



Citation for published version:

Augustin, NH, Musio, M, von Wilpert, K, Kublin, E, Wood, SN & Schumacher, M 2009, 'Modeling spatiotemporal forest health monitoring data', *Journal of the American Statistical Association*, vol. 104, no. 487, pp. 899-911.
<https://doi.org/10.1198/jasa.2009.ap07058>

DOI:

[10.1198/jasa.2009.ap07058](https://doi.org/10.1198/jasa.2009.ap07058)

Publication date:

2009

Document Version

Peer reviewed version

[Link to publication](#)

This is the author's accepted version. The published version is available from
<http://dx.doi.org/10.1198/jasa.2009.ap07058>

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Modelling spatio-temporal forest health monitoring data

Nicole H. Augustin, Monica Musio, Klaus von Wilpert,
Edgar Kublin, Simon N. Wood and Martin Schumacher*

May 5, 2010

Abstract

Forest health monitoring schemes were set up across Europe in the 1980's in response to concern about air pollution related forest die back (*Waldsterben*) and have continued since then. Recent threats to forest health are climatic extremes likely to be due to global climate change, 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 between 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 to use generalized additive mixed

*Nicole Augustin, Simon Wood: Department of Mathematical Sciences, University of Bath, Bath, UK; Monica Musio: Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy; Edgar Kublin, Klaus von Wilpert: Forest Research Centre Baden-Württemberg, Freiburg, Germany and Martin Schumacher: University Hospital Freiburg, University Freiburg, Freiburg, Germany.

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 hence 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 non-statisticians, such as environmental policy makers. The model incorporates a non-linear 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 modelled with the space-time smooth, we account for random temporal correlation at site level by an auto-regressive 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 semi-variogram and the empirical auto-correlation function respectively.

Our method provides, for the first time, the predicted spatial and temporal trends and their confidence intervals. These trends show that since 2004 there is significant evidence for an increased trend in defoliation in spruce. In addition they suggest that there was a recent switch in the primary drivers of damage: Recent damage can mainly be associated with drought years due to climate change and cumulative effects of pollution, whereas initial damage can solely be associated with pollution.

Keywords: air pollution, climate change, generalized additive mixed models (GAMMs), environmental monitoring, forest damage, Norway spruce (*picea abies* L.), tree defoliation, spatio-temporal model, tensor product smooths, space-time data.

1 Introduction

The main exogenous factors influencing European forest eco-systems are pollution, weather and biological pests. These three factors interact with each other and with soil characteristics, which can amplify adverse effects on the eco-system. In particular the effects of deposition of pollutants through the air and rainfall are long-term and accumulative. They cause acidification of soil which disequilibrates the soil chemistry, including the nutrient and metal availability, and this finally causes the washing out of essential alkaline macro nutrients in the root area. The symptoms seen on 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 this also makes biological pests more common, causing more direct stress. All of these three factors and their interaction have caused substantial damage to the forest eco-systems in Europe in the last four 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 which have brought down 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, probably due to anthropogenic global climate change, have recently increased forest damage.

We analyze yearly data on percentage tree defoliation from a monitoring survey carried out annually in Baden-Württemberg, Germany since 1983, the Terrestrial Crown Condition Inventory (TCCI). 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, so that substantial changes in

damage can be detected promptly. Ideally, the survey should facilitate the investigation of how forest damage is associated with 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 the trees and the soil (Augustin et al., 2007; Musio et al., 2007, 2008).

Here the investigation is restricted to relating the site characteristics to defoliation, since 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 aid in making recommendations on environmental policies.

The traditional method used for estimation in Baden-Württemberg calculates estimates of the mean defoliation separately from the TCCI data for each year. It is only suitable for temporal trend estimation and does not take any spatial or temporal structure of the data into account. Neither does it make use of possible predictor variables. Ignoring correlation in the data when estimating the trend of defoliation could result in biased estimates and consequently give wrong answers.

A new modelling approach is required which 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. Since it is also 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 modelled as non-linear effects. Recent methodological development on space-time modelling for point referenced data has been mainly on hierarchical Bayes models (see Banerjee et al. (2004) for an introduction) and geostatistical models. Applications are mainly in environmental pollution monitoring (Sahu et al., 2007; Shaddick and Wakefield, 2002), meteorology (Glasbey and Allcroft, 2008; Xu et al., 2005; Fuentes et al., 2005) and epidemiology (Richardson et al.,

2006; MacNab and Dean, 2001).

Generalized additive mixed models (GAMMs) (Lin and Zhang, 1999; Kamman and Wand, 2003; Ruppert et al., 2003) have been used to model spatio-temporal 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, which have no random effect term. In most recent applications of GAMMs to spatio-temporal data, the space and time dimensions are modelled separately. For example, Kneib and Fahrmeir (2006) present a generalized additive model 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) use a generalized additive mixed model for analyzing spatio-temporal mortality data by including effects for space and time in an additive manner. Here we use the general methodology of Wood (2004, 2006b) for constructing scale invariant tensor product smooths of the space-time dimension. The proposed method accommodates any response which 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 non-linear effects of covariates such as altitude, the space-time interaction can be modelled 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, hence allowing assessment of changes in trends of defoliation.

2 The Terrestrial Crown Condition Survey

The data analyzed are from the Terrestrial Crown Condition Inventory (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 the period 1983-1986, in 1989, 1991, 1994, 1997 and 2001, 8×8 in 1987, 1988, 2005-2007; and 16×16 in the remaining years. See Figure 1 where grid locations are shown 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 1475 sampling grid points. There are 7 natural growth regions which are used for forest management and reporting purposes. In the Black Forest area, the sampling grid is most dense because of over-proportionally dense forest cover compared to other areas. The survey is carried out in alignment with the European level I monitoring programme 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 which 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 4 subplots oriented among the main compass directions 50m 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 centre was 32m. Selected trees are permanently marked and re-assessed during subsequent surveys. Trees that are removed are replaced by newly selected trees.

The survey protocol has several features to avoid observer bias: The assessment is stan-

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 EU level, federal level and state level in order 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). Also, the survey is carried out with two independent observers per tree and estimates are pooled. As a further check, 20% of plots are re-assessed by experienced members of staff. A temporal trend in observer error can be excluded by this procedure since 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.) followed by beech, fir, pine and other species. Here we limit all subsequent analysis to spruce. For comparing variability within and between sampling grid points a random effects model is fitted to all data 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, b_i is a random effect parameter for the i th location with $i = 1, \dots, 1475$ with $j = 1, \dots, a_i$ spruce trees at location i , where a_i is a maximum of 24. We assume $b_i \sim N(0, \sigma_b^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. 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 asymptotic normality of restricted maximum likelihood estimates (σ^2 and σ_b^2 are distant from zero 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 within grid points in most years. Hence we would expect to explain more variability by introducing site rather than tree specific covariates in the model. In the years where the sampling grid resolution is 16×16 , the confidence interval of the between location variability, σ_b , is substantially wider than in years with a smaller grid resolution. In the year 2000 of the winter storm Lothar, variability between grid locations is substantially higher than within sampling location, compared to other years.

4 A spatio-temporal model for defoliation

4.1 Requirements

In the following we state the requirements of the model based on the underlying physical and biological processes. The most important aspect of the data the model needs to handle is the *space-time trend and interaction*. This should be modelled using a smooth three-dimensional function of space and time. The temporal trend of defoliation differs between areas because site characteristics such as geology, soil type, climatic conditions and pollution levels differ within the survey area and hence it is expected that the defoliation happens at different strengths and cycles depending on the location. For example, the Black Forest area has a number of characteristics which make it more susceptible to pollution. The geology is mainly granite or Triassic sandstone bedrock and soils on such stone are more prone to acidification. Also, a lot 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). Hence defoliation is expected to be more severe and rapid than in the Swabian Alb or the 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. Also, we expect the defoliation to depend on the previous year's defoliation in more subtle ways. For example trees which are damaged may be more susceptible to damage in subsequent years. Hence the correlation in time is made up from two parts: First, a temporal trend due to site characteristics and other conditions, which may be modelled by a smooth function of space and time. Second, the random variability in measurements at the tree and hence site level will be 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. In the survey data there is a large range of tree ages, from 5 to 208 years. Hence the model should contain a smooth function of age. In order 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 case there is *spatial correlation* between grid locations, not accounted for by the smooth space-time function, this is most likely to be caused by spatial trend in the site characteristics. Such trend should be modelled by functions of site characteristics such as geology, soil type, type of slope, and so forth. There is also likely to be spatial correlation within a sampling grid point where typically 24 trees are observed. Since the x and y coordinates are not available at the tree-level it is difficult to model the spatial correlation within a grid point. We can only use age of the three relevant tree specific covariates (social class, diameter and age). But social class was excluded 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 diameter was excluded because it is strongly

correlated with age. The other covariates are only available at the grid point resolution. Therefore we aggregate the data within grid points and hence the model only needs to 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, hence there are generally no big differences in terms of height and dominance. In addition, the model is to be used as an environmental monitoring tool and tree specific prediction is not of interest. Hence the aggregated response variable is the average defoliation \bar{y}_{it} at the i th grid point location with $i = 1, \dots, 1475$ and $j = 1, \dots, a_{it}$ trees, with a maximum of 24 trees at location i in year t , with $t = 1985, \dots, 2007$, where weights $1/a_{it}$ will be used in all analysis. We assume \bar{y}_{it} follows a normal distribution by making use of the central limit theorem. Note that to simplify notation we will refer to \bar{y}_{it} as y_{it} in the following.

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 above stated requirements: It completely ignores the longitudinal and spatial structure of the data and can be described as fitting a linear model to years separately:

$$y_{ij} = \alpha + \epsilon_{ij} \quad (1)$$

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 ϵ is distributed as $N(\mathbf{0}, \mathbf{I}\sigma^2)$. Hence 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 in-

dependence of the data in time and space, which is not given here, and consequently 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

The proposed model is a generalized additive mixed model (GAMM) (Lin and Zhang, 1999; Fahrmeir and Lang, 2001). It allows for a complex stochastic structure for dealing with the above requirements and is the result of some extensive model selection described in section 6:

$$\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_2(\text{no}_i, e_i, \text{year}_t). \quad (2)$$

The response, mean proportion defoliation, is

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

at location $i = 1, \dots, 1475$ and year $t = 1, \dots, 23$. 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 since 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 modelling 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. Since 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 below).

The residual error vector ϵ is distributed as $N(\mathbf{0}, \sigma^2 \mathbf{\Lambda})$, where $\mathbf{\Lambda}$ is block diagonal with the i^{th} subvector ϵ_i having covariance matrix $\mathbf{\Lambda}_i$ relating to residuals at one location over time. The $\mathbf{\Lambda}_i$ 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 ϵ_i reflect the temporal correlation in the error, modelled by a mixed autoregressive and moving average (ARMA) process, using the approach given in 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 zero. So $\mathbf{\Lambda}_i$ 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 above is a restricted version of the following model, which we use to describe the methodology in general:

$$\text{logit}E(y_{it}) = \mathbf{X}_{it}\boldsymbol{\theta} + f_1(\text{age}_{it}) + f_2(\text{no}_i, e_i, \text{year}_t) + f_3(w_{3it}) + \dots + f_k(w_{kit}) + \mathbf{Z}_{it}\mathbf{b} \quad (3)$$

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; $f_3 \dots f_k$ are smooth functions of covariates $w_3, \dots w_k$; \mathbf{Z}_{it} is a row of a random effects model matrix; the vector of random effects coefficients \mathbf{b} is distributed as $N(0, \boldsymbol{\psi})$ with unknown positive definite covariance

matrix ψ . We assume that the error vector $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda})$ with a covariance matrix $\mathbf{\Lambda}$ reflecting the assumed error structure. For example in Model 2 this is a block diagonal with the i^{th} subvector ϵ_i having covariance matrix $\mathbf{\Lambda}_i$ reflecting temporal correlation at location i , but no spatial correlation. Alternative structures are possible, such as an exponential spatial correlation between locations (Pineiro and Bates, 2000). Model 3 is extremely general and care needs to be taken not to overfit, since in most data situations there will be identifiability issues. For example, the spatial effect of $f_2(\cdot)$ will be confounded with other site specific effects. For this reason, we carry out a careful step-by-step model selection, as described in Section 6.

5.2 A 3-d 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 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:

$$f_{year}(year) = \sum_{i=1}^P \alpha_i a_i(year) = \mathbf{X}_{year} \boldsymbol{\alpha} \quad \text{and} \quad f_{no,e}(no, e) = \sum_{l=1}^L \beta_l b_l(no, e) = \mathbf{X}_{no,e} \tilde{\boldsymbol{\beta}},$$

where α_i and β_l are parameters while a_i and b_l are basis functions with $i = 1, \dots, P$ and $l = 1, \dots, L$. The \mathbf{X}_{year} and $\mathbf{X}_{no,e}$ are the marginal model matrices 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} could be converted into a smooth function of no and e . For this to happen 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:

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

and this 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 matrices \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 i th 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

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

The matrices \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} . For example, here the cubic penalty for f_{year} is $J_{year}(f_{year}) = \int f_{year}''(year)^2 dyear$. 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, $\beta_l(year)$,

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

Hence 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 re-expressed, with β as defined in equation 4:

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

In Model 2 the cubic regression spline basis is used for the temporal smooth, hence the elements of β are the actual function heights, which therefor 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 modelling software (Lin and Zhang, 1999; Wood, 2004). The mixed model approach gives a self consistent and computationally tractable way to handle smoothing and deal with autocorrelation simultaneously. It involves setting prior distributions on coefficients of $f_k()$, derived from the roughness penalties, and treating the coefficients as random effects. To avoid improper random effect distributions some re-parametrization is generally 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). Hence Model 3 can be rewritten as:

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

$$= \mathbf{X}_{it}^*\boldsymbol{\theta}^* + \mathbf{Z}_{it}^*\mathbf{b}^* \quad (6)$$

with $y_{it} = E(y_{it}) + \epsilon_{it}$. The fixed effect components of f_1, \dots, f_k , relating to unpenalized coefficients (with completely improper priors), $\mathbf{X}_{fit}\boldsymbol{\theta}_f$, are absorbed into $\mathbf{X}_{it}^*\boldsymbol{\theta}^*$ and the

components relating to penalized coefficients with proper priors, $\mathbf{X}_{fit}\boldsymbol{\theta}_f$, are absorbed into $\mathbf{Z}_{it}^*\mathbf{b}^*$. In this re-parametrized form of the model we have model matrices and vectors for fixed and random effects which contain, in addition to the usual fixed and random components, additional components from the smoothers. Then \mathbf{b}^* is distributed as $N(0, \boldsymbol{\psi}^*)$ with matrix $\boldsymbol{\psi}^*$ depending also 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 $N(\mathbf{0}, \sigma^2\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ contains the correlation parameters. Since the GAMM can be expressed as a generalized linear mixed model (GLMM) (as in equation 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:

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

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 , $v_{it} = \frac{1}{g'(\hat{\mu}_{it}^b)^2}$. Defining $\mu_{it}^b = E(\widehat{y_{it}}|\mathbf{b}^*)$, then \mathbf{u} is a vector of pseudo data, with elements

$$u_{it} = g'(\hat{\mu}_{it}^b)(y_{it} - \hat{\mu}_{it}^b) + (X_{it}^*\hat{\boldsymbol{\theta}}^* + Z_{it}^*\hat{\mathbf{b}}^*).$$

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

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

with \mathbf{b}^* distributed as $N(0, \boldsymbol{\psi}^*)$ and $\boldsymbol{\epsilon}$ distributed as $N(\mathbf{0}, \mathbf{W}^{-1}\sigma^2)$. 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 respectively, obtained by treating the

smooths as penalized fixed effects. Then, if \mathbf{y} denotes the response and $\hat{\boldsymbol{\mu}}$ the predictions of \mathbf{y} according to the estimated smooth model, we have

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

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

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

where $n = 7864$, 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 the covariance matrix $\boldsymbol{\Lambda}$ and are obtained in the usual way (see the end of Section 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 hence it is suitable in situations where there is a large sample size 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, in particular for data sampled irregularly in space.

5.4 Variance estimation

We explain the variance estimation in terms of the reparametrized version of the general model in equation (6). Let $\boldsymbol{\gamma}^T = (\boldsymbol{\theta}^T, \boldsymbol{\theta}_f^T, \mathbf{b}_f^T)$ 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 since we are interested in making inferences about model components in $\boldsymbol{\gamma}$, some of which are treated as random variables in the mixed modelling. In addition, the

Bayesian model representation gives a self-consistent basis for constructing confidence intervals, or more precisely credible intervals, since the posterior distribution of the model parameters is known. In comparison, frequentist confidence intervals are not as straightforward, see also (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, it is first necessary to calculate the covariance matrix for the response data implied by the estimated random effects structure *excluding* the smooth terms: $\mathbf{Q} = \mathbf{Z}^T \boldsymbol{\psi} \mathbf{Z} + \Lambda \sigma^2$, where \mathbf{Z} and $\boldsymbol{\psi}$ are defined as in equation (3). Then

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

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 in Lin and Zhang (1999). The degrees of freedom per element of $\boldsymbol{\gamma}$ can be estimated from the leading diagonal of $(\bar{\mathbf{X}}^T \mathbf{Q}^{-1} \bar{\mathbf{X}} + \mathbf{S})^{-1} \bar{\mathbf{X}}^T \mathbf{Q}^{-1} \bar{\mathbf{X}}$.

5.5 Trend estimation

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

$$\hat{y}_t = \frac{1}{n} \sum_{i=1}^n \text{logit}^{-1} \left(\mathbf{X}_{it}^* \hat{\boldsymbol{\theta}}^* + \mathbf{Z}_{fit} \hat{\mathbf{b}}_f \right)$$

where \mathbf{X}_{it}^* and \mathbf{Z}_{fit} are evaluated at the observed values. The random effects which 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 as random variation and taken at their (prior) expected value (i.e. zero). Note that we are not using 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 the estimates are based on, would downweight certain grid locations where the tree species are not entirely spruce, i.e. less than 24 spruce trees are observed and 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 $\gamma^T = (\boldsymbol{\theta}^T, \boldsymbol{\theta}_f^T, \mathbf{b}_f^T)$ with the posterior covariance matrix as in (7) and obtain from this a sample from the predictive distribution of the response. Then the p th draw from the posterior distribution is

$$\widehat{E}_p(y_{it}) = \hat{y}_{itp} = \text{logit}^{-1} \left(\mathbf{X}_{it}^* \hat{\boldsymbol{\theta}}_p^* + \mathbf{Z}_{fit} \hat{\mathbf{b}}_{fp} \right).$$

Averaging over i yields the predictive distribution of \hat{y}_t :

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

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{y}_{tp}) respectively (as presented in the results section below). If we want to predict defoliation on a regular grid and, for example, a standard age, then the \mathbf{X}_{it}^* and \mathbf{Z}_{fit} are evaluated for these prediction data. Otherwise the same procedure as described above is carried out.

5.6 Model diagnostics

In order to check whether the model has eliminated residual correlation we investigate residuals in space and time. For models where we do not assume an independent error structure, normalized residuals are created as described in Pinheiro and Bates (2000). In the case of the Model 2, let the vector of errors at location i be $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2 \boldsymbol{\Lambda}_i)$, as defined

in 4.3. Each Λ_i has the same parameters ϕ and ρ but varying structure due to missing years at location i . Then the residual vector \mathbf{r}_i is given by

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

with $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$ 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 semi-variogram 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 semi-variogram of the residuals by year. Then the residuals are permuted 999 times within year and envelopes are computed by taking, at each lag, the maximum and minimum values of the semi-variograms 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, $z(1 - \alpha/2)$ 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 above for low lags, because the normalized residuals are based on parameter estimates rather than 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 we check the space-time structure and finally we select covariates.

Step one: Initially models listed in Table 1 were checked for gross violations of distributional assumptions, using the type of residual diagnostic plots described above. As the

Table 1: Model selection *step one*: This shows the results under the traditional method (“trad”) and different versions of Model 2 (a, b, c, d), with respect to two kinds of temporal error structure. The first kind of 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.

model	residuals	residuals
	in time	in space
”trad” $y_{it} = \alpha_t + \epsilon_{it}$ with $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{A})$, diagonal \mathbf{A} with $1/a_{it}$ *	corre- lated	corre- lated
(a) $\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_2(n_i, e_i, \text{year}_t)$ $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{A})$, diagonal \mathbf{A} with $1/a_{it}$	corre- lated	uncorre- lated
(b) $\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_2(n_i, e_i, \text{year}_t)$ $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda})$, first order autoregressive (AR) with $\mathbf{\Lambda}_i(\phi)$ with $\epsilon_{it} = \phi \epsilon_{it-1}$ and along diagonal $1/a_{it}$	corre- lated	uncorre- lated
(c) Model 2 $\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_2(n_i, e_i, \text{year}_t)$ (Model 2) $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda})$, ARMA with $\mathbf{\Lambda}_i(\phi, \rho)$ and along diagonal $1/a_{it}$	uncorre- lated	uncorre- lated
(d) $\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_2(n_i, e_i, \text{year}_t) + b_i$ $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{A})$, \mathbf{A} as in (a) and random effect b_i i.i.d. $N(0, \sigma_b^2)$	corre- lated	uncorre- lated

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

worst possible model we fit an adapted version of the traditional method’s model, simultaneously to all years. In order to make the traditional method comparable to 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 account of the fact 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 greater than two years. The semi-variograms 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, modelled with the 3-dimensional function $f_2(n_i, e_i, year_t)$ in Model 2. An obvious question is whether we can replace $n_i, e_i, year_t$ by other covariates which 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 is currently not possible since these variables are not available. Another idea is to replace some of the space-time interaction modelled 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 since it would imply that the defoliation mechanism is changing with time. In order to check whether the space-time interaction is required, different types of functions of n_i, e_i and $year_t$ were compared by plotting normalized residuals versus time and smoothing residuals in time and in space. The results 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 carry out forward variable selection on Model 2. Forward selection is preferred to backward selection for computational reasons, and is defensible since 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 of the covariates 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 Model 2 as the best model, i.e. no other covariate is selected.

Table 2: Model selection *step two*: This shows the results under models with different types space-time interactions. For all models it is assumed that $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{\Lambda})$, $\mathbf{\Lambda}_i(\phi, \rho)$. 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.

space-time effect	residuals	
	in time	in space
$\text{logit}E(y_{it}) = f_1(\text{age}_{it})$	trend	trend
$\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_2(n_i, e_i, \text{year}_t)$ (Model 2)	no trend	no trend
$\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_3(n_i, e_i) + f_4(\text{year}_t)$	no trend	trend
$\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_3(n_i, e_i)$	trend	no trend
$\text{logit}E(y_{it}) = f_1(\text{age}_{it}) + f_4(\text{year}_t)$	no trend	trend

Note that we only allow for additive covariate effects, which may be a restriction here. Some covariates, e.g. soil type, may contribute to defoliation in certain weather conditions. In that case its effect would change over time.

Finally, we investigate the relative importance of the effects $f_1()$ and $f_2()$ in Model 2 (Table 3). Dropping age yields an increased AIC and BIC. The R^2 reduces from 0.77 to 0.43, confirming that age is an important covariate. Figure 4 shows the non-linear effect of age as estimated with Model 2: with increasing age the mean defoliation increases steeply until about 70 years where the effect of age starts to flatten. Dropping the space-time smooth $f_2()$ yields also an increased AIC. The R^2 reduces from 0.77 to 0.68 showing that $f_2()$ explains far less variability in the data compared to $f_1(\text{age})$.

Table 3: Final model results. The edf are effective degrees of freedom.

model	variance		edf		adj.	approx.	approx.
	$\hat{\Lambda}$	$\hat{\sigma}^2$	f_1	f_2	R^2	AIC	BIC
(2)	*	0.066	3.99	258.07	0.77	-18131.54	-16298.03
- $f_1(\text{age})$	**	0.145	-	221.09	0.43	-13818.61	-12270.63
- $f_2(\text{no}, e, \text{year})$	***	0.092	3.99	-	0.68	-16664.21	-16629.43

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

7 Trend estimates

The yearly estimated spatial trend maps of defoliation for spruce from Model 2 in Figure 1 are based on the mean age of 74 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, 1992 - 1995) the areas with defoliation above 30% are small and mostly in the Black Forest. In the third period of high defoliation (2000 - 2007) the pattern changes, areas with more than 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), both regions with loam and clay soils and the warmest climate in the state. Between 2005 and 2007 defoliation was higher than ever observed since 1985 and areas above 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, hence it is comparable to estimates of the traditional method which are also shown on the plot. The con-

confidence bands overlap with the confidence intervals of the traditional method estimates in all years except in 1996, and in the years 1987, 1993, 1999 and 2000 the overlap is relatively small. Nearly all years with high differences between the traditional method and the Model 2 are years with a grid resolution of 16 x 16 km. In 2000, the year of the winter storm 'Lothar' which wiped out a number of grid locations, there is also a large discrepancy between the two methods. On the whole, the width of the confidence intervals reflect the change in grid resolution over the years for both methods. In the case of the traditional method the yearly confidence intervals solely depend 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 ever observed, standardized for three ages: 50, 75 and 95 (these are the lower, median and upper quartile age in the data). The trend in the bottom plot is smoother than the trend estimated at fitted values in the top plot, but has 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 between 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 in any year since 1985. In 2005 to 2007 the defoliation was at 30% for trees at median age of 75 years and even higher for older trees. A mean defoliation of 30% is considered as medium damage. It is also interesting to see the strong effect of age; the mean defoliation for trees at 50 years is significantly lower in all years than for trees that are 75 or older.

Looking at the estimated temporal trends separately by growth regions in Figure 6 reveals, as the spatial trend plots, that the cyclic patterns are not synchronized in the different growth regions. Also, 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 the one 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 (Figure not shown).

8 Discussion

Our proposed spatio-temporal model allows an adequate assessment of the forest health status and the results show that the forest health is damaged. For a given age structure there is significant evidence for an increased trend in defoliation of spruce (*Picea abies* L.) since 2004 compared to the period 1985-2003. The spatial patterns of high crown damage in Figure 1 provide some indication of the main damaging factors. In the first periods of high defoliation the most severe crown defoliation was in the Black Forest, with soils on silicatic bedrocks 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. By contrast, in the most recent period of high defoliation (since 2000), the highest mean defoliation is in areas with loam and clay soils and a much warmer climate than the high elevation area of the Black Forest. This is suggestive of drought as a major factor in damage, since the fine texture and low water storage capacity of loam and clay soils intensifies the effect of drought on trees. This observed change in spatial pattern supports the following hypothesis that the main damaging factor has recently changed from acidification caused by pollution to drought, caused by extreme weather conditions, such as the extremely dry years of 2000 to 2003, combined with the cumulative effects of pollution, mostly nitrogen emissions from transportation and industrial processes.

Our approach is ideal for modelling 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 spatio-temporal trend has intuitive appeal, making it straightforward to explain and interpret when communicating results to environmental policy makers. The 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 with decisions about lime application to some areas as a countermeasure to acidification of the soil. In addition the proposed model has been used in simulations to optimize the monitoring network, under the financial constraints which preclude a 4x4km grid resolution in future (Augustin, 2006).

As 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 set-up allows different degrees of smoothness relative to the different covariate axes, because the tensor product smooth allows the use of different penalty matrices for each dimension. Due to the different penalty matrices the resulting smooth is also scale invariant. In addition our approach gives the flexibility of using different bases for the different dimensions. Here we use different bases in space and time, while still having space-time scale invariance.

Simulation results show that the proposed model can separate the smooth temporal trend from temporal correlation at site level and also ARMA parameters can reliably be recovered. In 100 response vectors of size 400 with autocorrelated errors following an (1) ARMA process ($\phi = 0.9$ and $\rho = -0.7$) and (2) AR1 process ($\phi = 0.99$) with (a) a strong temporal trend made up from a polynomial of time and (b) with 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 square error between accounting and not accounting for the ARMA correlation structure in the

GAMM is negative. Also the medians of 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 as a covariate, a well known risk factor for defoliation, 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, required 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 $f_2()$ is mainly driven by site-specific characteristics such as geology, soil type, nutrient balance, local pollution levels and local weather conditions. All these covariates are correlated in space, and therefore it is not possible to distinguish between site specific effects and spatial effects modelled using eastings and northings. Similar findings on such confounding have been made in Augustin et al. (2007). Hence, with Model 2, the emphasis is on prediction rather than on establishing relationships between explanatory and response variables. In environmental monitoring identifiability of model parameters is difficult, and often it is not possible to separate the spatial effect from other effects which vary smoothly in space, especially if important covariates, such as pollution levels and weather information, are not included. The lack of covariate effects in the model is disappointing, but unsurprising, since the available covariates are mostly constant in time and the analysis has clearly shown that there are space-time effects.

The model takes temporal and spatial correlation via the space-time smoother into account

and in addition there is the possibility of accounting for unexplained correlation via a very general structure of the error distribution and the inclusion of random effects. In Model 2 we model temporal correlation of errors within location as an ARMA process, rather than also allowing for a spatial correlation structure. This makes sense since we are dealing with averages of defoliation of trees which typically hold needles for seven 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. Since 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 which showed that the temporal correlation is much stronger than the spatial correlation. Our diagnostic plots of residuals showed that the error structure is modelled adequately. This means that inference based on the model can be relied upon for making management decisions.

Acknowledgments

This work was supported by the Federal Ministry for Education and Research (Germany) project number 0339985. The first author was financially supported by a Royal Society UK Relocation Fellowship and by the Forest Research Centre Baden-Württemberg (Germany).

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, Department of Mathematical Sciences, University of Bath, BATH, BA2 7AYD.

- 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 C - Applied Statistics*, 56 (1), 29–50.
- Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. 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. 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 C (Applied Statistics)*, 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 C - Applied Statistics*, 57, 343–355.

- Hastie, T. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- 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 C (Applied Statistics)*, 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, B*, 61, 381–400.
- MacNab, Y. and Dean, C. (2001). Autoregressive spatial smoothing and temporal spline smoothing 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 characteristics. *European Journal of Forest Research*, 126(1), 91–100. Online issue: <http://dx.doi.org/10.1007/s10342-006-0132-8>.
- Pinheiro, J. and Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer-Verlag, New York.

- 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*. R Foundation for Statistical Computing, Vienna, Austria. 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 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, Series 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, Series B*, 47, 1–52.

- Wahba, G. (1983). Bayesian confidence intervals for the cross validated smoothing spline. *Journal of the Royal Statistical Society, Series 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.
- Wood, S. (2006a). *Generalized Additive Models. An Introduction with R*. Chapman & Hall/CRC, Boca Raton.
- Wood, S. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4), 1025–1036.
- Wood, S. (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.

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, green a medium and yellow a high percentage of defoliation, with defoliation levels given on each plot by the isolines. The dots indicate the sampling locations in each year.

Figure 2: The 1475 sampling locations of the TCCI survey with growth areas.

Figure 3: Semi-variogram of scaled (normalized) Pearson residuals from (a) the traditional method (Model 1) and (b) the Model 2.

Figure 4: Age effect on the logit scale (Model 2).

Figure 5: Top: Estimated average defoliation of spruce with 95% confidence bands 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. Bottom: Estimated average defoliation of spruce with 95% confidence bands based on predictions on regular grid standardized for three ages: 50, 75 and 95 (these are the lower, median and upper quartile age in the data).

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.