Exploring Bias in a Generalized Additive Model for Spatial Air Pollution Data

Timothy Ramsay,[1] Richard Burnett,[2] and Daniel Krewski[1]

[1]R. Samuel McLaughlin Centre for Population Health Risk Assessment, Ottawa, Ontario, Canada; [2]Health Canada, Ottawa, Ontario, Canada

During the past few years, the generalized additive model (GAM) has become a standard tool for epidemiologic analysis exploring the effect of air pollution on population health. Recently, the use of the GAM has been extended from time-series data to spatial data. Still more recently, it has been suggested that the use of GAMs to analyze time-series data results in air pollution risk estimates being biased upward and that concurvity in the time-series data results in standard error estimates being biased downward. We show that concurvity in spatial data can lead to underestimation of the standard error of the estimated air pollution effect, even when using an asymptotically unbiased standard error estimator. We also show that both the magnitude and direction of the bias in the air pollution effect depend, at least in part, on the nature of the concurvity. We argue that including a nonparametric function of location in a GAM for spatial epidemiologic data can be expected to result in concurvity. As a result, we recommend caution in using the GAM to analyze this type of data. *Key words:* air pollution, bias, concurvity, generalized additive models, PM_{10}, $PM_{2.5}$, spatial data, sulfate particles. *Environ Health Perspect* 111:1283–1288 (2003). doi:10.1289/ehp.6047 available via *http://dx.doi.org/* [Online 3 April 2003]

Spatial statistical methods are commonly used to examine associations between two variables that demonstrate spatial variability, with concordance between the patterns of spatial variation providing evidence of an association between these two variables (Ferrándiz et al. 1999). For example, Pope et al. (2002) demonstrated a strong concordance between ambient levels of fine particulate matter and mortality in large U.S. cities. In a later reanalysis of these data, Krewski et al. (2000) examined a number of different statistical methods designed to assess concordance in spatial patterns between urban air pollution and mortality, taking into account spatial autocorrelation at different levels of geographic scale. Both analyses have demonstrated significant associations between (principally cardiopulmonary) mortality and both fine particulate matter and sulfate particles. Robust associations between the gaseous copollutant sulfur dioxide were also found.

Although the reanalysis made progress toward understanding the influence of spatial autocorrelation on the effects of sulfate particles on mortality, the analytic methods did not allow for the possibility that the relationship between air pollution and mortality may display nonstationarity over space, with the air pollution effect constrained to be the same at different locations within the United States. A more flexible modeling strategy is needed to assess nonstationary relationships. Also, reliance on a single autocorrelation parameter may have effectively filtered variables that operate at a broad regional scale, such as sulfate, but may not have controlled autocorrelation from pollutants such as sulfur dioxide, which exhibits a more spatially concentrated or local distribution (Krewski et al. 2000). Preliminary work with more advanced

spatial models has been conducted by Burnett et al. (2001) and Cakmak et al. (In press).

The spatial analysis technique introduced by Burnett et al. (2001) and Cakmak et al. (In press) is based on a semiparametric generalized additive model (GAM). Dominici et al. (2002) have recently shown that S-Plus (Insightful, Seattle, WA, USA), a widely used statistical software package, can produce biased estimates of the linear parameters in semiparametric GAMs, and they have proposed two methods for avoiding this bias. Ramsay et al. (2003) have shown that the estimated standard errors produced by S-Plus can systematically underestimate the true variability of fitted linear parameters in semiparametric GAMs.

Dominici et al. (2003) addressed the problem of biased standard error estimates by showing that an alternative estimator of standard error, proposed by Hastie and Tibshirani (1990), is asymptotically unbiased. Furthermore, they have implemented in S-Plus a function called "gam.exact" that fits GAMs and uses this alternative method to estimate standard errors for the fitted parameters.

In this article, we present the results of two simulation studies designed to explore the impact of concurvity on a particular data set taken from the literature (Burnett et al. 2001). The results suggest that, even with the stringent convergence criteria suggested by Dominici et al. (2002), increased concurvity in the data used to fit a semiparametric spatial GAM leads directly to increased downward bias in the estimated standard error of the fitted linear parameter and, consequently, to inflated type I error in the standard significance test of this effect. This result is true both for the S-Plus standard error estimate and for the asymptotically unbiased alternative estimate.

The Generalized Additive Model

Suppose that y is a vector of observed responses and that $x_1 \ldots x_p$ are p vectors of independent variables. The GAM (Hastie and Tibshirani 1990) postulates that the response is additively related to the independent variables via the equation

$$g(y) = \mu = \sum_{i=1}^{p} f_i(x_i) + \varepsilon ,\qquad [1]$$

where the function g is called the link function and ε is a random error term. Each function f_i is assumed to be in some predefined space of functions $\mathcal{H}_i$. The function spaces most commonly chosen for $\mathcal{H}_i$ are linear functions and nonparametric functions defined by smoothers such as loess (locally estimated polynomial regression) or smoothing splines.

Because of the flexibility provided by allowing the f_i to be nonparametric functions, as well as the fact that it is easy to fit these models using standard statistical software such as S-Plus, the GAM has gained favor as a powerful analytical tool. One popular form of the GAM is the semiparametric model in which one function, f_1, is chosen to be linear. This form of the model can be expressed as

$$g(y) = \mu = \beta x_1 + \sum_{i=2}^{p} f_i(x_i) + \varepsilon .\qquad [2]$$

Because of the ease of interpretation of a linear parameter, the semiparametric model is particularly attractive when x_1 is the primary variable of interest and $x_2 \ldots x_p$ are included to control for the effects of various nuisance variables. S-Plus provides a variance estimate for β, thus allowing the use of a standard *t*-test for the significance of the effect of x_1.

The back-fitting algorithm used by statistical software such as S-Plus to fit GAMs is iterative and uses several convergence parameters to determine when it terminates. Dominici et al. (2002) have recently shown that using the default convergence parameters

in S-Plus may lead to biased parameter estimates. This problem, they suggest, can be corrected either by using stricter convergence criteria or by substituting parametric functions for the nonparametric smoothers in the GAM definition. We (Ramsay et al. 2003) have shown that the S-Plus estimate of standard error for the linear parameter of a semiparametric model can be biased, thus invalidating the results of the corresponding significance test. This bias may also be avoided by using parametric, rather than nonparametric, smoothers.

The term "concurvity" refers to the presence of a specific type of (approximate) additive dependence among the independent variables in a GAM. For the general GAM in Equation 1, an additive dependence of the form

$$h_1(x_1) \approx \sum_{i=2}^{p} f_i(x_i) \qquad [3]$$

constitutes concurvity if each h_i is a function of the type specified by the preselected function space $\mathcal{H}_i$. This implies that concurvity is highly dependent on the particular GAM being considered. The approximate equality ($\approx$) in Equation 3 indicates that the two sides of the equation are positively correlated.

Because multicollinearity can be defined as the existence of a relation of the form

$$\beta_1 x_1 \approx \beta_0 + \sum_{i=2}^{p} \beta_i x_i,$$

concurvity is really just the nonparametric analogue of multicollinearity in regression analysis. In the case of the semiparametric model (Equation 2), a concurvity relation involving x_1 can be simplified to

$$x_1 \approx \sum_{i=2}^{p} g_i(x_i). \qquad [4]$$

An exact concurvity relation causes the GAM to be unidentifiable (Ramsay et al. 2003).

As with linear models and multicollinearity, concurvity leads to variance inflation for the fitted parameters in a GAM. This effect differs from that of multicollinearity, however, in that the variance inflation is not reflected in the standard error estimates produced by S-Plus. Furthermore, the S-Plus gam function does not provide any diagnostic tests for assessing the presence of concurvity. A consequence of this inability to detect concurvity is that using the S-Plus standard error estimate to test the significance of a fitted linear parameter can result in inflated type I error.

In theory, for any GAM in which the back-fitting algorithm converges, there exists a matrix $\mathbf{R}$ such that the fitted values $\hat{\mu}$ are given by the relation $\hat{\mu} = \mathbf{R}z$, where $z = g(y)$ is the vector of observations transformed by the link function. In addition, for each additive component of the model, there is a matrix $\mathbf{R}_i$ such that the fitted additive component $\hat{f}_i(x_i)$ can be expressed as $\hat{f}_i(x_i) = \mathbf{R}_i z$. Unfortunately, the back-fitting algorithm does not produce these matrices, and they are therefore not available for estimating the standard error of the fitted additive components.

Hastie and Tibshirani's (1990) monograph on GAMs proposed a method for estimating the matrices $\mathbf{R}$ and $\mathbf{R}_i$, but this algorithm was considered too computationally expensive to be implemented in the S-Plus gam function. Instead, the gam function uses an approximation to this variance (Chambers and Hastie 1992); this approximation appears to perform well in the absence of concurvity, but it systematically overestimates the precision when concurvity is present.

In the wake of the discovery of the impact of concurvity on the standard error estimates of the gam functions, Dominici et al. (2003) have developed a new S-Plus function, gam.exact, which fits univariate GAMs and uses Hastie and Tibshirani's (1990) approximations to the matrices $\mathbf{R}$ and $\mathbf{R}_i$ to estimate the standard errors of the fitted parameters. They show that the standard error estimates produced by the gam function are asymptotically biased, but that those produced by gam.exact are not. The gam.exact function software is available on the Internet (Internet-based Health & Air Pollution Surveillance System 2002). Unfortunately, the current implementation of gam.exact will not fit GAMs containing functions of more than one variable.

Ramsay et al. (2003) suggest that in GAMs for time-series data, the standard error bias of GAMs may be avoided by using fully parametric smoothers in place of nonparametric smoothers. This solution is somewhat difficult to apply to spatial GAMs, unfortunately, because the gam function provides no parametric method for smoothing two variables (the *x*- and *y*-coordinates) simultaneously. Also, it is our experience that bivariate parametric smoothers often require substantially more degrees of freedom than do bivariate nonparametric smoothers.

Exploring the Impact of Concurvity

The simulation studies presented in this section were designed to explore the impact of concurvity on a particular spatial data set previously analyzed by our research team (Burnett et al. 2001) as the second stage of a two-stage analysis relating air pollution to mortality. In the first stage of the analysis, survival data from a cohort of more than 500,000 subjects in 144 U.S. metropolitan statistical areas were fit, via a Cox proportional hazards model, to a collection of both individual-level and community-level variables. The fitted values consisted of the log-relative risk of mortality,

along with a variance estimate, for each of the 144 cities. The second stage employed a spatial GAM to relate mortality to air pollution by modeling the log-relative risks as the sum of a nonparametric function of location and a linear function of (average) airborne sulfate particle concentration.

The model used was

$$r = f(x, y) + \beta p + \eta + \varepsilon, \qquad [5]$$

where (x,y) was the spatial location of a given city, r was the associated log-relative risk of mortality, and p was the average concentration of airborne sulfate particles. The parameter of interest, β, reflects the effect of air pollution on mortality. The function f was included to control for unknown confounding effects that vary smoothly over space. In a simple linear regression model, these confounding effects produced autocorrelated residuals. The first of the two components of variance, η, was assumed to have mean zero and variance θ. The error due to estimating the relative risks from the stage 1 model, ε, was assumed to have mean zero and variance ν. The variance parameter θ was estimated from the data, but ν was taken to be the model-based variance estimate of $\hat{r}$ obtained from the fitted Cox proportional hazards model.

Burnett et al. (2001) used the gam function in S-Plus to fit a smooth function f and a linear parameter $\beta = 0.0087$ to the data, with an estimated standard error of 0.00277. The fitted parameter β represents the estimated increase in log-relative risk due to increasing the concentration of airborne sulfate particles by 1 μg/cm^3. The estimation procedure was iterative, with a Newton-Raphson update of the unknown variance parameter θ computed at each iteration.

To test for concurvity, as defined in Equation 4, we fit the model $p = h(x,y)$ to these data, with h modeled as a loess function. The fitted function $\hat{h}$ was highly significant, with a squared correlation coefficient of 0.57 between p and $\hat{h}(x,y)$. This value, which we refer to as the concurvity $\mathbf{R}^2$ statistic for p in Equation 5, indicates that 57% of the variation in p can be explained by a smooth function of location and thus tells us that there is a moderate degree of concurvity in the data. In fact, because common sense tells us that air pollution varies smoothly over space, some degree of concurvity is to be expected in these data relative to Equation 5.

Methodology: simulated air pollution data. Both of our simulation studies investigate the effect of unknown concurvity by fitting a slightly simplified version of Equation 5 to a large number of simulated data sets with random concurvity. Because the mechanical details of the two studies are similar, we begin by giving the details of our method for simulating

GAM models and data sets with random concurvity. Although we focus exclusively on the data analyzed by Burnett et al. (2001), this method could easily be modified to suit data for any nonparametric GAM.

The simulation procedure consists of five distinct steps: steps 1–3 generate independent variables with concurvity, and steps 4 and 5 generate response data.

Step 1 generates 144 random locations $(\bar{x}_i, \bar{y}_i)$, for $i = 1, \ldots 144$, to represent the cities.

Step 2 generates a set of completely random air pollution variables, $\bar{p}_i$, and a smooth function $\bar{h}$. The $\bar{p}_i$ are mutually independent pollution values for the 144 cities and hence display no concurvity. The function $\bar{h}$, generated using the random surface generation algorithm given in the Appendix, defines a set $\bar{h}(\bar{x}_i, \bar{y}_i)$ of spatially dependent air pollution values with extreme concurvity.

Step 3 combines the two sets of pollution variables into a single set of new pollution values $\bar{p}_{ij}$. This step depends on the value of the concurvity parameter j to determine the degree of concurvity in the simulated data. After assigning j a value between 0 and 1, the $\bar{p}_{ij}$ are defined by the relation

$$\bar{p}_{ij} = \left(1 - \sqrt{j}\right)\bar{p}_i + \sqrt{j}\ \bar{h}\left(\bar{x}_i, \bar{y}_i\right). \qquad [6]$$

Setting $j = 0$ implies that $\bar{p}_{ij} = \bar{p}_i$ and, hence, that the $\bar{p}_{ij}$ are concurvity-free. Conversely, when $j = 1$, the function $\bar{h}$ explains 100% of the variation in the $\bar{p}_{ij}$. The smooth function $\bar{h}$ is not a loess surface of the sort used to fit f in Equation 5, because this would imply perfect concurvity and an unidentifiable model. However, setting $j = 1$ still results in a data set with extreme concurvity (on average, $R^2 = 0.94$). The definition in Equation 6 is motivated by the observation that the $\bar{p}_{ij}$ lie on a smooth surface whose spatial autocorrelation function is equal to j times that of $\bar{h}$.

Step 4 prepares to generate log-relative risks by randomly generating a function $\bar{f}$, again using the algorithm in the Appendix.

Step 5 generates a set of independent, normally distributed random errors $\bar{\varepsilon}_i$ with mean zero. These random errors are combined with the function $\bar{f}$, pollution variables $\bar{p}_{ij}$ and a constant slope $\bar{\beta}$ to define log-relative risks via the model

$$\bar{r}_i = \bar{f}\left(\bar{x}_i, \bar{y}_i\right) + \bar{\beta}\ \bar{p}_{ij} + \bar{\varepsilon}_i. \qquad [7]$$

Note that Equation 7 is a simplification of Equation 5: it is missing the second variance component η.

The end result is a randomly generated data set with a specified degree of concurvity, for which the true value of each of the parameters in Equation 7 is known. For convenience, we will refer to the combination $\{\bar{h}, \bar{x}_i, \bar{y}_i, \bar{p}_i\}$ used to generate air pollution values with

concurvity as the preconcurvity data set, to the combination $\{\bar{f}, \bar{x}_i, \bar{y}_i, \bar{p}_{ij}\}$ used to generate the responses as the simulation data set, and to the combination $\{\bar{r}_i, \bar{x}_i, \bar{y}_i, \bar{p}_{ij}\}$ of variables used to fit the model as the simulated data set. This distinction will be important when we describe the second simulation study.

All of the randomly generated quantities in this procedure were sampled from distributions that approximately mirrored the distributions of the original variables. In step 1, the locations $(\bar{x}_i, \bar{y}_i)$ were chosen to be independently and uniformly distributed over the range of the 144 city locations. In step 2, the independent pollution measurements were uniformly distributed over the range of the observed pollution values, and the function $\bar{h}$ was scaled so that the range of the $\bar{h}(\bar{x}_i, \bar{y}_i)$ coincided with the range of the observed pollution values. The final pollution values $\bar{p}_{ij}$ were again scaled so that their range was the same as that of the observed pollution. The fitted model from Burnett et al. (2001) is used to define the surface $\bar{f}$ and errors $\bar{\varepsilon}_i$ in steps 4 and 5. The function $\bar{f}$ was scaled to have the same range as the fitted surface $\hat{f}$. The variance of the random errors of step 5 was set to be the sum of the fitted variance parameters $\nu + \theta$.

First simulation study: standard error. Although we have previously shown that the S-Plus standard error estimates can be biased (Ramsay et al. 2003), we did not clearly show that this bias is affected by the degree of concurvity in the data. The first simulation study was conducted to verify that increased concurvity does result in increased standard error bias and, therefore, to inflated type I error. For each j in the set $\{0, 0.1, 0.2, \ldots, 1\}$, we independently generated 1,000 simulated data sets in which the true value of β was zero. By fitting Equation 7 to each data set and testing the null hypothesis that $\beta = 0$, we obtained an estimate of the type I error as a function of j.

The results of this set of simulations are presented in Figure 1. The x-axis in both Figure 1A and Figure 1B is the concurvity parameter j from Equation 6. The type I errors in Figure 1A were obtained by testing the significance of β using the S-Plus standard error estimate at the 0.05 level of significance; the type I error increases dramatically with the degree of concurvity. Interestingly enough, the observed type I error at $j = 0$ was 0.081, which, viewed as a binomial statistic ($n = 1,000$, $p = 0.05$), corresponds to a p-value of about 10^{-5}. This indicates that the type I error is somewhat inflated even when there is no concurvity in the data.

In Figure 1A, 95% of the observed concurvity statistics fell within the interval defined by the error bars. Recall that the concurvity statistic is the squared correlation between the $\{\hat{p}_{ij}\}$ and the $\bar{p}_{ij}$, where the $\hat{p}_{ij}$ are the fitted

values obtained by modeling the $\bar{p}_{ij}$ as a loess function of the $(\bar{x}_i, \bar{y}_i)$. It should be noted that although both the type I error and the concurvity R^2 statistic are measured on a scale of 0 to 1, the two curves in Figure 1A do not measure the same thing. We include them in the same graph to suggest that the concurvity R^2 statistic may be a useful, albeit imperfect, tool in assessing the presence of concurvity in data.

One argument that the R^2 statistic shows significant concurvity in the data is as follows. The mean observed value of R^2 in those models for which $j = 0$ was 0.08, with a largest observed value of slightly less than 0.2. Because the R^2 statistic for the data of Burnett et al. (2001) exceeded 0.2, it is unlikely that these data were concurvity-free.

In Figure 1B, the sample standard deviation in $\hat{\beta}$ as a function of the concurvity coefficient is in contrast to the S-Plus estimate of the standard deviation of $\hat{\beta}$, and indicates that the true (empirical) standard error of the estimate of β increases substantially with increasing concurvity, whereas the estimated standard error produced by the S-Plus gam function is increased by comparatively little. Again, the error bars include 95% of the observed standard error estimates. When the degree of concurvity was high, the S-Plus gam function dramatically underestimated the true standard error.

The average concurvity R^2 statistic for $j = 0.3$ was 0.56, about the same as that observed in the data analyzed by Burnett et al. (2001). For the 1,000 data sets simulated with $j = 0.3$, the type I error was 0.246, the sample standard deviation was about 0.0063, the average estimated standard error estimate was about 0.0036, and 95% of the standard error estimates were < 0.0052. For the entire set of 11,000 fitted models, the estimated number of degrees of freedom used to fit the model ranged from 8.3 to 9.5 (nonparametric models can use fractional degrees of freedom), so overfitting was unlikely to have been an issue.

Second simulation study: bias. The results of the second simulation study revealed three important features of the effect of simulated concurvity. First, the bias in $\hat{\beta}$ varied greatly with the parameters used to generate the response. Second, the alternative standard error estimator, although better than the default estimate provided by S-Plus, still did not provide a satisfactory estimate of the standard deviation of $\hat{\beta}$. Third, most of the variance observed in $\hat{\beta}$ in the first simulation study was due to bias.

This simulation study can be viewed as (roughly) following a two-way factorial design, with one factor consisting of 30 different preconcurvity models and the other consisting of 11 different degrees of concurvity. The precise design is as follows. First, we executed steps 1 and 2 of the simulation procedure (described

above) 30 times. Then, for each of these 30 preconcurvity data sets, we executed the third step 11 times (once for each j in the set {0,0.1,0.2, ... ,1}), and executed step 4 once. This resulted in 11 simulation data sets for each preconcurvity data set, differing only in the value of the concurvity parameter used to construct the air pollution values. Then, for each of the 330 resulting simulation data sets, we executed the fifth step 100 times, to produce 100 different simulated data sets. This time, we used the fitted value of 0.00887 from the original analysis for the slope $\hat{\beta}$.

Besides the fact that $\hat{\beta}$ is not zero, this simulation study differs from the previous one in that the first simulation did not replicate response data for a given set of simulation data. In other words, the first simulation generated one set of responses from each simulation data set, whereas the second simulation generated 100 different data sets from each simulation data set. As a result, we were able to explore how the bias in $\hat{\beta}$ changed with different patterns of concurvity as well as with different degrees of concurvity. In addition to the fitted slope and standard error, we also computed the alternative standard error estimate (Hastie and Tibshirani 1990) for each of the 33,000 fitted models. The results of this simulation are summarized in Figures 2 and 3. The small points represent one hundred replications from a single simulation. Simulation data sets constructed from the same preconcurvity data set are connected by lines, and the large blue points in Figure 3 represent the overall type I error for a single value of the concurvity parameter.

Figure 2 shows the observed relative bias for each of the 30 models at each of the 11 levels of concurvity. The variance of the bias increases with concurvity and the direction of the bias can be positive or negative.
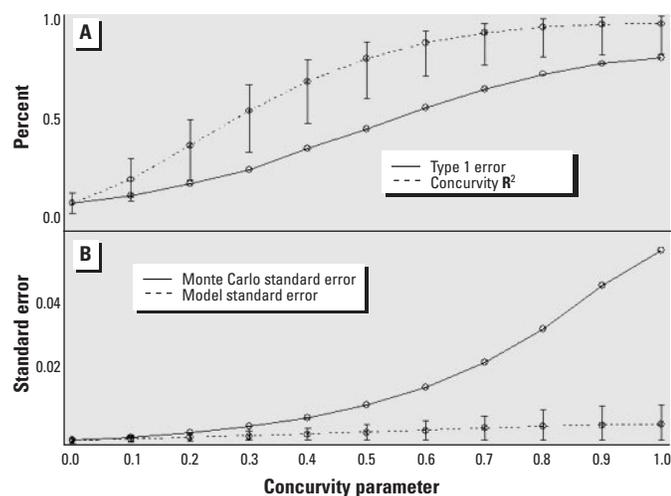
Figure 3 shows the type I errors observed when using t-tests based on the S-Plus estimate and the alternative estimate, respectively, to test the true null hypothesis that $\beta = 0.0087$ at the 0.05 level of significance. The performance of the alternative standard error estimate was superior to that of the estimate provided by the S-Plus gam function; at each level of concurvity, the test based on the S-Plus estimator resulted in a larger overall type I error. One interesting feature of these results is that using the S-Plus estimator to test the significance of the pollution effect resulted in an inflated overall type I error even when there was no concurvity. The alternative estimator, on the other hand, performed as expected when there was no concurvity in the data. At each level of concurvity, there were some estimators for which the alternative estimator did not inflate the type I error rate; for several of the models the type I error was not inflated at any level of concurvity. We conclude that, although the alternative estimator was better than the S-Plus estimator, neither estimator was reliable in presence of concurvity.

As mentioned above, the average R^2 statistic was about the same as that observed in the real data when $j = 0.3$. The results for the 30 simulated models with $j = 0.3$, therefore, seem likely to best represent the type of bias we might expect to be present in the fitted GAM produced from those data. The observed relative bias for those 30 models ranged from –1.53 to 1.83; 21 of 30 had magnitude > 0.2, and 12 had magnitude > 0.5. Two-thirds of the relative biases were negative, and one-third were positive (over all values of j, about 61% were negative).

Using a set of 30 independent Kolmogorov-Smirnov goodness-of-fit tests, we failed to reject the null hypothesis that the fitted $\hat{\beta}$ values were normally distributed for a given

model when $j = 0.3$ (only two were significant at the 0.05 level, corresponding to a binomial p-value of 0.19). We therefore assumed normality and tested the 30 biases for significance using 30 separate t-tests. The null hypothesis of no bias was rejected for 27 of the 30 models at the 0.01 level of significance.

On average, when we executed the same simulation using the standard convergence criteria provided by S-Plus, the observed relative bias was about 40% larger than that observed with the stricter convergence criteria suggested by Dominici et al. (2002). This suggests that if the bias is caused by a lack of convergence, then the likelihood surface in the neighborhood of the maximum likelihood must be extremely flat.

Using the S-Plus standard error estimate to test the true null hypothesis that $\beta = 0.0087$ resulted in an observed type I error rate > 0.05 (the expected value) for 24 of the 30 models for which $j = 0.3$. For 13 of the models, the observed error rate was > 0.2, and for 6 it was at least 0.5; the overall type I error rate was 0.29. By contrast, when the alternative standard error estimate was used in place of that provided by S-Plus, only 6 of the 30 models had an observed error rate > 0.05 (five had error rates > 0.2, and three were at least 0.5), and the overall type I error was 0.12.

Interestingly enough, both the S-Plus and the alternative standard error estimators appeared to systematically overestimate the standard deviation of $\hat{\beta}$. On average, for the 30 models where $j = 0.3$, the S-Plus standard error estimate was about 17% larger than the sample standard deviation, and the alternative estimate was about 100% larger.

At first glance, the observation that the variability of $\hat{\beta}$ was underestimated in the first simulation study but overestimated in the second may appear to be contradictory. The

Figure 1. Results of first simulation study as function of concurvity. (*A*) Type I error and concurvity statistic. (*B*) Empirical and estimated standard error. Error bars indicate the interval containing 95% of the observations.
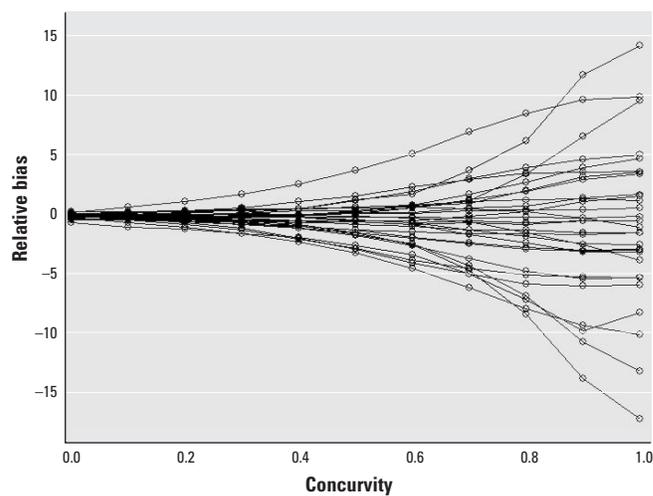
Figure 2. Relative bias, second simulation study. Points represent bias for one simulation data set (100 replications). Lines connect simulation data sets constructed from the same preconcurvity data set.

explanation is that, because of the bias in $\hat{\beta}$, the sample standard deviation of $\hat{\beta}$ is measuring different types of variability in the two simulation studies. The standard deviations in the second study represent the variability of $\hat{\beta}$ conditional on the simulation model, whereas the standard deviations in the first reflect the variability in $\hat{\beta}$ over all possible models.

The difference between the standard deviations in the two simulations can perhaps be more easily understood if we express the unconditional variance of $\hat{\beta}$ in terms of its conditional distribution. Letting Θ denote the set $\{\hat{f},\hat{h},\bar{x},\bar{y},\bar{p}\}$, consisting of the simulation data together with the concurvity relationship, we can write

$$V\big[\hat{\beta}\big] = V\big[E\big(\hat{\beta}|\Theta\big)\big] + E\big[V\big(\hat{\beta}|\Theta\big)\big],$$

where V is variance and E is expectation. The first term on the right-hand side of this expression is just the variance of the bias. The observed variance in the first simulation study was the unconditional variance $V[\hat{\beta}]$, whereas the observed variance in the second simulation study was $V[\hat{\beta}|\Theta]$.

Discussion

The results of our simulation studies depend on the methodology we chose to generate the artificial data. Although the surfaces produced by our random surface algorithm represent plausible patterns of spatial variation, there are other equally plausible patterns that it cannot reproduce. In the context of GAMs for time-series data, Dominici et al. (2002) found that the bias varied inversely with the

size of β. It is likely, therefore, that the bias in spatial GAMs would be negligible if the true effect of interest is large enough.

Nonetheless, we can still use these results to draw some general conclusions about concurvity and spatial GAMs. Concurvity in the data for a spatial GAM can lead to bias in the fitted linear parameter; this bias, if it occurs, probably depends on complex interrelationships among the independent variables. Because of the variability in the bias, the variance of the fitted linear parameter can be inflated in the presence of concurvity. Because this inflation is not properly captured by either of the currently available standard error estimators, the use of either estimate of the standard error to perform a *t*-test of significance on the fitted linear parameter can result in an inflated type I error.

Positive spatial autocorrelation in a variable α implies that some of the variation in α can be explained by modeling α as a smooth function of location. If a linear function (or any smooth function) of a spatially autocorrelated variable α and a smooth function of location are both included as additive components in a GAM, there will be some degree of concurvity in the data. Because most epidemiologic variables can be expected to exhibit spatial autocorrelation, it follows that concurvity is a particular problem for spatial GAMs using a smooth function of location to account for autocorrelation in the residuals.

In our study, the data sets that seemed to most resemble the data analyzed by Burnett et al. (2001) were those in which the concurvity parameter *j* had the value 0.3. For those simulations, the relative bias ranged from –1.53 to

1.83, and the overall type I error when using the S-Plus standard error estimate to test the significance of $\hat{\beta}$ (at the 0.05 level) was 0.25. Using the alternative estimator of Hastie and Tibshirani (1990) resulted in a type I error of 0.12. Because our model was a simplification of that used in the original analysis, these results may not accurately reflect the bias in their fitted model. For example, more than 10% of the R^2 statistics observed when *j* = 0.4 were less than the statistic for the original data. For those models with *j* = 0.4, the relative bias ranged from –2.2 to 2.7, the S-Plus type I

Appendix: Random Smooth Surfaces

To generate spatial data with truly random concurvity, it is necessary to generate random surfaces. Although any method for doing so can only generate a small subset of all possible smooth surfaces, we believe that the surfaces produced by the following method are both realistic enough and varied enough to provide good insight into the effect of unknown concurvity on the GAM estimation process.

The algorithm defines a random surface as the sum of a number of gaussian "bump" functions of the form

$$b(x,y) = a \exp\left\{\frac{1}{h}\left[\frac{(x-c_1)^2}{v_1} + \frac{(x-c_2)^2}{v_2}\right]\right\}.$$

[8]

The parameter *a* determines the amplitude of the bump and was sampled from the uniform distribution on the union of the intervals [0.5,1] and [–1,–0.5]. The parameter (c_1,c_2) determines where the bump is centered and was sampled from the uniform distribution on the unit square. The parameter (v_1,v_2) determines the ellipticity of the bump, or how far the bump deviates from being perfectly circular, and was sampled from the uniform distribution on the square [0.2,1] × [0.2,1]. Finally, the parameter *h* determines the width of the bump and was sampled from the uniform distribution on the interval [0.05,1].

Each random surface was generated from 10 independent random bumps of the form described in Equation 8. Each bump was randomly rotated by an angle selected from the uniform distribution on [0,2π], and the resulting bumps were summed to make one function. This function was evaluated on the data by transforming the unit square to cover the range of the $(\bar{x}_i,\bar{y}_i)$ values found in the data and was then rescaled to have the correct range. The rescaling step merely involved multiplication by a constant.
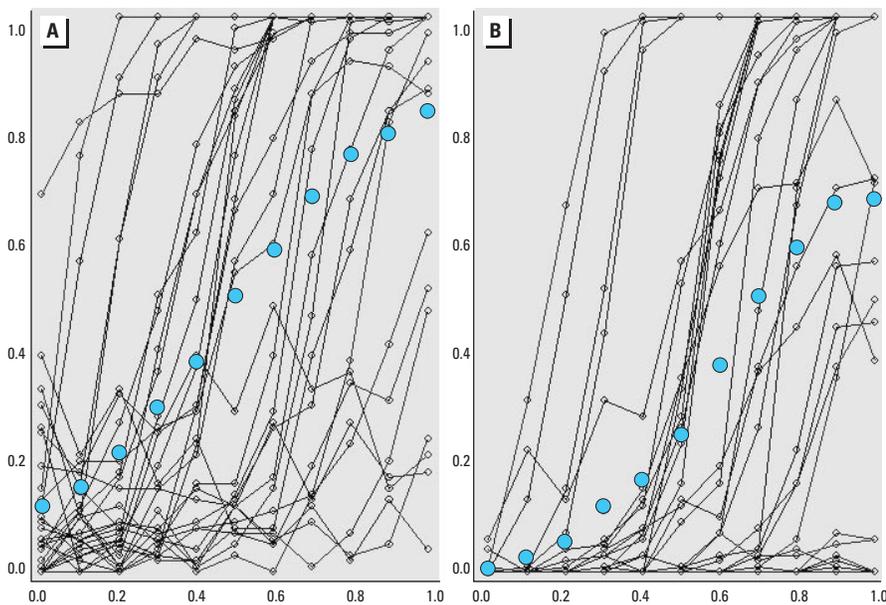
Figure 3. Type I error, second simulation study. (*A*) Default standard error versus concurvity. (*B*) Alternative standard error versus concurvity. Blue data points represent overall type I error for a single value of the concurvity parameter, and black points represent individual simulation data sets.

error was about 0.35, and the alternative type I error was about 0.17.

Although S-Plus has figured prominently in the discussions of Dominici et al. (2002, 2003) and Ramsay et al. (2003) concerning bias in GAMs, there is no reason to believe that these bias issues are specific to the S-Plus software. Rather, we believe the bias problems to be intrinsic features of nonparametric GAMs fitted in the presence of concurvity. In view of the conclusions of Dominici et al. (2002)—that using parametric functions in the time-series analyses of air pollution minimizes the bias issue—one possible solution to the bias problem in spatial GAMs is to devise a way to use parametric functions of location. This approach may turn out to be of limited value, however, because we have found that switching to parametric smoothers often requires more degrees of freedom than may be available in spatial data sets with relatively few observations.

Another possible solution to the problem might be to restructure the independent variables in some way to remove the concurvity. This strategy, in the form of principal components regression, has certainly proven useful for dealing with multicollinearity in linear and generalized linear models. Until such time as a viable solution to concurvity-induced bias in spatial GAMs is developed, the present results suggest that the GAM is not an appropriate method for dealing with spatial autocorrelation in air pollution data.

REFERENCES

Burnett R, Ma R, Jerrett M, Goldberg MS, Cakmak S, Pope A III, et al. 2001. The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data. Environ Health Perspect 109(suppl 3):375–380.

Cakmak S, Burnett R, Jerrett M, Goldberg M, Pope A, Ma R, et al. In press. Spatial regression models for large cohort studies linking community air pollution and health. J Toxicol Environ Health.

Chambers JM, Hastie TJ. 1992. Statistical Models in S. London:Chapman & Hall.

Dominici F, McDermott A, Hastie T. 2003. Issues in Semi-parametric Regression with Applications in Time Series Models for Air Pollution and Mortality. Baltimore, MD:Johns Hopkins University. Available: http://biosun01.biostat.jhsph.edu/~fdominic/trevorpaper.pdf [accessed 30 May 2003].

Dominici F, McDermott A, Zeger S, Samet J. 2002. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 156:3:1–11.

Ferrándiz J, López A, Sanmartín P. 1999. Spatial regression models in epidemiological studies. In: Disease Mapping and Risk Assessment for Public Health (Lawson A, Biggeri A, Böning D, Lesaffre E, Viel J-F, Bertollini R, eds). Chichester, UK:John Wiley and Sons, 203–215.

Hastie T, Tibshirani R. 1990. Generalized Additive Models. London:Chapman & Hall.

Internet-based Health & Air Pollution Surveillance System. 2002. GAM Exact S.E. Available: http://ihapss.biostat.jhsph.edu/software/gam.exact [accessed 6 June 2003].

Krewski D, Burnett RT, Goldberg MS, Hoover K, Siemiatycky J, Jerrett M, et al. 2000. Re-analysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Part II: Sensitivity Analysis. Cambridge, MA: Health Effects Institute, 129–240.

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:9:1132–1141.

Ramsay T, Burnett R, Krewski D. 2003. Underestimation of standard errors in generalized additive models linking mortality to ambient particulate matter. Epidemiology 14:18–23.