Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Modeling Spatiotemporal Forest Health Monitoring Data

Nicole H. Augustin[a], Monica Musio[a], Klaus von Wilpert[a], Edgar Kublin[a], Simon N. Wood[a] & Martin Schumacher[a]

[a] Nicole Augustin is Lecturer and Simon Wood is Professor, Department of Mathematical Sciences, University of Bath, Bath, U.K. Monica Musio is Lecturer, Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy. Edgar Kublin is Scientist and Klaus von Wilpert is Chief Scientist, Forest Research Centre Baden-Württemberg, Freiburg, Germany. Martin Schumacher is Professor, University Hospital Freiburg, University Freiburg, Freiburg, Germany. This work was supported by the German Federal Ministry for Education and Research (project 0339985). The first author was supported by a Royal Society U.K. Relocation Fellowship and by the Forest Research Centre Baden-Württemberg.
Published online: 01 Jan 2012.

Modeling Spatiotemporal Forest Health Monitoring Data

Nicole H. AUGUSTIN, Monica MUSIO, Klaus VON WILPERT, Edgar KUBLIN, Simon N. WOOD, and Martin SCHUMACHER

Forest health monitoring schemes were set up across Europe in the 1980s in response to concerns about air pollution-related forest dieback (*Waldsterben*) and have continued since then. Recent threats to forest health are climatic extremes likely due to global climate change and increased ground ozone levels and nitrogen deposition. We model yearly data on tree crown defoliation, an indicator of tree health, from a monitoring survey carried out in Baden-Württemberg, Germany since 1983. On a changing irregular grid, defoliation and other site-specific variables are recorded. In Baden-Württemberg, the temporal trend of defoliation differs among areas because of site characteristics and pollution levels, making it necessary to allow for space–time interaction in the model. For this purpose, we propose using generalized additive mixed models (GAMMs) incorporating scale-invariant tensor product smooths of the space–time dimensions. The space–time smoother allows separate smoothing parameters and penalties for the space and time dimensions and thus avoids the need to make arbitrary or ad hoc choices about the relative scaling of space and time. The approach of using a space–time smoother has intuitive appeal, making it easy to explain and interpret when communicating the results to nonstatisticians, such as environmental policy makers. The model incorporates a nonlinear effect for mean tree age, the most important predictor, allowing the separation of trends in time, which may be pollution-related, from trends that relate purely to the aging of the survey population. In addition to a temporal trend due to site characteristics and other conditions modeled with the space–time smooth, we account for random temporal correlation at site level by an autoregressive moving average (ARMA) process. Model selection is carried out using the Bayes information criterion (BIC), and the adequacy of the assumed spatial and temporal error structure is investigated with the empirical semivariogram and the empirical autocorrelation function.

KEY WORDS: Air pollution; Climate change; Environmental monitoring; Forest damage; Generalized additive mixed model; Norway spruce (*Picea abies* L.); Space–time data; Spatiotemporal model; Tensor product smooth; Tree defoliation.

1. INTRODUCTION

The main exogenous factors influencing European forest ecosystems are pollution, weather, and biological pests. These three factors interact with one another and with soil characteristics, which can amplify adverse effects on the ecosystem. In particular, the effects of deposition of pollutants through the air and rainfall are long-term and accumulative. They cause acidification of soil, which disequilibriates the soil chemistry, including the nutrient and metal availability, eventually leading to the washing out of essential alkaline macronutrients in the root area. The symptoms seen in trees due to this type of damage are defoliation and yellowing of foliage. Extreme climatic conditions, such as heat and droughts, impose a direct stress on the forest and also makes biological pests more common, causing more direct stress. These three factors and their interactions have caused substantial damage to the forest ecosystems in Europe over the last 4 decades. Severe damage by air pollution was first noticed in the 1970s, and as a consequence forest health has been monitored in Europe since the early 1980s (Seidling 2001). Air pollution has been partly reduced through emission cuts that have decreased industrial sulfur oxide emissions. But a more recent cause of damage is the increase in nitrogen oxide emissions from transportation by road and industry. This indirectly causes forest damage due to increased ground ozone levels and nitrogen deposition. In addition, climatic extremes, likely due to anthropogenic global climate change, also have recently increased forest damage.

We analyze yearly data on percentage of tree defoliation from the Terrestrial Crown Condition Inventory (TCCI), a monitoring survey carried out annually in Baden-Württemberg, Germany since 1983. The sampling grid is irregular with changing levels of coarseness. The response variable defoliation, other site-specific explanatory variables (e.g., soil type) and tree-specific variables (e.g., age) are recorded yearly. The main purpose of the survey is to monitor any changes in damage over time and space to allow prompt detection of significant changes in damage. Ideally, the survey should facilitate the investigation of the associations between forest damage and its possible causative factors. The relationship between nutrient availability and defoliation has been investigated using spatial data from other surveys, complementary to the TCCI, which concentrate on nutrient content in trees and soil (Augustin et al. 2007; Musio, von Wilpert, and Augustin 2007; Musio, Augustin, and von Wilpert 2008).

Here the investigation is restricted to relating the site characteristics to defoliation, because currently no data on deposition of pollutants and climate are available at the same temporal and spatial resolution. The results of the investigations are required for forest management, and on a larger scale will help guide recommendations on environmental policies.

The traditional estimation method in Baden-Württemberg calculates estimates of the mean defoliation separately from the TCCI data for each year. This method is suitable only for temporal trend estimation and does not take into account any spatial or temporal structure of the data. It also does not make use

Nicole Augustin is Lecturer (E-mail: *n.h.augustin@bath.ac.uk*) and Simon Wood is Professor, Department of Mathematical Sciences, University of Bath, Bath, U.K. Monica Musio is Lecturer, Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy. Edgar Kublin is Scientist and Klaus von Wilpert is Chief Scientist, Forest Research Centre Baden-Württemberg, Freiburg, Germany. Martin Schumacher is Professor, University Hospital Freiburg, University Freiburg, Freiburg, Germany. This work was supported by the German Federal Ministry for Education and Research (project 0339985). The first author was supported by a Royal Society U.K. Relocation Fellowship and by the Forest Research Centre Baden-Württemberg.

Journal of the American Statistical Association
September 2009, Vol. 104, No. 487, Applications and Case Studies
DOI: 10.1198/jasa.2009.ap07058

of possible predictor variables. Ignoring correlation in the data when estimating the trend of defoliation could result in biased estimates and thus give wrong answers.

A new modeling approach is required that adequately accounts for possible spatial and temporal correlation and incorporates important predictors. The new approach needs to be able to deal with data on an irregular grid, with different subsets of grid locations sampled over time. Because it is very unlikely that the spatial trend of defoliation is additive in time, the new approach needs to account for a space–time interaction. Finally, the effects of important predictor variables, such as mean tree age, need to be modeled as nonlinear effects. Recent methodological developments in space–time modeling for point-referenced data have involved mainly hierarchical Bayes models (see Banerjee, Carlin, and Gelfand 2004 for an introduction) and geostatistical models, and applications have been mainly in environmental pollution monitoring (Shaddick and Wakefield 2002; Sahu, Gelfand, and Holland 2007), meteorology (Fuentes et al. 2005; Xu, Wikle, and Fox 2005; Glasbey and Allcroft 2008) and epidemiology (MacNab and Dean 2001; Richardson, Abéllan, and Best 2006).

Generalized additive mixed models (GAMMs) (Lin and Zhang 1999; Kamman and Wand 2003; Ruppert, Wand, and Carroll 2003) have been used to model spatiotemporal data and provide the general framework to fulfill the model requirements. Generalized additive models (GAMs; Hastie and Tibshirani 1990) are a special case of GAMMs that have no random-effect term. In most recent applications of GAMMs to spatiotemporal data, the space and time dimensions are modeled separately; for example, Kneib and Fahrmeir (2006) presented a GAM for ordered categorical data with an application to tree defoliation monitoring data. The spatial and temporal effects enter the model additively, and the spatial correlation is dealt with via a random effect for location using, for example, a Markov random field prior. MacNab and Dean (2001) used a GAMM to analyze spatiotemporal mortality data by including effects for space and time in an additive manner. Here we use the general methodology of Wood (2004, 2006b) to construct scale-invariant tensor product smooths of the space–time dimension. Our proposed method accommodates any response that follows a distribution from the exponential family. It does not rely on a regular grid and allows incorporation of a wide range of correlation structures. Besides one-dimensional smooth functions accounting for nonlinear effects of covariates, such as altitude, the space–time interaction can be modeled using scale-invariant tensor product smooths, where the smoothness parameters are estimated and also do not depend on the different scales of the covariate axes. These tensor product smooths also allow combinations of different basis functions most suitable for the dimensions of space and time, as well as time-varying spatial estimates. Our approach provides marginal estimates of average defoliation over time and space with confidence bands, thereby allowing assessment of changes in trends of defoliation.

2. THE TERRESTRIAL CROWN CONDITION SURVEY

The data analyzed are from the TCCI, which has been carried out yearly in the forests of Baden-Württemberg since 1983. The survey's spatial resolution has varied during the years; it is 4×4 km in 1983–1986, 1989, 1991, 1994, 1997, and 2001; 8×8 in 1987, 1988, 2005–2007; and 16×16 in the remaining years. Figure 1 shows grid locations for each year. The data are essentially yearly repeated measures on an irregular spatial grid with different subsets of locations missing, depending on the year. We restrict the analysis to the years 1985–2007, because in the years 1983 and 1984 the survey protocol differed substantially from subsequent years. Figure 2 shows the complete set of 1,475 sampling grid points. Seven natural growth regions are used for forest management and reporting purposes. The sampling grid is densest in the Black Forest area, because of overproportionally dense forest cover compared with other areas. The survey is carried out in alignment with the European level I monitoring program of the International Cooperative Programme on Forests and Integrated Monitoring of Ecosystems.

The response variable recorded is percentage defoliation in the crown, estimated by eye in 5% classes for each individual tree using binoculars. The sampling design excludes all parts of crowns that are under direct influence of shadow from surrounding trees. Only the upper crown is assessed, and trees with biotic damage are excluded. At each grid location, several trees are systematically selected using four subplots oriented among the main compass directions 50 m away from the grid point. On each subplot, the 6 trees nearest to the subplot centre are selected as sample trees, resulting in 24 sample trees per plot. In 2007, the maximum distance of a selected tree from the subplot center was 32 m. Selected trees are permanently marked and reassessed during subsequent surveys. Trees that are removed are replaced by newly selected trees.

The survey protocol has several features designed to avoid observer bias. The assessment is standardized every year through intercalibration courses, where only observers with a between-observer error $< 5\%$ are selected, using the observations of an experienced permanent member of staff as a benchmark. The courses take place at the European Union level, federal level, and state level, to avoid regional biases. Every observer has a photoguide (Evers et al. 1997) for comparing the observed defoliation with different percentages of defoliation in photos (photos are in 5%–15% levels). In additon, the survey is carried out with two independent observers per tree, and estimates are pooled. As a further check, 20% of plots are reassessed by experienced members of staff. A temporal trend in observer error can be excluded by this procedure, because the basis for these intercalibration courses is the photo guide. As covariates we consider year, northing, easting, altitude, age, slope gradient, geology, slope direction, situation, soil texture, soil type, soil depth, relief type, and nutrient balance in soil. Except for year and age, these covariates are available at the grid point level only.

3. EXPLORATORY ANALYSIS

Between 1985 and 2007, more than half of the observed trees were Norway spruce (*Picea abies* L.), with the rest beech, fir, pine, and other species. Here we limit all subsequent analysis to spruce. To compare variability within and between sampling grid points, we fit a random-effects model to all data

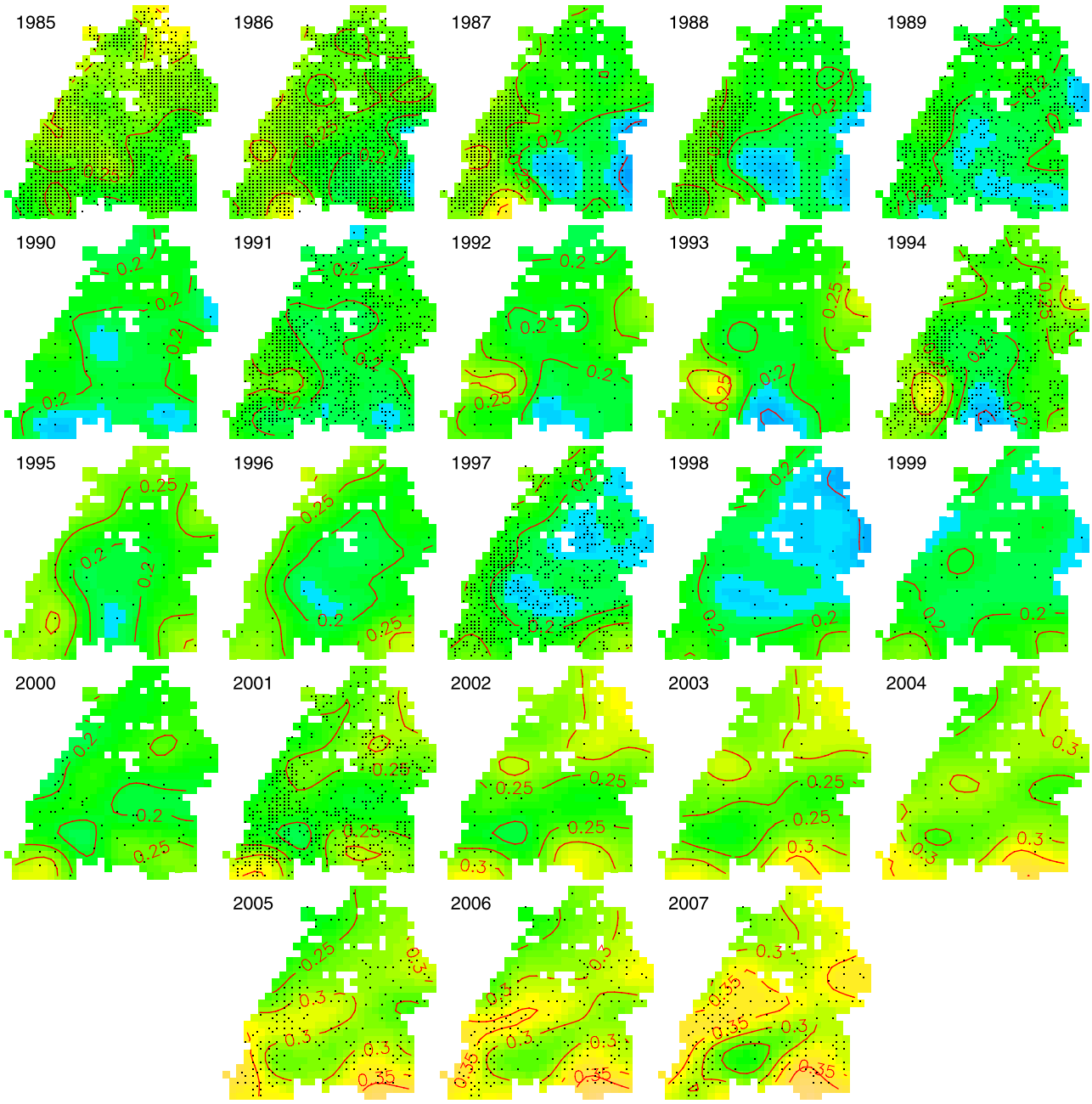

Figure 1. Predicted surface of average defoliation by year (1985–2007) for spruce [model (2)]. The mean age of 74 years was used for all predictions. Blue indicates a low percentage, green indicates a medium percentage, and yellow indicates a high percentage of defoliation, with defoliation levels given on each plot by the isolines. The dots indicate the sampling locations in each year.

within a particular survey year (for all years) for spruce, $y_{ij} = \alpha + b_i + \epsilon_{ij}$, where y_{ij} is the observed proportion defoliation of spruce tree j at grid location i, α is a fixed-effects parameter for the population mean, and b_i is a random-effects parameter for the ith location with $i = 1, \ldots, 1{,}475$ with $j = 1, \ldots, a_i$ spruce trees at location i, where a_i is a maximum of 24. We assume $b_i \sim \mathrm{N}(0, \sigma_b^2)$ and $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$. The b_i and ϵ_{ij} are assumed to be mutually independent. The resulting confidence intervals of σ_b and σ for each year are computed relying on the asymptotic normality of restricted maximum likelihood estimates. (σ^2 and σ_b^2 are distant from 0 here.) The 95% confidence intervals indicate that the variance component relating to between-grid location variability, σ_b, tends to be larger in most years than the variance component relating to within-site variability, σ, and in 19 of 23 years $\hat{\sigma}_b$ is significantly greater than $\hat{\sigma}$. This implies that variability between grid points is greater than variability within grid points in most years. Thus we would expect to explain more variability by introducing site-specific rather than tree-specific covariates in the model. In the years in which the sampling grid resolution is 16×16, the confidence interval of the between-location variability, σ_b, is substantially wider than that in years with a smaller grid resolution. In 2000, the year of the winter storm Lothar, variability between grid locations was substantially higher than variability within sampling location compared with other years.

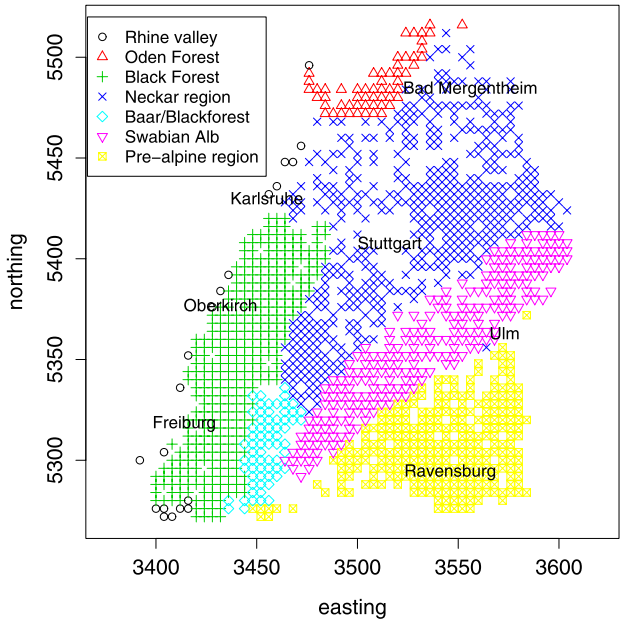

Figure 2. The 1,475 sampling locations of the TCCI survey with growth areas.

4. A SPATIOTEMPORAL MODEL FOR DEFOLIATION

4.1 Requirements

In what follows we state the requirements of the model based on the underlying physical and biological processes. The most important aspect of the data that the model needs to handle is the *space–time trend and interaction*. This should be modeled using a smooth three-dimensional function of space and time. The temporal trend of defoliation differs between areas because of differing site characteristics (e.g., geology, soil type, climatic conditions pollution levels) within the survey area. Thus defoliation is expected to occur at different strengths and cycles depending on the location. For example, the Black Forest area has various characteristics that make it more susceptible to pollution. The geology is mainly granite or Triassic sandstone bedrock; soils on such stone are more prone to acidification. In addition, much of the area is at high altitude, where acid input from rain and fog precipitation can be high due to a frequent meteorological inversion layer (Schöpfer and Hradetzky 1984). Thus defoliation is expected to be more severe and rapid here than in the Swabian Alb or pre-alpine region, where the geology is alkaline limestone with a higher buffer capacity.

Given the sampling scheme and the biology of trees, it is very likely that we observe *correlation in time* in the response, the proportion of defoliation. A spruce tree typically holds needles from the last 7 years, so we expect the observed defoliation to be a moving average. We also expect the defoliation to depend on the previous year's defoliation in more subtle ways. For example, trees that are damaged may be more susceptible to damage in subsequent years. Thus the correlation in time is composed of two parts: a temporal trend due to site characteristics and other conditions, which may be modeled by a smooth function of space and time. Thus the random variability in measurements at the tree level and site level are correlated in time.

It is well known that age is a major predictor for defoliation (Anon 2001) and that the effect of age on defoliation is not linear. The survey data contain a wide range of tree ages, from 5 to 208 years. Thus the model should contain a smooth function of age. To control for age when estimating the spatial and temporal trend of defoliation, we need to be able to separate trends in time, which may be pollution-related, from trends that relate purely to the aging of the survey population.

In cases where *spatial correlation* between grid locations is not accounted for by the smooth space–time function, it is most likely caused by spatial trends in the site characteristics. Such trends should be modeled by functions of site characteristics, such as geology, soil type, type of slope, and so on. There also is likely to be spatial correlation within a sampling grid point, where typically 24 trees are observed. Because the x and y coordinates are not available at the tree-level, modeling the spatial correlation within a grid point is difficult. We can use only age of the three relevant tree- specific covariates (social class, diameter, and age). But we excluded social class, because the survey protocol restricts the social class to the two upper social classes (class 1, very dominant trees, and class 2, dominant trees) to avoid competition effects, and excluded diameter because it is strongly correlated with age. The other covariates are available only at the grid point resolution. Therefore, we aggregate the data within grid points, and thus the model need only account for the temporal correlation within, and spatial correlation between, sampling grid points. Aggregation also makes sense because most of the forest is heavily managed, and at any grid location trees will almost always have the same age and will have been planted in a regular pattern; thus there generally are no significant differences in terms of height and dominance. Moreover, the model is to be used as an environmental monitoring tool, and tree-specific prediction is not of interest. Thus the aggregated response variable is the average defoliation $\bar{y}_{it}$ at the ith grid point location with $i = 1, \ldots, 1{,}475$ and $j = 1, \ldots, a_{it}$ trees, with a maximum of 24 trees at location i in year t, with $t = 1985, \ldots, 2007$, where weights $1/a_{it}$ are used in all analysis. We assume that $\bar{y}_{it}$ follows a normal distribution by making use of the central limit theorem. Note that to simplify notation, we refer to $\bar{y}_{it}$ as y_{it} in what follows.

4.2 Traditional Method for Trend Estimation

The method traditionally used for trend estimation of average defoliation in Baden-Württemberg does not fulfill any of the aforementioned requirements. It completely ignores the longitudinal and spatial structure of the data and can be described as fitting a linear model to years separately:

$$y_{ij} = \alpha + \epsilon_{ij} \tag{1}$$

where y_{ij} is the observed defoliation of tree j at grid location i, α is a fixed-effects parameter for the population mean, and the error $\boldsymbol{\epsilon}$ is distributed as $\mathrm{N}(\mathbf{0}, \mathbf{I}\sigma^2)$. Thus the traditional method simply takes an average of the observed defoliation by year and species over the observed area, that is, the whole of Baden-Württemberg or a specific growth area. An informal test on a significant change in trend involves checking whether confidence intervals for the mean defoliation between different years overlap. This test assumes independence of the data in time and space, which is not given here, and thus may lead to incorrect conclusions. In addition, the traditional method does not make use of explanatory variables, for example, tree age, which is known to be a very strong predictor.

4.3 The Proposed Model

Our proposed model is a GAMM (Lin and Zhang 1999; Fahrmeir and Lang 2001) that allows for a complex stochastic structure for dealing with the foregoing requirements and is the result of some extensive model selection described in Section 6:

$$\text{logit}\, E(y_{it}) = f_1(age_{it}) + f_2(no_i, e_i, year_t). \quad (2)$$

The response, mean proportion defoliation, is

$$y_{it} = E(y_{it}) + \epsilon_{it}$$

at location $i = 1, \ldots, 1{,}475$ and year $t = 1, \ldots, 23$. The logit link is applied to ensure that fitted values are bounded in (0, 1), while keeping the Normal assumption for the error term. The observations are rarely near the boundary (0, 1), so this should not be a problem. We use this model because we are dealing with proportions of needle loss estimated by eye; that is, the counts are observed with error and without knowing the total number of needles, excluding the possibility of modeling the data with a binomial distribution.

The function f_1 is a one-dimensional smooth function of age_{it} represented using a cubic regression spline basis. The function f_2 is a multidimensional smooth function of northing (no), easting (e), and year, allowing smoothing parameter selection to be independent of the different scales of the covariate axes. We want both the time and space dimension to have an "optimal" degree of smoothness in terms of the bias variance trade-off. Because the units of time (years) and space (km) are different, the smoother needs to be invariant to their relative scaling, which is essentially arbitrary. This can be achieved by using multidimensional tensor product smooths with different penalties for each marginal basis. For the two spatial dimensions no and e, an isotropic thin-plate regression spline basis function is used, because in this case the smoother should not depend on the coordinate system used. For the time dimension, a cubic regression spline basis function is used (details given later).

The residual error vector $\boldsymbol{\epsilon}$ is distributed as $N(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ is block diagonal with the ith subvector $\boldsymbol{\epsilon}_i$ having covariance matrix $\boldsymbol{\Lambda}_i$ relating to residuals at one location over time. The $\boldsymbol{\Lambda}_i$ also contains on the diagonal the weights $1/a_{it}$, where a_{it} is the number of spruce trees sampled at location i and time t. The $\boldsymbol{\epsilon}_i$ reflect the temporal correlation in the error, modeled by a mixed autoregressive and moving average (ARMA) process, using the approach given by Pinheiro and Bates (2000). Here we use a first-order ARMA process that assumes the following dependence structure in errors: $\epsilon_{it} = \phi\epsilon_{it-1} + \rho c_{it-1} + c_{it}$, with correlation parameters ϕ and ρ and where c_{it} is random Gaussian noise with an expected value of 0. Thus $\boldsymbol{\Lambda}_i$ has entries $\lambda_{t,t-k} = \phi + \rho/(\rho^2 + 1)$ for $k = 1$ and $\lambda_{t,t-k} = \phi^k(\phi + \rho/(\rho^2 + 1))$ for $k > 1$.

5. METHODOLOGY DETAILS

5.1 The Full Model

Model (2) described earlier is a restricted version of the following model, which we use to describe the methodology in general:

$$\text{logit}\, E(y_{it}) = \mathbf{X}_{it}\boldsymbol{\theta} + f_1(age_{it}) + f_2(no_i, e_i, year_t) + f_3(w_{3it}) + \cdots + f_k(w_{kit}) + \mathbf{Z}_{it}\mathbf{b} \quad (3)$$

with $y_{it} = E(y_{it}) + \epsilon_{it}$. The vector $\boldsymbol{\theta}$ contains fixed parameters; $\mathbf{X}_{it}$ is a row of a fixed effects model matrix; f_1 is as described for model (2); $f_3, \ldots, f_k$ are smooth functions of covariates $w_3, \ldots, w_k$; $\mathbf{Z}_{it}$ is a row of a random-effects model matrix; and the vector of random effects coefficients $\mathbf{b}$ is distributed as $N(\mathbf{0}, \boldsymbol{\psi})$ with unknown positive definite covariance matrix $\boldsymbol{\psi}$. We assume that the error vector $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$ with a covariance matrix $\boldsymbol{\Lambda}$ reflecting the assumed error structure. For example, in model (2) this is a block diagonal, with the ith subvector $\boldsymbol{\epsilon}_i$ having covariance matrix $\boldsymbol{\Lambda}_i$, reflecting temporal correlation at location i but no spatial correlation. Alternative structures are possible, such as an exponential spatial correlation between locations (Pinheiro and Bates 2000). Model (3) is extremely general, and care must be taken to not overfit, because in most data situations identifiability issues exist. For example, the spatial effect of $f_2(\cdot)$ will be confounded with other site-specific effects. For this reason, we perform careful step-by-step model selection, as described in Section 6.

5.2 A Three-Dimensional Tensor Product Smoother for Space and Time

For constructing a three-dimensional tensor product smooth of space and time, we start from a marginal smooth for time f_{year} and a two-dimensional marginal smooth for space $f_{no,e}$. With the marginal smooths, we have associated quadratic penalties measuring their roughness. Assuming that we have two low-rank bases of any type (with mixing of types possible) available for representing smooth functions for f_{year} and $f_{no,e}$, we can write

$$f_{year}(year) = \sum_{i=1}^{P} \alpha_i a_i(year) = \mathbf{X}_{year}\boldsymbol{\alpha} \quad \text{and}$$

$$f_{no,e}(no, e) = \sum_{l=1}^{L} \beta_l b_l(no, e) = \mathbf{X}_{no,e}\tilde{\boldsymbol{\beta}},$$

where α_i and β_l are parameters, whereas a_i and b_l are basis functions with $i = 1, \ldots, P$ and $l = 1, \ldots, L$. The $\mathbf{X}_{year}$ and $\mathbf{X}_{no,e}$ are the marginal model matrixes evaluating the basis functions with the corresponding parameter vectors $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\beta}}$. Now consider how the smooth function of $year$, f_{year}, can be converted into a smooth function of no and e. For this to occur, we need the temporal smooth f_{year} to vary smoothly within the space dimensions no and e. This can be achieved by letting its coefficients vary with no and e. Using the basis setup for $f_{no,e}$ we can write

$$\alpha_i(no, e) = \sum_{l=1}^{L} \beta_{il} b_l(no, e),$$

which gives

$$f_{year,no,e}(year, no, e) = \sum_{i=1}^{P}\sum_{l=1}^{L} \beta_{il} b_l(no, e) a_i(year). \quad (4)$$

For any particular set of observations of $year$, no, and e, there is a simple relationship between the model matrix $\mathbf{X}_f$ evaluating the tensor product smooth at these observations and the model matrixes $\mathbf{X}_{year}$ and $\mathbf{X}_{no,e}$ that would evaluate the marginal smooths at the same observations. By ordering the β_{il} appropriately into a vector $\boldsymbol{\beta}$, it can be shown that the ith row of $\mathbf{X}_f$

is $\mathbf{X}_{fi} = \mathbf{X}_{year,i} \otimes \mathbf{X}_{no,e,i}$, where $\otimes$ is the Kronecker product. In fact, this construction can be continued for as many covariates as are required (see Wood 2006a for a general introduction).

For the roughness penalty associated with this "tensor product" basis, we also start from the marginal smooth functions f_{year} and $f_{no,e}$. Suppose that each smooth has an associated functional, J, that measures its roughness and can be written as a quadratic form in its coefficients. The J obviously depends on the basis function type. Specifically, suppose that

$$J_{year}(f_{year}) = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{S}_{year}\boldsymbol{\alpha} \qquad \text{and} \qquad J_{no,e}(f_{no,e}) = \tilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{S}_{no,e}\tilde{\boldsymbol{\beta}}.$$

The matrixes $\mathbf{S}_{year}$ and $\mathbf{S}_{no,e}$ contain known coefficients, and $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\beta}}$ are the parameters of the marginal smooths. Note that we distinguish between $\tilde{\boldsymbol{\beta}}$ containing the marginal smooth parameters and $\boldsymbol{\beta}$ containing the tensor product smooth parameters β_{il}. Here the cubic penalty for f_{year} is $J_{year}(f_{year}) = \int f''_{year}(year)^2\, d\, year$. To obtain an overall penalty, we apply the penalties of the spatial smooth to the spatially varying coefficients of the marginal temporal smooth, $\alpha_i(no, e)$,

$$\sum_i^P J_{no,e}\{\alpha_i(no, e)\},$$

and, equivalently, we apply the penalties of the temporal smooth to the temporally varying coefficients of the marginal spatial smooth, $\beta_l(year)$,

$$\sum_l^L J_{year}\{\beta_l(year)\}.$$

Thus the roughness of $f_{year,no,e}$ can be measured by the sum of the two penalties weighted by smoothness parameters for time λ_{year} and space $\lambda_{no,e}$:

$$J(f_{year,no,e}) = \lambda_{no,e}\sum_i^P J_{no,e}\{\alpha_i(no, e)\} + \lambda_{year}\sum_l^L J_{year}\{\beta_l(year)\}.$$

This can be reexpressed, with $\boldsymbol{\beta}$ as defined in eq. (4), as

$$J(f_{year,no,e}) = \lambda_{no,e}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{I}_P \otimes \mathbf{S}_{no,e}\boldsymbol{\beta} + \lambda_{year}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_{year} \otimes \mathbf{I}_L\boldsymbol{\beta}.$$

In model (2) the cubic regression spline basis is used for the temporal smooth. Thus the elements of $\boldsymbol{\beta}$ are the actual function heights, which vary smoothly in the space dimension.

5.3 Parameter Estimation

Here we use the fact that the smooth model terms can be represented as random effects, allowing their estimation via standard mixed modeling software (Lin and Zhang 1999; Wood 2004). The mixed model approach provides a self-consistent and computationally tractable way to handle smoothing and deal with autocorrelation simultaneously. It involves setting prior distributions on coefficients of $f_k(\cdot)$, derived from the roughness penalties, and treating the coefficients as random effects. To avoid improper random-effects distributions, some reparameterization generally is required so that each smooth is represented using a small number of fixed effects and a larger number of random effects with a proper distribution (Wood 2006b). Thus model (3) can be rewritten as

$$\operatorname{logit} E(y_{it}) = \mathbf{X}_{it}\boldsymbol{\theta} + \mathbf{X}_{fit}\boldsymbol{\theta}_f + \mathbf{Z}_{it}\mathbf{b} + \mathbf{Z}_{fit}\mathbf{b}_f \tag{5}$$

$$= \mathbf{X}^\star_{it}\boldsymbol{\theta}^\star + \mathbf{Z}^\star_{it}\mathbf{b}^\star \tag{6}$$

with $y_{it} = E(y_{it}) + \epsilon_{it}$. The fixed-effects components of $f_1, \ldots, f_k$, relating to unpenalized coefficients (with completely improper priors), $\mathbf{X}_{fit}\boldsymbol{\theta}_f$, are absorbed into $\mathbf{X}^\star_{it}\boldsymbol{\theta}^\star$, and the components relating to penalized coefficients with proper priors, $\mathbf{X}_{fit}\boldsymbol{\theta}_f$, are absorbed into $\mathbf{Z}^\star_{it}\mathbf{b}^\star$. In this reparameterized form of the model, we have model matrixes and vectors for fixed and random effects that contain, in addition to the usual fixed and random components, additional components from the smoothers. Then $\mathbf{b}^\star$ is distributed as $\mathrm{N}(\mathbf{0}, \boldsymbol{\psi}^\star)$ with matrix $\boldsymbol{\psi}^\star$ also depending on the smoothing parameters λ_{age}, λ_{year}, and $\lambda_{no,e}$ of the one- and three-dimensional smoothers in addition to the variance of the random effects. The error $\boldsymbol{\epsilon}$ is distributed as $\mathrm{N}(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ contains the correlation parameters. Because the GAMM can be expressed as a generalized linear mixed model (GLMM) [as in eq. (6)], parameter estimation can be carried out as for a GLMM using penalized quasi-likelihood (PQL) (Breslow and Clayton 1993). This amounts to iteratively maximizing an (approximate) Laplace approximation of the marginal likelihood, the likelihood of a weighted linear model,

$$L^*(\boldsymbol{\theta}^\star, \sigma^2, \boldsymbol{\psi}^\star) \propto |\boldsymbol{\psi}^\star|^{-1/2}\int \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{W}^{1/2}(\mathbf{u} - \mathbf{X}^\star\boldsymbol{\theta}^\star - \mathbf{Z}^\star\mathbf{b}^\star)\right\|^2 - \frac{1}{2}\mathbf{b}^{\star\mathrm{T}}\boldsymbol{\psi}^{\star-1}\mathbf{b}^\star\right) d\mathbf{b}^\star,$$

where $\mathbf{W} = \mathbf{V}^{1/2}\boldsymbol{\Lambda}\mathbf{V}^{1/2}$, with the diagonal weights matrix, $\mathbf{V}$, induced by the logit link function $g(\cdot)$, with entries for location i and time t, $v_{it} = \frac{1}{g'(\hat{\mu}^b_{it})^2}$. Defining $\mu^b_{it} = \widehat{E(y_{it}|\mathbf{b}^\star)}$, then $\mathbf{u}$ is a vector of pseudodata with elements

$$u_{it} = g'(\hat{\mu}^b_{it})(y_{it} - \hat{\mu}^b_{it}) + (X^\star_{it}\hat{\boldsymbol{\theta}}^\star + Z^\star_{it}\hat{\mathbf{b}}^\star).$$

At each iteration, the following linear mixed-effects model is estimated:

$$\mathbf{u} = \mathbf{X}^\star\boldsymbol{\theta}^\star + \mathbf{Z}^\star\mathbf{b}^\star + \boldsymbol{\epsilon},$$

with $\mathbf{b}^\star$ distributed as $\mathrm{N}(\mathbf{0}, \boldsymbol{\psi}^\star)$ and $\boldsymbol{\epsilon}$ distributed as $\mathrm{N}(\mathbf{0}, \mathbf{W}^{-1}\sigma^2)$. We use the `gamm()` function of the `mgcv` R package, which iteratively calls the `lme()` function of the `nlme` R package (Pinheiro et al. 2008) for maximization.

For model comparisons without "conventional" random effects, we use a generalized version of the Akaike information criterion (AIC) and BIC, obtained by treating the smooths as penalized fixed effects. Then, if $\mathbf{y}$ denotes the response and $\hat{\boldsymbol{\mu}}$ denotes the predictions of $\mathbf{y}$ according to the estimated smooth model, we have

$$\hat{L} = \frac{1}{(2\pi)^{n/2}|\hat{\sigma}^2\hat{\boldsymbol{\Lambda}}|^{1/2}}\exp\left(-\frac{1}{2\hat{\sigma}^2}(\mathbf{y} - \hat{\boldsymbol{\mu}})^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}})\right) = \frac{1}{(2\pi)^{n/2}|\hat{\sigma}^2\hat{\boldsymbol{\Lambda}}|^{1/2}}\exp(-n/2),$$

where $\hat{\mathbf{\Lambda}}$ is the estimated correlation matrix and

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \hat{\boldsymbol{\mu}})^{\mathrm{T}} \hat{\mathbf{\Lambda}}^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}})}{n},$$

where $n = 7{,}864$, the total number of observations, that is, the sum of all the sampling points over all years. The effective degrees of freedom (edf) of the penalized model also account for parameters of the covariance matrix $\mathbf{\Lambda}$ and are obtained in the usual way (see the end of Sec. 5.4), and the approximate AIC is then $AIC = -2\log(\hat{L}) + 2edf$ and the BIC is, accordingly, $BIC = -2\log(\hat{L}) + \log(n)edf$. The BIC can be used in an approximation of the Bayes factor (Kass and Raftery 1995), and thus it is suitable in situations with large sample sizes with respect to the number of parameters, which is the case in our application. In practice, due to the smaller penalty term, the AIC tends to keep more terms in the model than the BIC, and to avoid overfitting, we use the BIC in our model selection. Simulation results of Huang et al. (2007) show that the BIC outperforms the AIC in space–time model selection, particularly for data sampled irregularly in space.

5.4 Variance Estimation

We explain the variance estimation in terms of the reparameterized version of the general model in eq. (6). Let $\boldsymbol{\gamma}^{\mathrm{T}} = (\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\theta}_f^{\mathrm{T}}, \mathbf{b}_f^{\mathrm{T}})$ be the parameter vector containing all fixed effects and the random effects for the smooth terms only, and let $\bar{\mathbf{X}}$ be the corresponding model matrix. Let $\boldsymbol{\psi}$ be the covariance matrix of random effects excluding $\mathbf{b}_f$, the random effects relating to the smoothers. We use the Bayesian representation of the frequentist mixed model because we are interested in making inferences about model components in $\boldsymbol{\gamma}$, some of which are treated as random variables in the mixed modeling. In addition, the Bayesian model representation gives a self-consistent basis for constructing confidence intervals or, more precisely, credible intervals, because the posterior distribution of the model parameters is known. In comparison, frequentist confidence intervals are not as straightforward (see Wood 2006a, p. 189). We use Bayesian credible intervals, as developed by Wahba (1983) and Silverman (1985); see also Wood (2004, 2006c), for their computational implementation. A Bayesian posterior covariance matrix for the coefficients of the smooth terms and fixed effects can be obtained. Conditioning on the parameter estimates for the random effects $\mathbf{b}$, excluding the random effects for the smooth terms, we first need to calculate the covariance matrix for the response data implied by the estimated random- effects structure *excluding* the smooth terms, $\mathbf{Q} = \mathbf{Z}^{\mathrm{T}} \boldsymbol{\psi} \mathbf{Z} + \mathbf{\Lambda}\sigma^2$, where $\mathbf{Z}$ and $\boldsymbol{\psi}$ are defined as in eq. (3). Then

$$\boldsymbol{\gamma} | \mathbf{y} \dot{\sim} \mathrm{N}(\hat{\boldsymbol{\gamma}}, (\bar{\mathbf{X}}^{\mathrm{T}} \mathbf{Q}^{-1} \bar{\mathbf{X}} + \mathbf{S})^{-1}), \tag{7}$$

where $\hat{\boldsymbol{\gamma}}$ is the vector of estimates or predictions of the elements of $\boldsymbol{\gamma}$. The matrix $\mathbf{S}$ is block diagonal with blocks $\lambda_{no,e}/\sigma^2 \mathbf{I}_p \otimes \mathbf{S}_{no,e}$, $\lambda_{year}/\sigma^2 \mathbf{I}_L \otimes \mathbf{S}_{year}$, and $\lambda_{age}/\sigma^2 \mathbf{S}_{age}$ in the case of model (2). This is essentially the approach taken by Lin and Zhang (1999). The degrees of freedom per element of $\boldsymbol{\gamma}$ can be estimated from the leading diagonal of $(\bar{\mathbf{X}}^{\mathrm{T}} \mathbf{Q}^{-1} \bar{\mathbf{X}} + \mathbf{S})^{-1} \bar{\mathbf{X}}^{\mathrm{T}} \mathbf{Q}^{-1} \bar{\mathbf{X}}$.

5.5 Trend Estimation

Using the reparameterized version (6) of the general model (3), the mean needle loss at time t, averaged over the whole survey area, is estimated as

$$\hat{\bar{y}}_t = \frac{1}{n} \sum_{i=1}^{n} \mathrm{logit}^{-1}(\mathbf{X}_{it}^{\star} \hat{\boldsymbol{\theta}}^{\star} + \mathbf{Z}_{fit} \hat{\mathbf{b}}_f),$$

where $\mathbf{X}_{it}^{\star}$ and $\mathbf{Z}_{fit}$ are evaluated at the observed values. The random effects that are part of the space–time trend and the smooth function of age, $\mathbf{Z}_{fit}\hat{\mathbf{b}}_f$, are included with their predicted expected (posterior) value, and the "conventional" random effects, $\mathbf{Z}_{it}\hat{\mathbf{b}}$, are considered random variation and taken at their (prior) expected value (i.e., 0). Note that we do not use the number of trees sampled at location i as weights in the estimator. Weighting the individual estimates at each location i by their respective variance, which is proportional to the proportion of spruce trees on which the estimates are based, would downweight certain grid locations where the tree species are not entirely spruce; that is, fewer than 24 spruce trees are observed. This would be an undesirable feature of the estimator.

Alternatively, the predictive distribution of y_{it} can be used to provide estimates. We obtain a posterior sample of the distribution of estimates $\boldsymbol{\gamma}^{\mathrm{T}} = (\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\theta}_f^{\mathrm{T}}, \mathbf{b}_f^{\mathrm{T}})$ with the posterior covariance matrix as in (7), and from this obtain a sample from the predictive distribution of the response. Then the pth draw from the posterior distribution is

$$\widehat{E_p(y_{it})} = \hat{y}_{itp} = \mathrm{logit}^{-1}(\mathbf{X}_{it}^{\star} \hat{\boldsymbol{\theta}}_p^{\star} + \mathbf{Z}_{fit} \hat{\mathbf{b}}_{fp}).$$

Averaging over i yields the predictive distribution of $\hat{\bar{y}}_t$,

$$\hat{\bar{y}}_{tp} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_{itp}).$$

Then the required summary statistics, in this case the median and lower and upper 95% quantiles, are computed for the spatial ($\hat{y}_{itp}$) and temporal trend ($\hat{\bar{y}}_{tp}$) (as presented in Section 7). If we wanted to predict defoliation on a regular grid and, for example, a standard age, then we would evaluate the $\mathbf{X}_{it}^{\star}$ and $\mathbf{Z}_{fit}$ for these prediction data. Otherwise, we would carry out the same procedure described earlier.

5.6 Model Diagnostics

To check whether the model has eliminated residual correlation, we investigate residuals in space and time. For models in which an independent error structure is not assumed, normalized residuals are created as described by Pinheiro and Bates (2000). In the case of model (2), let the vector of errors at location i be $\boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{\Lambda}_i)$, as defined in Section 4.3. Each $\mathbf{\Lambda}_i$ has the same parameters ϕ and ρ but varying structure because of missing years at location i. Then the residual vector $\mathbf{r}_i$ is given by

$$\mathbf{r}_i = \hat{\sigma}^{-1} \left(\hat{\mathbf{\Lambda}}_i^{-1/2} \right)' (\mathbf{y}_i - \hat{\mathbf{y}}_i) \dot{\sim} \mathrm{N}(\mathbf{0}, \mathbf{I}),$$

with $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, where T is the number of years observed at location i. For visual investigation of the adequacy of the assumed spatial variance–covariance structure, the empirical semivariogram can be used. For estimation, we use the `variogram` function of the `geoR` package (Ribeiro Jr. and

Diggle 2001) in R (R Development Core Team 2008). We then calculate the empirical semivariogram of the residuals by year. We permute the residuals 999 times within year and compute the envelopes by taking, at each lag, the maximum and minimum values of the semivariograms for the permuted residuals.

To investigate the adequacy of the temporal variance–covariance structure, we calculate the empirical autocorrelation function for the residuals within each location and the corresponding approximate two-sided critical bounds for autocorrelations of random white noise, given by $z(1-\alpha/2)/\sqrt{N(l)}$, where $N(l)$ is the number of pairs of residuals at lag l in years and $z(1-\alpha/2)$ is the standard normal quantile of the critical value $1-\alpha/2$. For this, we use the ACF function of the nlme R package (Pinheiro et al. 2008). The critical bounds are approximate and in fact narrower than given earlier for low lags, because the normalized residuals are based on parameter estimates rather than on known parameters (Box and Pierce 1970).

6. MODEL SELECTION

We carry out the model selection in three steps. We first select the appropriate error structure, then check the space–time structure, and finally select covariates.

Step one. First, we check the models listed in Table 1 for gross violations of distributional assumptions, using the residual diagnostic plots described earlier. As the worst possible model, we fit an adapted version of the traditional method's model simultaneously to all years. To make the traditional method comparable with the other models, we use the aggregated defoliation y_{it} rather than individual tree defoliation. The other models, (a)–(c), have the same terms but differing error distribution assumptions, and the last model (d) includes also a random effect for grid point i, taking into account that the data are repeated measures on grid points. The diagnostic plots show that only model (2)(c) eliminates both the temporal and spatial residual correlation. Temporal correlation is strongest for the traditional method and model (a), both with independent errors. The residual temporal correlation of the random-effects model (d) is negative for lags exceeding 2 years. The semivariograms of normalized residuals for three typical years shown in Figure 3 confirm that the traditional method does not eliminate residual spatial correlation, whereas model (2) mostly succeeds, with 1986 as an exception.

Step two. In the second step we investigate the adequacy of the type of space–time effect, modeled with the three-dimensional function $f_2(no_i, e_i, year_t)$ in model (2). An obvious question is whether we can replace $no_i, e_i, year_t$ by other covariates that have a mechanistic relationship with defoliation. In particular, we would expect site-specific time-varying covariates on weather and pollution to be more adequate. But this option currently is not possible, because these variables are not available. Another idea is to replace some of the space–time interaction modeled in f_2 with a site-specific time–covariate interaction, with constant in time covariates (e.g., geology). But this would be very hard to interpret, because it would imply that the defoliation mechanism is changing with time. To check whether the space–time interaction is required, we compared different types of functions of *no*, *e*, and *year* by plotting normalized residuals versus time and smoothing residuals in time and in space. The results, given in Table 2, show that only model (2) is acceptable.

Step three. Having selected the error structure and the type of space–time smoother, we next carry out forward variable selection on model (2). Forward selection is preferred to backward selection for computational reasons and is defensible because we have already shown that model (2) gives an adequate fit to the data. The set of covariates includes altitude, slope gradient, geology, slope direction, situation, soil texture, soil type, soil depth, relief type, and nutrient balance. In the forward selection, we add each covariate of the set to model (2) in turn. A covariate is selected if the fit yields a lower BIC than model (2). The forward selection confirms that model (2) is the best model; that is, no other covariate is selected. Note that we allow only for additive covariate effects, which may be a restriction here. Some covariates (e.g., soil type) may contribute

Table 1. Model selection: step one

Model	Residuals	
	In time	In space
"trad" $y_{it} = \alpha_t + \epsilon_{it}$ with $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{A})$, diagonal $\mathbf{A}$ with $1/a_{it}$*	correlated	correlated
(a) $\mathrm{logit}\, E(y_{it}) = f_1(age_{it}) + f_2(no_i, e_i, year_t)$, $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{A})$, diagonal $\mathbf{A}$ with $1/a_{it}$	correlated	uncorrelated
(b) $\mathrm{logit}\, E(y_{it}) = f_1(age_{it}) + f_2(no_i, e_i, year_t)$, $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$, first-order autoregressive (AR) with $\boldsymbol{\Lambda}_i(\phi)$ with $\epsilon_{it} = \phi\epsilon_{it-1}$ and along diagonal $1/a_{it}$	correlated	uncorrelated
(c) Model (2) $\mathrm{logit}\, E(y_{it}) = f_1(age_{it}) + f_2(no_i, e_i, year_t)$ [model (2)], $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$, ARMA with $\boldsymbol{\Lambda}_i(\phi, \rho)$ and along diagonal $1/a_{it}$	uncorrelated	uncorrelated
(d) $\mathrm{logit}\, E(y_{it}) = f_1(age_{it}) + f_2(no_i, e_i, year_t) + b_i$, $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{A})$, $\mathbf{A}$ as in (a) and random effect b_i iid $\mathrm{N}(0, \sigma_b^2)$	correlated	uncorrelated

NOTE: The table shows the results under the traditional method ("trad") and different versions of model (2) (a, b, c, and d) with respect to two kinds of temporal error structure. The first structure has independence over time ("trad"; a), and the second has correlation over time (b, c, and d). If the residuals are uncorrelated over time, then the entry in the second column is "uncorrelated"; otherwise, the entry is "correlated." A similar convention is used for correlation in space, reported in the third column.

*a_{it} is the number of trees sampled at site i and time t.

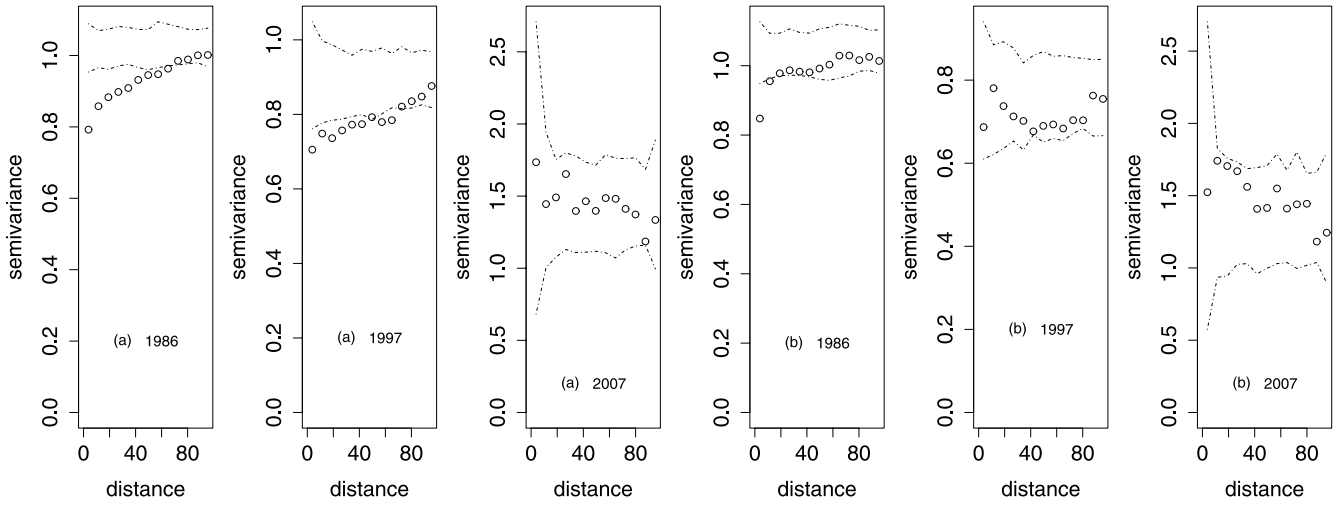

Figure 3. Semivariogram of scaled (normalized) Pearson residuals from (a) the traditional method [model (1)] and (b) model (2).

Table 2. Model selection: step two

	Residuals	
Space–time effect	In time	In space
$\text{logit}\, E(y_{it}) = f_1(age_{it})$	trend	trend
$\text{logit}\, E(y_{it}) = f_1(age_{it}) + f_2(no_i, e_i, year_t)$ [model (2)]	no trend	no trend
$\text{logit}\, E(y_{it}) = f_1(age_{it}) + f_3(no_i, e_i) + f_4(year_t)$	no trend	trend
$\text{logit}\, E(y_{it}) = f_1(age_{it}) + f_3(no_i, e_i)$	trend	no trend
$\text{logit}\, E(y_{it}) = f_1(age_{it}) + f_4(year_t)$	no trend	trend

NOTE: The table gives the results under models with different types of space–time interactions. For all models, it is assumed that $\boldsymbol{\epsilon} \sim \text{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda})$, $\boldsymbol{\Lambda}_i(\phi, \rho)$. If the residuals exhibit a trend in time, then the entry in the second column is "trend"; otherwise, the entry is "no trend." A similar convention is used for trend in space, reported in the third column.

Table 3. Final model results

			edf				
Model	$\hat{\boldsymbol{\Lambda}}$	Variance $\hat{\sigma}^2$	f_1	f_2	Adj. R^2	Approx. AIC	Approx. BIC
(2)	*	0.066	3.99	258.07	0.77	−18,131.54	−16,298.03
$-f_1(age)$	**	0.145	–	221.09	0.43	−13,818.61	−12,270.63
$-f_2(no, e, year)$	***	0.092	3.99	–	0.68	−16,664.21	−16,629.43

*ARMA, $\hat{\boldsymbol{\Lambda}}_i(\hat{\phi} = 0.94, \hat{\rho} = -0.73)$; **ARMA, $\hat{\boldsymbol{\Lambda}}_i(\hat{\phi} = 0.96, \hat{\rho} = -0.65)$; ***ARMA, $\hat{\boldsymbol{\Lambda}}_i(\hat{\phi} = 0.89, \hat{\rho} = -0.58)$.

to defoliation in certain weather conditions. In that case, its effect would change over time.

Finally, we investigate the relative importance of the effects $f_1(\cdot)$ and $f_2(\cdot)$ in model (2) (Table 3). Dropping age yields increased AIC and BIC. The R^2 decreases from 0.77 to 0.43, confirming that age is an important covariate. Figure 4 shows the nonlinear effect of age as estimated with model (2); with increasing age, mean defoliation increases steeply until about 70 years, at which point the effect of age starts to flatten. Dropping the space–time smooth $f_2(\cdot)$ also yields an increased AIC. The R^2 decreases from 0.77 to 0.68, demonstrating that $f_2(\cdot)$ explains far less variability in the data compared with $f_1(age)$.

7. TREND ESTIMATES

The yearly estimated spatial trend maps of defoliation for spruce from model (2) in Figure 1 are based on a mean age of 74

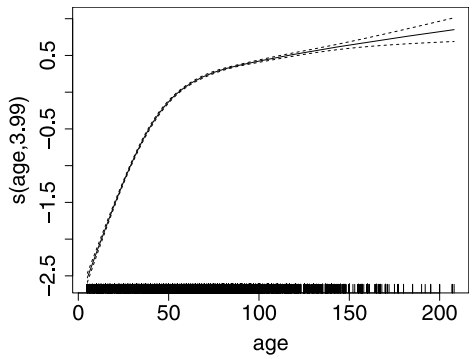

Figure 4. Age effect on the logit scale [model (2)].

years. The plots clearly show a cyclic pattern of mean defoliation. It is apparent that the predicted surface of mean defoliation is not additive in time. In the first periods of high defoliation (1985–1987 and 1992–1995), the areas with defoliation $> 30\%$ are small and mostly in the Black Forest. In the third period of high defoliation (2000–2007), the pattern changes; areas with $>30\%$ defoliation become larger and are mainly in the hills and planes of the Neckar regions and in the south of Baden Württemberg (southern Black Forest and the area close to the Lake of Konstanz), regions with loam and clay soils and the warmest climate in the state. Between 2005 and 2007, defoliation was greater than any observed since 1985, and areas with $>30\%$ defoliation extend over most of Baden-Württemberg.

Figure 5 shows the temporal trend for spruce averaged over the whole of Baden-Württemberg estimated with model (2). The confidence bands were constructed using the predictive distribution of the mean yearly defoliation, as described in Section 5.5. The top plot shows the estimated average yearly defoliation only at fitted values; thus it is comparable to estimates

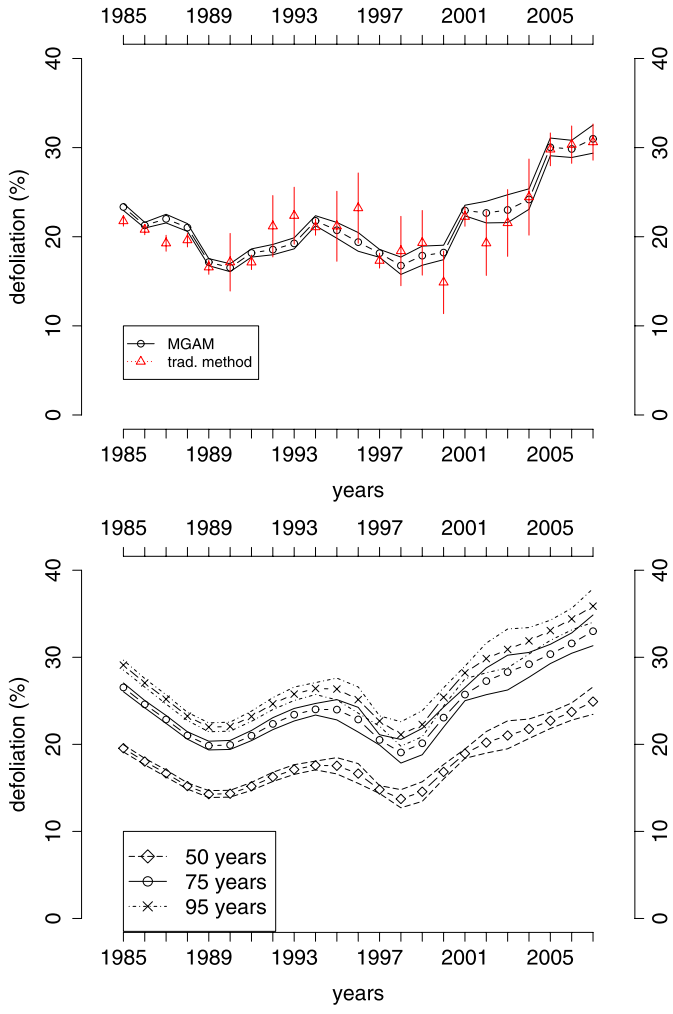

Figure 5. Estimated average defoliation of spruce with 95% confidence bands. (a) Estimates at fitted values (with age as observed) between 1985 and 2007 in Baden-Württemberg [model (2)]. The triangles indicate the estimates of the traditional method with 95% confidence intervals. (b) Estimates based on predictions on regular grid standardized for three ages: 50, 75, and 95 (the lower, median, and upper quartile ages in the data).

of the traditional method, which are also shown on the plot. The confidence bands overlap with the confidence intervals of the traditional method estimates in all years except 1996, and the overlap is relatively small in the years 1987, 1993, 1999, and 2000. Nearly all years with high differences between the traditional method and model (2) are years with a grid resolution of 16×16 km. In 2000, the year of the winter storm Lothar, which wiped out a number of grid locations, there also is a large discrepancy between the two methods. On the whole, the widths of the confidence intervals reflect the changes in grid resolution over the years for both methods. In the case of the traditional method, the yearly confidence intervals depend solely on the grid resolution in that year. The width of confidence bands based on model (2) changes smoothly, also depending on the previous and later years' grid resolutions. The bottom plot of Figure 5 shows the mean yearly defoliation predicted at all grid points observed, standardized for three ages: 50, 75, and 95 (the lower, median, and upper quartile ages in the data). The trend shown in the bottom plot is smoother than the trend estimated at fitted values in the top plot, but the two plots have the same general features. The trend plot standardized for median age is comparable in absolute value to the trend plot at fitted values. The cyclic pattern from 1985 to around 1998, with a period of approximately 10 years, does not appear to repeat itself in recent years. In fact, since 1998, the mean defoliation is increasing, and since 2004, the mean defoliation is significantly higher than that in any year since 1985. In 2005–2007, the defoliation is 30% for trees at a median age of 75 years and even greater for older trees. A mean defoliation of 30% is considered medium damage. The strong effect of age is interesting as well; mean defoliation is significantly lower in all years for trees age 50 than for trees age 75 or older.

Looking at the estimated temporal trends separately by growth regions in Figure 6 reveals, as in the spatial trend plots, that the cyclic patterns are not synchronized in the different growth regions. In addition, the patterns are different for the different regions. For instance, in the Black Forest there is a clear cyclic pattern, whereas in the pre-alpine region the mean defoliation is slowly increasing. In the Swabian Alb the pattern is similar to that in the Black Forest, but less pronounced in the earlier years. Comparing the estimated temporal trends with age, as observed at the fitted values with the raw regional trends and their confidence intervals as calculated with the traditional method, shows that the space–time model estimates tie in well with the raw regional trends. The confidence bands mostly overlap (data not shown).

8. DISCUSSION

Our proposed spatiotemporal model allows adequate assessment of forest health status, and the results demonstrate that forest health is damaged. For a given age structure, there is significant evidence of an increased trend in defoliation of spruce (*Picea abies* L.) since 2004 compared with the period 1985–2003. The spatial patterns of high crown damage shown in Figure 1 provide some indication of the main damaging factors. In the first periods of high defoliation, the most severe crown defoliation occurred in the Black Forest in soils on silicatic bedrock, such as granite. These soils are particularly susceptible to acidification. This suggests that acid deposition and soil acidification caused by pollution may have been the major factors contributing to damage at that stage. In contrast, in the most recent period

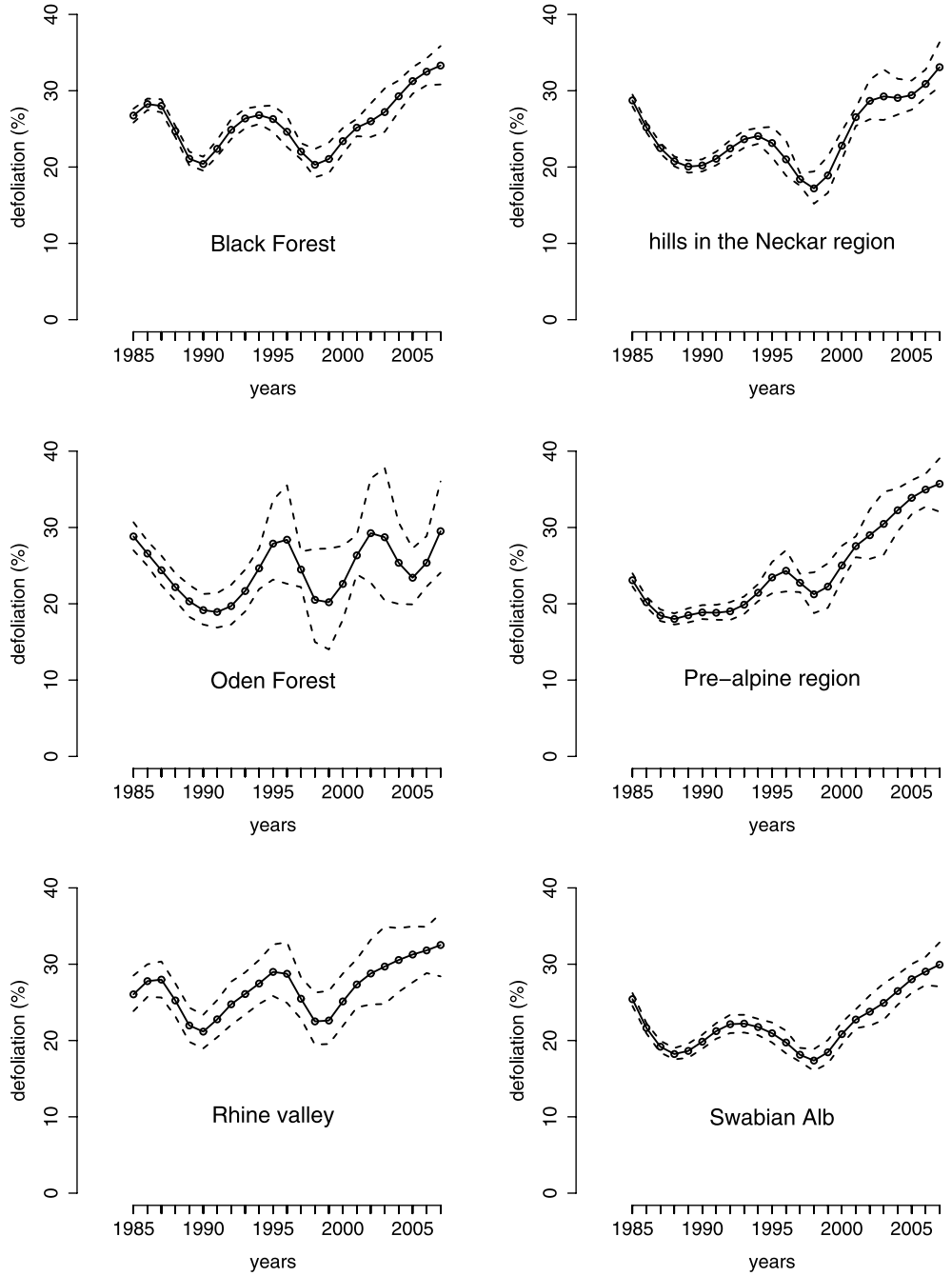

Figure 6. Estimated average defoliation of spruce with 95% confidence bands at a regular grid standardized for mean age 74 years for the different growth areas [model (2)]. The areas Black Forest and Baar/Black Forest are combined.

of high defoliation (since 2000), the greatest mean defoliation occurred in areas with loam and clay soils and a much warmer climate than the high-elevation area of the Black Forest. This suggests drought as a major factor in damage, because the fine texture and low water storage capacity of loam and clay soils intensify the effect of drought on trees. This observed change in spatial pattern supports the hypothesis that the main damaging factor recently changed from acidification caused by pollution to drought caused by extreme weather conditions, such as the extremely dry years of 2000–2003, combined with the cumulative effects of pollution, mostly nitrogen emissions from transportation and industrial processes.

Our approach is ideal for modeling space–time monitoring data observed on an irregular grid, where typically the spatial effect is not additive in time. The tensor product smooth for the spatiotemporal trend has intuitive appeal, making it straightforward to explain and interpret when communicating results to environmental policy makers. Our method provides the predicted maps and time trends, both with confidence bands, essential for environmental monitoring, such as checking for short-term spatial changes in the estimated mean defoliation. The maps also help guide decisions about lime application to some areas as a countermeasure to acidification of the soil. In addition, our proposed model has been used in simulations

to optimize the monitoring network, under the financial constraints that preclude a 4×4 km grid resolution in the future (Augustin 2006).

Because the space–time smoother is set up using the scale-invariant marginal spatial and temporal bases, we avoid having to deal with different scales in time and space. This setup allows different degrees of smoothness relative to the different covariate axes, because the tensor product smooth allows the use of different penalty matrixes for each dimension. Because of the different penalty matrixes, the resulting smooth also is scale-invariant. In addition, our approach provides the flexibility of using different bases for the different dimensions. Here we use different bases in space and time while still maintaining space–time scale invariance.

Simulation results show that our proposed model can separate the smooth temporal trend from temporal correlation at site level and that ARMA parameters can reliably be recovered. In 100 response vectors of size 400 with autocorrelated errors following (1) an ARMA process ($\phi = 0.9$ and $\rho = -0.7$) and (2) an AR1 process ($\phi = 0.99$) with a strong temporal trend composed of (a) a polynomial of time and (b) a linear temporal trend, the GAMM fitted with the temporal smooth and an ARMA correlation structure outperformed a GAMM fitted with the temporal smooth and an independent error structure. For both models (1a) and (1b) and models (2a) and (2b), the median difference in mean squared error between accounting for and not accounting for the ARMA correlation structure in the GAMM is negative; in addition, the medians of the parameter estimates of ϕ and ρ are fairly close to the true values.

As a reviewer pointed out, a natural extension of our tensor product smooth is to consider two separate smoothing parameters in the two spatial directions, allowing for anisotropy. We did consider two separate smoothing parameters in the two spatial directions, but the spatial residual plots did not improve with this model.

Our model includes mean age, a well-known risk factor for defoliation, as a covariate, allowing the prediction of trends in mean defoliation standardized for age. The predicted temporal trend standardized for age shown in Figure 5 (bottom) is much clearer than the estimated temporal trend of the traditional method (top). The model allows extensions, such as adding site-specific covariates on pollution and weather, needed for a more thorough investigation of the defoliation processes. In particular, it would be useful to disentangle the effects of pollution, climate, and weather.

The spatial part of the space–time smoother $f_2(\cdot)$ is driven mainly by site-specific characteristics, such as geology, soil type, nutrient balance, local pollution levels, and local weather conditions. All of these covariates are correlated in space, making it impossible to distinguish between site-specific effects and spatial effects modeled using eastings and northings. Similar findings on such confounding have been reported by Augustin et al. (2007). Thus with model (2), the emphasis is on prediction rather than on establishing relationships between explanatory and response variables. In environmental monitoring, identifying model parameters is difficult, and often it is impossible to separate the spatial effect from other effects that vary smoothly in space, especially if important covariates (e.g., pollution levels and weather information) are not included. The lack of covariate effects in the model is disappointing but not surprising, because the available covariates are mostly constant in time, and the analysis clearly demonstrates space–time effects.

The model takes into account temporal and spatial correlation via the space–time smoother. In addition, the possibility exists of accounting for unexplained correlation via a very general structure of the error distribution and the inclusion of random effects. Model (2) models temporal correlation of errors within location as an ARMA process, rather than also allowing for a spatial correlation structure. This makes sense because we are dealing with averages of defoliation of trees that typically hold their needles for 7 years. Adding a further *spatial* correlation structure for the errors or random effects at location i would lead to confounding between spatial trend of the space–time smoother and the assumed spatial correlation. Because we are aiming to model the spatial trend using easting, northing, and other covariates, these explicit spatial error models are suitable here. Our model choice in terms of correlation structure is backed up by preliminary work demonstrating that the temporal correlation is much stronger than the spatial correlation. Our diagnostic plots of residuals show that the error structure was modeled adequately; this means that inference based on the model can be relied on for making management decisions.

[Received February 2007. Revised February 2009.]

REFERENCES

Anon (2001), "Forest Condition in Europe, UNECE and EC," technical report, Federal Research Centre for Forestry and Forest Products (BFH).

Augustin, N. (2006), "Terrestrial Inventory Survey (TWI): Simulation Study to Investigate Different Sampling Schemes," technical report, University of Bath, Dept. of Mathematical Sciences, Bath, U.K.

Augustin, N., Lang, S., Musio, M., and von Wilpert, K. (2007), "A Spatial Model for the Needle Losses of Pine Trees in the Forests of Baden-Württemberg: An Application of Bayesian Structured Additive Regression," *Journal of the Royal Statistical Society, Ser. C*, 56 (1), 29–50.

Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman & Hall/CRC.

Box, G., and Pierce, D. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.

Evers, J., Franz, C., Körver, F., and Ziegler, C. (1997), *Bildserien zur Einschätzung von Kronenverlichtungen bei Waldbäumen. Arbeitsgemeinschaft Dauerbeobachtungsflächen der Länder und des Bundes in Deutschland*, Kassel, Germany: Verlag M. Faste.

Fahrmeir, L., and Lang, S. (2001), "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors," *Journal of the Royal Statistical Society, Ser. C*, 50, 201–220.

Fuentes, M., Chen, L., Davis, J., and Lackmann, G. (2005), "Modeling and Predicting Complex Space–Time Structures and Patterns of Coastal Wind Fields," *Environmentrics*, 16 (5), 449–464.

Glasbey, C., and Allcroft, D. (2008), "A Spatiotemporal Auto-Regressive Moving Average Model for Solar Radiation," *Journal of the Royal Statistical Society Series, Ser. C*, 57, 343–355.

Hastie, T., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman & Hall.

Huang, H., Martinez, F., Mateu, J., and Montez, F. (2007), "Model Comparison and Selection for Stationary Space–Time Models," *Computational Statistics & Data Analysis*, 51 (9), 4577–4596.

Kamman, E. E., and Wand, M. P. (2003), "Geoadditive Models," *Journal of the Royal Statistical Society, Ser. C*, 52, 1–18.

Kass, R., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90 (430), 773–795.

Kneib, T., and Fahrmeir, L. (2006), "Structured Additive Regression for Categorical Space–Time Data: A Mixed Model Approach," *Biometrics*, 62, 109–118.

Lin, X., and Zhang, D. (1999), "Inference in Generalized Additive Mixed Models by Using Smoothing Splines," *Journal of the Royal Statistical Society, Ser. B*, 61, 381–400.

MacNab, Y., and Dean, C. (2001), "Autoregressive Spatial Smoothing and Temporal Spline Smootheing for Mapping Rates," *Biometrics*, 57, 949–956.

Musio, M., Augustin, N. H., and von Wilpert, K. (2008), "Geoadditive Bayesian Models for Forestry Defoliation Data: A Case Study," *Environmetrics*, 19 (6), 630–642.

Musio, M., von Wilpert, K., and Augustin, N. (2007), "Crown Condition as a Function of Soil, Site and Tree Chracteristics," *European Journal of Forest Research*, 126 (1), 91–100. Online issue at *http://dx.doi.org/10.1007/s10342-006-0132-8*.

Pinheiro, J., and Bates, D. (2000), *Mixed-Effects Models in S and S-PLUS*, New York: Springer-Verlag.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Core Team (2008), *nlme: Linear and Nonlinear Mixed Effects Models*, R package version 3.1-89.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0.

Ribeiro, Jr., P. J., and Diggle, P. J. (2001), "geoR: A Package for Geostatistical Analysis," *R-NEWS*, 1 (2), 14–18. ISSN 1609-3631.

Richardson, S., Abéllan, J., and Best, N. (2006), "Bayesian Spatio-Temporal Analysis of Joint Patterns of Male and Female Lung Cancer Risks in Yorkshire (UK)," *Statistical Methods in Medical Research*, 15, 385–407.

Ruppert, D., Wand, M., and Carroll, R. (2003), *Semiparametric Regression*, Cambridge: Cambridge University Press.

Sahu, S., Gelfand, A., and Holland, D. (2007), "High Resolution Space–Time Ozone Modeling for Assessing Trends," *Journal of the American Statistical Association*, 102, 109–118.

Schöpfer, W., and Hradetzky, J. (1984), "Der Indizienbeweis: Luftverschmutzung massgebliche Ursache der Walderkrankung, *Forstwissenschaftliches Centralblatt*, 103, 231–248.

Seidling, W. (2001), "Integrated Studies of Forest Ecosystem Conditions: Multivariate Evaluations on Tree Condition for Two Areas With Distinct Deposition Gradients," technical report, UNECEEC Flemish Community, Geneva, Brussels, Gent.

Shaddick, G., and Wakefield, J. (2002), "Modelling Daily Multivariate Pollutant Data at Multiple Sites," *Journal of the Royal Statistical Society, Ser. C*, 51, 351–372.

Silverman, B. (1985), "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting," *Journal of the Royal Statistical Society, Ser. B*, 47, 1–52.

Wahba, G. (1983), "Bayesian Confidence Intervals for the Cross Validated Smoothing Spline," *Journal of the Royal Statistical Society, Ser. B*, 45, 133–150.

Wood, S. (2004), "Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models," *Journal of the American Statistical Association*, 99 (467), 673–686.

——— (2006a), *Generalized Additive Models. An Introduction With R*, Boca Raton: Chapman & Hall/CRC.

——— (2006b), "Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models," *Biometrics*, 62 (4), 1025–1036.

——— (2006c), "On Confidence Intervals for Generalized Additive Models Based on Penalized Regression Splines," *Australian and New Zealand Journal of Statistics*, 48 (4), 445–464.

Xu, K., Wikle, C., and Fox, N. (2005), "A Kernel-Based Spatio-Temporal Dynamical Model for Nowcasting Weather Radar Reflectivities," *Journal of the American statistical Association*, 100 (472), 1133–1144.