acquitting the police officers accused

DOI: 10.1002/sim

of beating Rodney King. The civil unrest resulted in 53 deaths, 2325 reported injuries, more than 600 buildings were completely destroyed by fire, and approximately $ 735 million in total damages [14]. Many of the buildings destroyed were serving as alcohol outlets. As a result, a total of 279 liquor licenses were surrendered across 144 census tracts in Los Angeles County due to interruption of their services. In the wake of the civil unrest, an effort to halt the rebuilding of off-sale alcohol outlets was successful in restricting the re-licensing of outlets with a history of problems (e.g. assaults, homicides, drug sales) around their premises. The effort was responsible, in part, for over 150 outlets permanently closing in the civil unrest area. These events provide a natural experiment setting to test various hypotheses regarding the effect of closure of alcohol outlets in 144 tracts, compared with 336 tracts also exposed to the civil unrest where outlets were not affected [15].

The 1992 civil unrest occurred over a large area of South Central Los Angeles. In the present study we include only those census tracts in the area that were affected by the civil unrest, thereby controlling a possible global effect of the unrest on outcomes. To define the study area in this manner, we used the definition established by the Rose Institute of California State and Local Government at Claremont McKenna College (http://ccdl.libraries.clare- mont.edu/col/ric/) to study the economic impact of the civil unrest [16]. A total of 480 census tracts comprise the unrest area. These tracts contained 2 641 320 people in 1990, of whom 48 per cent were Hispanic and 27 per cent were African American. A total of 2240 unique addresses were damaged in the 480 tracts, while 144 tracts had one or more off-sale liquor outlets whose license was surrendered. The majority of the damaged addresses were commercial businesses. Immediately following the civil unrest, there emerged a grass roots effort among the affected communities to halt the rebuilding of alcohol outlets based on the finding that an over-concentration of outlets existed in the low socio-economic status areas prior to the civil unrest [17–19].

2.2. Data

The study time frame is 1990–1999. The purpose of this study is to check whether a sudden decreasing in alcohol outlet is related to a deeper drop in assault rates in the following years, and we want to check the general relationship between the alcohol outlet density and assault rates. We do this study at the census tract level. To do the analysis, we should consider other related factors such as the race distribution in the census tract. We should also take into account the possible spatial association among adjacent tracts.

Assaultive violence: Our measure of assault was obtained from the Los Angeles Police Department. Uniform crime report (UCR) offenses involving assault (i.e. murder, rape, robbery, and assault) were obtained for the years 1990–1999. A summary measure of the count of all violent offenses was generated for each census tract for all the study years by geocoding the data that contained the street address of the offense location.

Alcohol outlet density: Three measures of exposure to alcohol outlets were included in the analysis: (1) a dichotomous measure of whether or not a liquor license was surrendered in the census tract following the May 1992 civil unrest, (2) the percentage of all off-sale and on-sale liquor licenses surrendered in the census tract, and (3) the annual off-sale and on-sale outlet densities from 1990–1999. Annual counts of liquor outlet licenses came from the California Department of Alcohol Beverage Control (ABC). A list of outlets that surrendered their licenses and a list of stores damaged as a result of the 1992 Los Angeles civil unrest were also obtained from the ABC. Alcohol outlets were classified based on their license to sell alcohol for on-premise (bars

DOI: 10.1002/sim

and restaurants) or off-premise (liquor stores, grocery stores, and convenience stores) consumption using license codes provided by the ABC. All unique address listings were geo-coded and mapped to the 1990 census-defined areas, and individual data sources were matched by census tract. Ninety-eight per cent of addresses were matched using Arcview 3.2 GIS software (ESRI Inc., Redlands, CA) along with Los Angeles County Topographically Integrated Geographic Encoding and Referencing (TIGER) street files from the 2000 census. The addresses that the computer was unable to match were hand placed with the help of an Internet mapping site (Mapquest) and a Thomas Guide map book.

Additional covariates: Additional tract-level covariates included in the analysis were: (1) annual percentage of African American residents, (2) annual percentage of Hispanic residents, (3) annual percentage of all males between the ages of 15 and 30 years, (4) extent of physical damage in the census tract, and (5) population density. The first four covariates are annual estimates available for the years 1990–1999 and are included to control for changes in tract composition over time, an endogenous change that could explain temporal changes in assault rates. For example, it is possible that changes in assault are the result of the movement of populations at higher or lower risk for violence into or out of particular study tracts over the course of the observation period.

The annual estimates of population distributions by age, race, and sex were obtained from the Los Angeles County Department of Health Services, with actual counts available for 1990 and 1995 and counts for the other years estimated from the birth and death records. The remaining socio-demographic data were obtained from the 1990 Census data for Los Angeles County. Information on damaged buildings [20] came, directly or indirectly, from four different sources: the Los Angeles City Department of Building and Safety, the Korean Central Daily, the California Insurance Commission, and the Compton Department of Building and Safety. Physical damage is measured as a binary indicator of any damage to a property in a tract due to the civil unrest. We also derived a measure of damage density (i.e. damage per square mile) calculated as the ratio of the number of unique addresses damaged in the 1992 civil unrest to the amount of land in the tract used for the commercial purposes. The denominator corresponds to land used for the commercial purposes because most of the damaged property was commercial. To compute the denominator, we estimated the proportion of commercial space in the tract using a land use file and multiplied it by the area of the tract in 1990 in square miles.

3. HIERARCHICAL ADDITIVE MODELING

In this section, the basic ideas of the MART and CAR methods are reviewed and then the hierarchical additive model is proposed. We describe the related algorithm used in model building and show how to explain the models.

3.1. Multiple additive regression trees

MART is a special case of the generic gradient boosting approach developed by Friedman [11]. Given n observations of the form $\{y_i, \mathbf{x}_i\}_1^n = \{y_i, x_{i1}, \dots, x_{ip}\}_1^n$ and any differentiable loss function $L(y, F(\mathbf{x}))$, MART considers the common problem of finding a function $F(\mathbf{x})$ mapping a p dimensional input vector $\mathbf{x}$ to response variable y, such that over the joint distribution of all $(y, \mathbf{x})$ values, the expected loss, $L(y, F(\mathbf{x}))$, is minimized. MART approximates the target function

DOI: 10.1002/sim

$F(\mathbf{x})$ by an additive expansion of trees

$$\hat{f}(x)=\sum_{m=1}^{M} v b_H(\mathbf{x};\boldsymbol{\gamma}_m) \tag{1}$$

where $b_H(\mathbf{x};\boldsymbol{\gamma}_m)$ is an H-terminal node tree (which partitions the input space into H-disjoint regions); $\boldsymbol{\gamma}_m$ is the parameter vector in building tree *m,* and $v\in(0,1)$ is the 'shrinkage' parameter $(0<v\leqslant 1)$ that controls the *learning rate* of the procedure. Empirical results have shown (see e.g. [11, 21] that small values of v *always* lead to smaller generalization error. The detailed algorithm of MART (for regression) is the following.

Algorithm 3.1 (*MART algorithm* (*Friedman* [*11*]))

(1) $\hat{f}_0(\mathbf{x})=\arg\min_{\gamma}\sum_{i=1}^{n} L(y_i,\gamma)$.

(2) Repeat for $m=1,2,\ldots,M$:

  (a) $\tilde{y}_i=-\left[\partial L(y_i,f(\mathbf{x}_i))/\partial f(\mathbf{x}_i)\right]_{f(\mathbf{x})=\hat{f}_{m-1}(\mathbf{x})}$, $i=1,2,\ldots,n$.

  (b) $\{R_{hm}\}_1^H=H$-terminal node tree on $\{\tilde{y}_{im},\mathbf{x}_i\}_1^n$.

  (c) $\gamma_{hm}=\arg\min_{\gamma}\sum_{\mathbf{x}_i\in R_{hm}} L(y_i,\hat{f}_{m-1}(\mathbf{x}_i)+\gamma)$.

  (d) $\hat{f}_m=\hat{f}_{m-1}+v\cdot\gamma_{hm}I(\mathbf{x}\in R_{hm})$.

(3) End algorithm.

Within each iteration m, a regression tree, whose splitting scheme $\{R_{hm}\}_{h=1}^H$ is optimized based on the negative gradient $\{\tilde{y}_i\}$ at its current estimate $\hat{f}_{m-1}$ (which is closely related to the *steepest-descent* minimization approach in function optimization), is fitted with an estimate γ_{hm} in each region. The value of M, i.e. the number of iterations or trees, can be chosen based on either cross-validation or monitoring the prediction performance on 'out-of-bag' samples with subsampling in each iteration (see [22]). Note that in practice, we can pre-specify the maximum depth D for an individual tree instead of the number of terminal nodes H. For example, the tree with $D=1$ (single-split trees with only two terminal nodes) fits an additive model without interaction and the MART-fitted model with $D=3$ is able to account for at most three-way interactions. For details of MART and gradient boosting, we refer the readers to the original paper by Friedman [11]. In this paper, MART is run by using the **gbm** package in **R**, produced by the Greg Ridgeway.

3.2. Conditional autoregressive model

We use the vector $\{\phi_{T_i,C_i}\}$ to capture spatial autocorrelations among areas C_i at time T_i, where $i=1,\ldots,n$ and n is the total number of observations; $T_i=1,\ldots,T$, where T is the total number of time slots and $C_i=1,\ldots,C$ where C is the total number of locations. A popular model for the spatial correlation arises by assuming that an area C_i is correlated with only the areas that are adjacent to it. Let $f(\mathbf{x})$, a function of the covariate vector $\mathbf{x}$, explores the association between $\mathbf{x}$ and the response variable y. We have the following model for y with a hierarchical structure on its mean function:

$$y_i\sim \mathrm{N}(\mu_i,\sigma^2) \quad \text{and} \quad \mu_i=f(\mathbf{x}_i)+\phi_{T_i,C_i} \tag{2}$$

DOI: 10.1002/sim

where y_i is the observed value of the response variable in the area C_i at time T_i. We apply a conditional autoregressive (CAR; [23]) structure for the spatial term ϕ_{T_i,C_i}. Let $j \sim i$ denote adjacency of regions i and j, and n_j be the number of tracts adjacent to tract j. The hierarchical CAR structure for $\{\phi_{T_i,C_i}\}$ has the form

$$\phi_{T_i,C_i=j}|\phi_{T_i,C_i\neq j} \sim \mathrm{N}\left(\sum_{k\sim j}\frac{1}{n_j}\phi_{T_i,k}, \frac{1}{n_j\tau_{T_i}}\right) \tag{3}$$

where τ_{T_i} is the precision parameter controlling the degree of spatial smoothing in T_i. We employ the Winbugs software to search for the estimates of ϕ. We use equation (3) as the prior distribution for ϕ and also, we assign the temporal smoothing term τ_{T_i} and the random precision term $1/\sigma^2$ as a noninformative uniform hyper distribution ranging from 0 to ∞. The maximum *a posteriori* (MAP) estimate of ϕ with this prior is equivalent to the MLE of our spatial terms with a hierarchical CAR structure. The estimates are called 'generalized MLE'. We obtain posterior distributions for ϕ via Markov chain Monte Carlo (MCMC) algorithms implemented in WinBUGS (free software available at http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml). Notice that we can also obtain MLE for ϕ through other iterative optimization algorithms.

3.3. The two-stage iteration algorithm

To combine the MART, which explores the variable relationship, with the CAR, which identifies a spatial autoregressive structure, we propose a two-stage iterative algorithm. In the first stage, explore the relationship between the covariates and response variable. In the second stage, we smooth the residual spatial correlations that have not been explained by the covariates. To avoid redundancy, the observations for those years that the covariates can satisfactorily explain most of the spatial correlation in the response variables are not analyzed in the second stage. We use the test statistic *Moran's I* [24] to test whether there is a spatial correlation remaining in y. An additive structure was assumed between $f(\mathbf{x})$, i.e. the effects on y from the covariate $\mathbf{x}$, and the remaining spatial correlations. In the algorithm, $\mathbf{x}$ is the vector of possible covariates, q is the count of iterations, Δ is a small constant set beforehand to control convergence, and δ is used to test convergence, i.e. measuring the relative difference in the expected values of the responses from the last iteration to the current iteration.

Algorithm 3.2 (*Two-stage analysis*)

(1) Let $\phi^0_{T_i,C_i}=0$ where $C_i\in\{1,\ldots,C\}$, $T_i\in\{1,\ldots,T\}$; $q=0$, $\delta=1000$, $\mu_{1i}=0$ and $i=1,\ldots,n$.
(2) If $\delta<\Delta$, go to step (3), otherwise $q=q+1$ and

- (a) Let $z_i=y_i-\phi^{[q-1]}_{T_i,C_i}$. Fit MART $f^{[q]}(\mathbf{x})$ on z and the covariates $\mathbf{x}$.
- (b) Let $e_i=y_i-f^{[q]}(\mathbf{x}_i)$, calculate the Moran's I of e_i within time slots 1 to T. Let S be the collection of time slots in which the spatial correlation test show a p-value smaller than 0.01 (this indicates a strong spatial correlation in e_i).
- (c) If S is empty, let $\phi^{[q]}_{T_i,C_i}=0$ and go to step (3); otherwise
  for the observations $i\in\{i:T_i\in S\}$, let the $f(\mathbf{x}_i)$ in equation (2) be $f^{[q]}(\mathbf{x}_i)$ and calculate the generalized MLEs of $\hat{\phi}_{T_i,C_i}$. Let $\phi^{[q]}_{T_i\in S,C_i}=\hat{\phi}_{T_i,C_i}$ and $\phi^{[q]}_{T_i\notin S,C_i}=0$.

DOI: 10.1002/sim

(d) Let $\mu_{0i} = \mu_{1i}$, $\mu_{1i} = f^{[q]}(\mathbf{x}_i) + \phi^{[q]}_{T_i, C_i}$ and let $\delta = \sum_{i=1}^{n} (\mu_{1i} - \mu_{0i})^2 / \sum_{i=1}^{n} \mu_{1i}^2$, go back to (2).

(3) Output the results from step q.

Little is known of the convergence properties of the above procedure. In proposing the iterated conditional modes, Besag [25] discussed the complication in parameter estimation when there are other parameters to be estimated besides the variance structure. Ideally, $f(\mathbf{x})$ and ϕ are estimated from the training data alone and the estimated values are used in the subsequent reconstructions. Typically, no training data are available and it is necessary to estimate $f(\mathbf{x})$ and ϕ as a part of the restoration procedure. Meng and Rubin [26] showed that under certain conditions, iterative conditional maximization converges to local maximizers. But the problem is even more complicated in the current analysis as the whole function (building of multiple trees) have to be estimated rather than just a few parameters. Our algorithm is essentially a backfitting process [13] with the MART and the CAR variance structure as two additive components. Buja *et al.* [27] proved the convergence of the backfitting process for a certain class of fixed, nonadaptive operators and the algorithm was well behaved in general [13]. In our analysis, we let $\Delta = 10^{-7}$. That is, if the relative difference, δ, in response means between two sequent iterations is less than 10^{-7}, we conclude that the algorithm converges.

3.4. Interpretation

One of the most important aspects in the interpretation of the results involves the identification of the relative importance of covariates in terms of their relative strength in predicting the response, and understanding their joint effects on the response. For tree-based methods, Breiman *et al.* [12] proposed a measure of importance $I_j^2(b_H)$ for each variable x_j, based on the number of times that variable was selected for splitting in the tree b_H weighted by the squared improvement to the model as a result of each of those splits. Friedman [11] generalized this importance measure to additive tree expansions by taking the average over the trees

$$I_j^2 = \frac{1}{M} \sum_{m=1}^{M} I_j^2(b_H(\mathbf{x}, \gamma_m)) \tag{4}$$

The measure (4) turns out to be more reliable than a single tree as it is stabilized by averaging. Since these measures are relative, we scale the measure so that the importance of all the variables sum to 100 per cent.

In addition to the importance measure, Friedman [11] also introduced a concept called *partial dependence* to describe the dependence of the fitted model on a subset of variables. Given any subset $\mathbf{x}_s$ of the input variables indexed by $s \subset \{1, \ldots, p\}$, the partial dependence is defined as

$$F_s(\mathbf{x}_s) = E_{\mathbf{x}_{\backslash s}}[f(\mathbf{x})] \tag{5}$$

where $E_{\mathbf{x}_{\backslash s}}[\cdot]$ means expectation over the joint distribution of all the input variables not indexed in s. In practice, partial dependence can be estimated from the data by

$$\hat{F}_s(\mathbf{x}_s) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(\mathbf{x}_s, \mathbf{x}_{i\backslash s}) \tag{6}$$

where $\{\mathbf{x}_{i\backslash s}\}_1^n$ are the data values of $\mathbf{x}_{\backslash s}$. The bootstrap method provides confidence intervals, and then we can make inferences on both the partial dependence and the importance measure.

DOI: 10.1002/sim

To explore the spatial autocorrelations, we draw a map with the MLEs of spatial random effects. The spatial random effects are residuals that are spatially correlated and cannot be explained by the covariates in the model. The map of spatial random effects could suggest various spatially varying covariates that are still missing in the model.

4. THE HIERARCHICAL ADDITIVE MODEL APPLICATION TO THE ALCOHOL-RELATED CRIME STUDY

In this section, the hierarchical additive model developed in Section 3 is used to analyze the data introduced in Section 2. The purpose of the analysis is to determine whether there is an association between the alcohol outlet measures and the assault. Furthermore, if there is an association, how it is related. In addition, we want to search for the other factors that are important in predicting the assault rates. In this analysis, the response variable is defined to be

$$y_i = \log\left[\frac{(\text{number of assault})_i + 0.0001}{\text{population}_i} \times 1000\right]$$

the assault rate per 1000 people. We add 0.0001 to the number of assault to avoid the complication when the number of assault is 0. The area unit in this analysis is census tract, ranging from 1 to 290, and the time unit is year.

We have 10 years data from 1990 to 1999. An exploratory analysis suggests that we use the possible covariates: (1) heterogeneity in race—the covariates represent the annual proportion of residents that are black, white, Asian, and Hispanic (the variable names in the model are 'black', 'white', 'asian', and 'hispanic'); (2) annual proportion of young males in census tract ('male_15_30'); (3) percentage of households in poverty ('poverty') in 1990; (4) damage level measuring the relative damage caused by civil unrest—the covariate is called 'damage', calculated as the number of damaged addresses in the tract divided by the tract area in square miles; (5) the years ('year') and (6) the alcohol outlet measures, including the on-premise, off-premise, total alcohol outlet density, and the indicator variables of the liquor license surrender during the civil unrest.

4.1. Relative variable importance and partial dependence

The following two models are fitted. In model A, the variable 'totaldensity', which is defined as the number of on-premise and off-premise alcohol outlets per roadway mile, is used to test the association between the alcohol outlets and assault. We also include the variables 'pctonsurryn' and 'pctoffsurryn', the percentages of on-premise and off-premise alcohol licenses surrendered in the 1992 civil unrest separately, to check if a sudden loss of alcohol outlets led to a change in assault violence rates. In model B, we use 'onsale' and 'offsale', the on-premise and off-premise alcohol outlet densities per roadway mile, to check whether on-premise or off-premise alcohol availability is more important in predicting the assault. In this model, we use two indicator variables 'onsurryn' and 'offsurryn', to indicate whether there is on-premise or off-premise alcohol license surrenders in the corresponding census tract in the 1992 civil unrest. In MART, we set the learning rate v at 0.001 and the maximum depth for each individual tree at 3, i.e. model counting up to three-way interactions.

DOI: 10.1002/sim

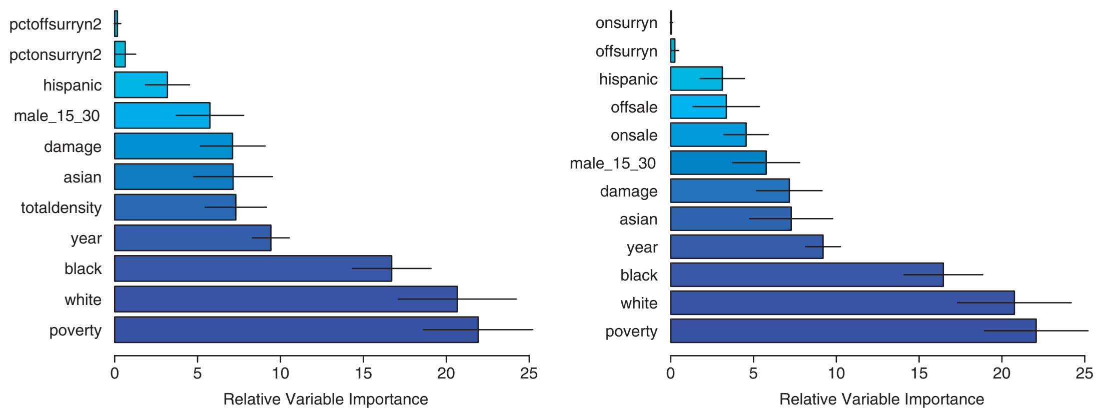

Figure 1. Relative variable importance in MART-fitted model A (left) and B (right). The straight lines show the 95 per cent confidence intervals of the importance measures.

Figure 1 shows the relative variable importance in MART-fitted models in Models A (left) and B (right), which consist of 4780 and 4519 individual trees, respectively. Note that the relative variable importance measures are based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and then averaged over all the trees (see Section 3.4). In Figure 1, we see that compared with other covariates in the model, the variable 'poverty' is the most important variable in predicting assault rates in both models. From the left panel of Figure 1, we find that alcohol availability ('totaldensity') is a relatively important variable, ranked fifth in the model, more important than the proportion of young males, the proportion of Asians and Hispanics and the damage level in the Civil Unrest. The percentages of on-premise or off-premise licenses surrendered in the civil unrest are relatively unimportant in the model. That is, we do not find important influence from a sudden drop in alcohol outlet density on the assaultive violence. When we consider the effect of on-premise and off-premise alcohol availabilities separately in model B, we find on-premise alcohol outlet density is a little bit more important than the off-premise outlet density (the right panel of Figure 1) in explaining the assault rates. But the difference is not significant by the overlap of confidence intervals of the importance measures of the two variables.

After establishing the relative importance of the explanatory variables, the nature of the dependence in the fitted model on any subset of explanatory variables is of interest. The partial dependence function can help us to graphically examine the dependence of a fitted model on a small subset of the variables. Figure 2 based on model A shows the partial dependence plots for the six most important variables. We establish the positive relationship between alcohol outlet density and assault rates in the Los Angeles area over the years 1990–1999: higher total alcohol outlet density is associated with higher assault rates. Also census tracts with relatively more blacks as well as poorer tracts tend to have a higher assaultive risk. The assault rates decrease with the percentages of young males in the tract and then increase a little (the partial dependence reaches the minimum when the percentage of young males is between 0.056 and 0.156) before stabilizing. Furthermore, the assault rates decreased over the years from 1990 to 1999 with a steeper decrease after 1993 (1 year after the civil unrest). The steeper decrease in assault rates lasted until the year 1998. This finding matches that from [9], where they obtain the results by model selection. Based on Figure 2,

DOI: 10.1002/sim

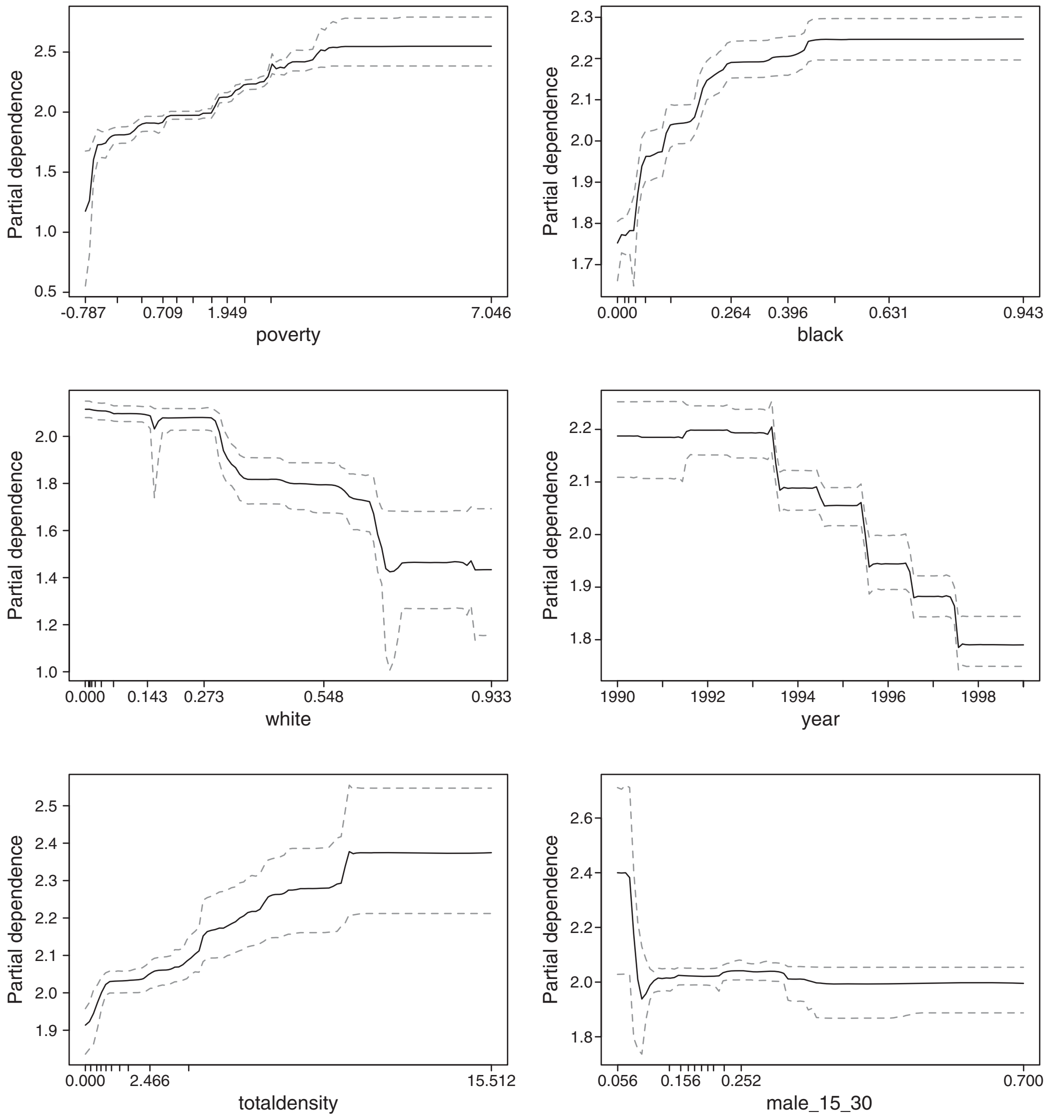

Figure 2. Partial dependence plots in MART-fitted model A. Dotted lines are the 95 per cent confidence intervals for the partial dependence.

detailed inference on how covariates are related with the assault rates could be drawn from the partial dependence of each covariates and their confidence intervals.

Figure 3 shows the two-dimensional partial dependence plot for *poverty* and *onsale* in the MART-fitted model B. We see that both *poverty* and *onsale* variables act positively on the response with no obvious interaction pattern. Friedman and Popescu [28] developed techniques that allow us to test the total interaction strength for each input variable. The procedure is essentially a variant of a permutation test. For details of the test procedure, we refer the readers to Section 8 in [28]. The

DOI: 10.1002/sim

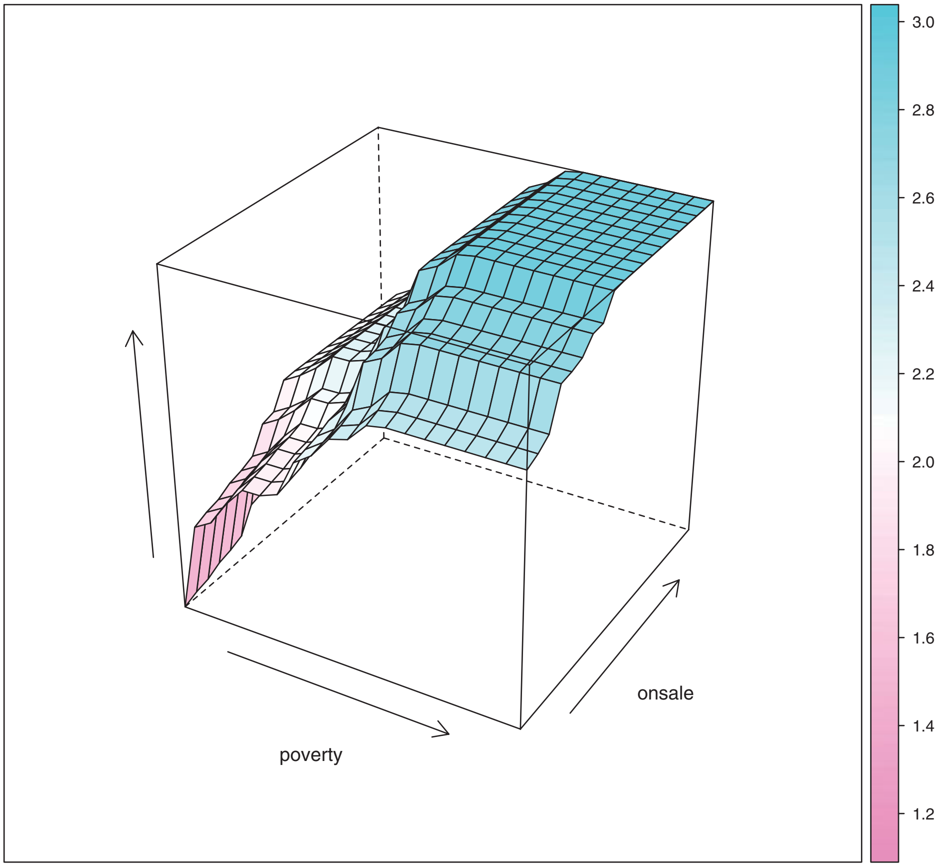

Figure 3. Two-dimensional partial dependence plot of *poverty* and *onsale* in MART-fitted model B.

procedure is applied to the fitted MART models A and B. We find that no significant interaction effect was observed for all the predictor variables.

4.2. *The spatial correlations*

In this section, all analyses are based on the MART-fitted model A. To check the spatial heterogeneity explained by our models, Table I shows the Morans' I, the index of spatial correlation, of the original log assault rate, the residuals after fitting with covariates, and the residuals after both spatial smoothing and covariate fitting over the 10 years. Note that after MART fitting with covariates, the remaining residuals in the years 1991 and 1996 have no significant spatial correlation, so the observations in those 2 years are not used in the second stage to fit the spatial errors. After the second stage of spatial smoothing, we find that most spatial correlations are explained by the hierarchical additive model. There are still spatial correlations remaining for the years 1990, 1992, and 1999, but the spatial correlation becomes negative. It seems that in these years, the spatial

DOI: 10.1002/sim

Table I. Moran's I and P-values of Spatial Correlation Testing: the 'Origin' column is the Moran's I of the original log assault rate; 'Res1' is the Moran's I of the remaining residuals after the covariates fitting, $y_i - f(x_i)$; and 'Res2' is the Moran's I of the remained residuals after the two-stage fitting, $y_i - \mu_i$.

Year	Origin	P-value	Res1	P-value	Res2	P-value
1999	0.46	<0.0001	0.13	1.23e−04	−0.09	0.017
1998	0.52	<0.0001	0.29	0.00e−00	−0.04	0.277
1997	0.50	<0.0001	0.25	1.46e−12	−0.02	0.688
1996	0.29	<0.0001	0.05	4.78e−02	0.05	0.048
1995	0.49	<0.0001	0.23	2.35e−11	−0.01	0.484
1994	0.49	<0.0001	0.24	4.98e−12	−0.04	0.335
1993	0.52	<0.0001	0.21	4.83e−09	−0.01	0.825
1992	0.30	<0.0001	0.14	1.77e−07	−0.16	0.000
1991	0.18	<0.0001	0.02	4.43e−01	0.02	0.443
1990	0.44	<0.0001	0.16	1.26e−06	−0.11	0.001

The P-values of the spatial correlation tests are shown in the right column to the corresponding Moran's I.

associations are overfitted. Note that these years are the years when the p-values of spatial tests are relatively large after fitting covariates, meaning there is less spatial correlation in the residuals. Also note that the year 1992 is when the civil unrest occurred.

Figure 4 maps the fitted assault rates in the civil unrest area in 1998. The upper panels separately map the fitted log assault rates using the hierarchical additive models, $f(x_i)+\phi_{T_i,C_i}$ as in model 2, and those fitted through MART, $f(x_i)$, only. The left upper panel reveals a higher level of assault rates in the middle to east areas, which comprise downtown LA and its immediate neighborhoods. The lower panels of Figure 4 present map residuals. The left panel is the fitted spatial error (ϕ_{1998,C_i}) and the right panel is the random residuals after all the model fitting, i.e. the raw log assault rate, y_i, minus the log-fitted assault rate, μ_i, at each census tract in 1998. We see that no obvious spatial correlations remain in the residuals after the two-stage analysis. The remaining spatial correlation in the residuals after the fitting of covariates using MART suggests the presence of unmeasured spatially varying covariates. One obvious candidate here would be the distance of each tract from the origin of the civil unrest (the intersection of Florence and Normandy). Additional possible factors are more social in nature, and are related to the fact that the southern region of the civil unrest area tends to be the most disadvantaged in the city. While measure of social economic status and ethnicity are already included in the model, other factors associated with concentrated disadvantage (i.e. family structure, ethnic isolation, low social capital) were not included and may contribute to the observed pattern. The maps from other years can also be drawn and analyzed to figure out the possible missing variables.

4.3. *Comparison of methods*

Yu *et al.* [9] analyzed the same data set using hierarchical linear regression models, with assumed linear effects on transformed variables and possible interactions. The spatial errors were also modeled using the CAR strategy in their paper. In their analysis, several models were compared, with the best model chosen in terms of Deviance Information Criterion (DIC) [29]. Here we compare their Bayesian hierarchical linear model with our hierarchical additive model in terms of the efficiency of explaining the spatial correlations and model explanation.

DOI: 10.1002/sim

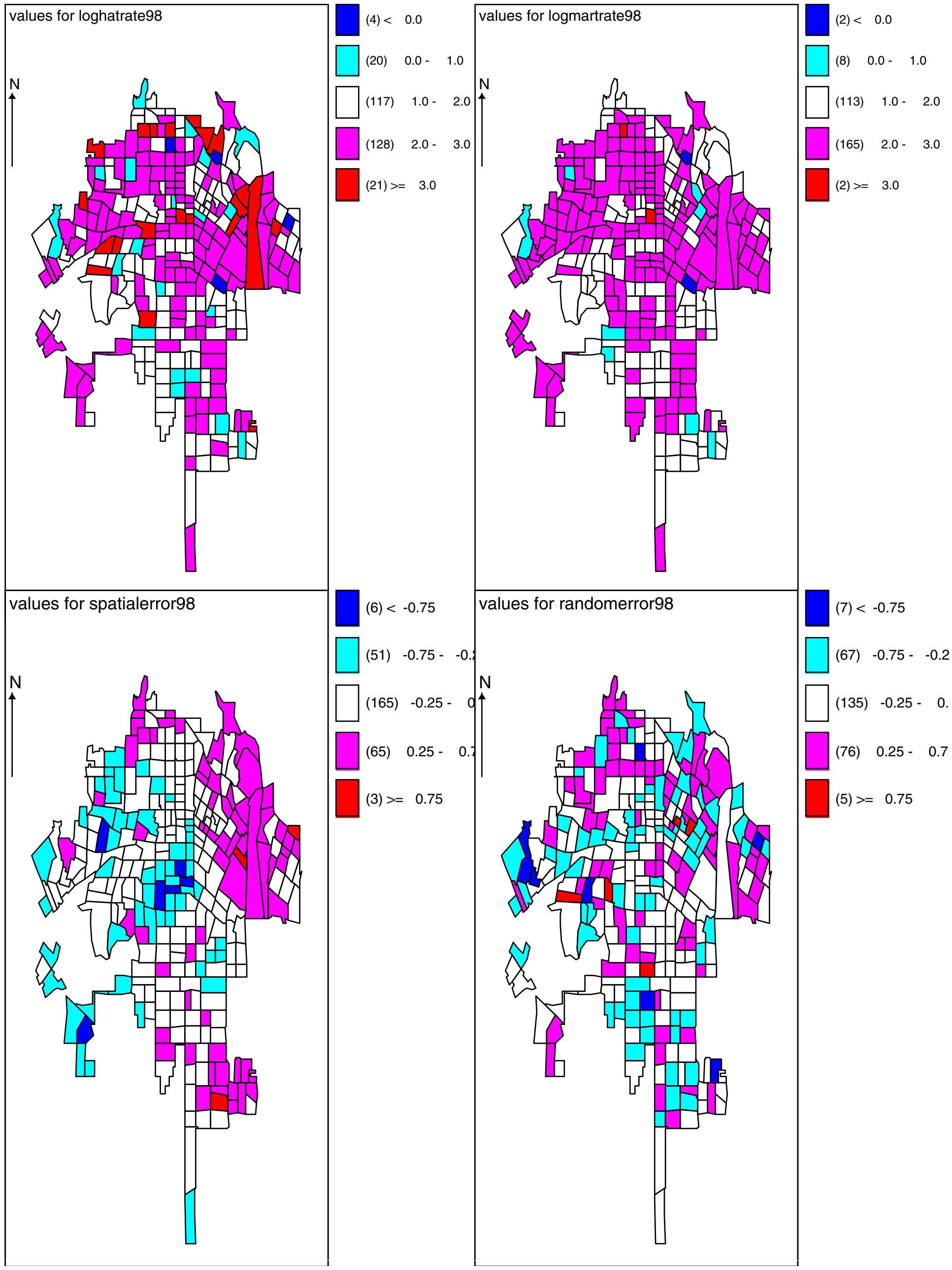

Figure 4. Maps of log assault rates on fitted data. The left upper map is the two-stage-fitted assault rate in 1998 and the right upper panel maps the MART-fitted log assault rates. The left lower panel maps the fitted spatial error and the right lower panel shows the random residuals after taking out the spatial errors.

Table II. Moran's I and P-values of spatial correlation testing for the residuals from the hierarchical linear model.

Year	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990
Moran's I	0.14	0.24	0.18	0.04	0.13	0.20	0.17	0.11	0.02	0.16
P-value	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.39	0.00

Moran's I is used to test the spatial correlation. Moran's I and the test p-value of the residuals after model fitting with the hierarchical linear model are shown in Table II, which should be compared with the last two columns of Table I. We see that there are still spatial correlations remaining in the residuals of the hierarchical linear regression. The hierarchical additive model does a better job in explaining the spatial correlations in this data set.

We also notice that the pD, explained as the effective number of model parameters [29], from the hierarchical linear model is 2356, while that from the hierarchical additive model is only 873. Both pDs account for the local shrinkage of the spatial random effects only. This means that most variances in the assault rates are explained by the spatial errors in the linear model, while covariates more efficiently explain the assault rates in the hierarchical additive model.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a hierarchical additive model strategy—using nonparametric method to look for relationships among the variables and utilizing a CAR model to address the spatial heterogeneity. We use this strategy to explore the relationship between the alcohol outlet density and the assault rates. We have demonstrated that the total alcohol outlet density is positively related to the assault rate over the study period. Decreasing alcohol outlet density was associated with decreasing assault rates. We capitalized on a natural experiment by the 1992 Los Angeles civil unrest, but did not find important consequences for the alcohol license surrenders on the assault rates when alcohol outlet density measures were included in the model. Maps were presented to show the distribution of the fitted assault rates, as well as residual maps to suggest possible missing covariates. Our method has been compared with the hierarchical linear model and showed superior performance in exploring important variables that explain the change in the assault rates.

As mentioned above, many other variables could have been included in our model and some lagged effect for coefficients could also be used to smooth the remaining spatial correlations. In addition, it might be of greater interest to model other types of assaultive violences with assault simultaneously. Our future research might include multivariate models to analyze different alcohol-related violences such as assault, homicide, rape, and robbery together. A possible solution is to use MART to explore the relationships between the different assaultive violence outcomes and the covariates separately and then use the multivariate intrinsic Gaussian CAR hyper-distribution on the random effect vector of different types of assaultive violence outcomes to explore the resulting residuals together. This could be easily realized through the 'mv.car' function in Winbugs. A final area of interest is to study the association between the alcohol outlet and mortality rates in the study region using the hierarchical additive model.

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

This paper is supported in part by the National Institute of Alcohol Abuse and Alcoholism, grant no. R01 AA013810_03 and the Publishing Arts Research Council, grant no. 98-1846389.

REFERENCES

1. Scribner RA, MacKinnon DP, Dwyer JH. The risk of assaultive violence and alcohol availability in Los Angeles county. *The American Journal of Public Health* 1995; **85**:335–340.
2. MacKinnon DP, Scribner RA, Taft K. Development of a city-level alcohol availability and alcohol problems database. *Statistics in Medicine* 1995; **14**:591–604.
3. Watts RK, Rabow J. Alcohol availability and alcohol-related problems in 213 California cities. *Alcoholism*: *Clinical and Experimental Research* 1983; **7**:47–58.
4. Scribner RA, Cohen D, Kaplan S, Allen SH. Alcohol availability and homicide in New Orleans: conceptual considerations for small area analysis of the effect of alcohol outlet density. *Journal of Studies on Alcohol* 1999; **60**:310–316.
5. Speer PW, Gorman DM, Labouvie EW, Ontkush MJ. Violent crime and alcohol availability: relationships in an urban community. *Journal of Public Health Policy* 1998; **19**:303–318.
6. Gorman DM, Speer PW, Gruenewald PJ, Labouvie EW. Spatial dynamics of alcohol availability, neighborhood structure and violent crime. *Journal of Studies on Alcohol* 2001; **62**:628–636.
7. Gruenewald PJ. Small area estimates from a geostatisical perspective: current status and future directions. *The Epidemiology of Alcohol Problems in Small Geographic Areas*. NIAAA Research Monograph, vol. 36. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism: Bethesda, MD, 2000; 245–270.
8. Zhu L, Gorman DM, Horel S. Alcohol outlet density and violence: a geospatial analysis. *Alcohol* 2004; **39**: 369–375.
9. Yu Q, Richard RA, Carlin BP, Theall K, Simonsen N, Ghosh-Dastidar B, Cohen D, Mason K. Multilevel spatiotemporal dual changepoint models for relating alcohol outlet destruction and changes in neighborhood rates of assaultive violence. *Geospatial Health* 2008; **2**(2):161–172.
10. Gruenewald PJ, Remer L. Changes in outlet densities affect violence rates. *Alcoholism*: *Clinical and Experimental Research* 2006; **30**:1184–1193.
11. Friedman JH. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* 2001; **29**:1189–1232.
12. Breiman L, Friedman JH, Olshen R, Stone C. *Classification and Regression Trees*. Wadsworth: Pacific Grove, 1984.
13. Hastie T, Tibshirani R. Bayesian backfitting. *Statistical Science* 2002; **15**:196–223.
14. Evans C. Public health impact of the 1992 Los Angeles civil unrest. *Public Health Reports* 1993; **108**:265–272.
15. Cohen D, Ghosh-Dastidar B, Scribner R, Miu A, Scott M, Robinson P, Farley TA, Bluthenthal RN, Brown-Taylor D. Alcohol outlets, gonorrhea, and the Los Angeles civil unrest: a longitudinal analysis. *Social Science and Medicine* 2006; **62**:3062–3071.
16. Hubler S. Extensive survey of riot area finds some surprises. *Los Angeles Times*, Los Angeles, 2002; 1.
17. Grills C, Bass K, Brown D, Akers A. Empowerment evaluation: building upon a tradition of activism in the African American community. In *Empowerment Evaluation*: *Knowledge and Tools for Self-Assessment and Accountability*, Fetterman D, Kaftarian S, Wandersman A (eds). Sage Publications: Thousand Oaks, CA, 1996; 123–140.
18. Berestein L. Rebuilding rules prevail merchants must hire guards, shorten hours if they want to sell liquor. *Los Angeles Times*, Los Angeles, 1994; B3.
19. Kang C. Store owners to fight restrictions on reopening business. *Los Angeles Times*, Los Angeles, 1994; B3.
20. Ong P. Losses in the Los Angeles civil unrest, April 29–May 1. *1992*: *Lists of the Damaged Properties and Korean Merchants and the L. A. Riot Rebellion*. UCLA: Los Angeles, 1993.
21. Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine* 2003; **22**:1365–1381.
22. Friedman JH. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 2002; **38**:367–378.
23. Besag J. Spatial interaction and the statistical analysis of lattice systems (with Discussion). *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*) 1974; **36**:192–236.

24. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall, CRC Press: London, Boca Raton, 2004; 71–75.
25. Besag J. On the statistical analysis of dirty pictures (with Discussion). *Journal of the Royal Statistical Society*: *Series B* 1986; **48**:259–302.
26. Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 1993; **80**:267–278.
27. Buja A, Hastie T, Tibshirani R. Linear smoothers and additive models. *The Annals of Statistics* 1989; **17**:453–555.
28. Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Technical Report*, Stanford University, 2005.
29. Spiegelhalter DJ, Best NG, Carlin BP, Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*) 2002; **64**:583–639.

DOI: 10.1002/sim