An Investigation of Humus Disintegration by Spatial-Temporal Regression Analysis

Roland H. FRIED
Department of Statistics, University of Dortmund, Germany
fried@statistik.uni-dortmund.de

We examine the hypothesis of an increase of humus disintegration by analyzing chemical substances measured in the seepage water of a German forest. Problems arise because of a large percentage of missing observations. We use a regression model with spatial and temporal effects constructed in an exploratory data analysis. Spatial dependencies are modelled by random effects and an autoregressive structure for observations in distinct soil depths resulting in a recursive linear mixed model structure. Temporal dependencies are included by an autoregressive structure of the random effects. For parameter estimation an EM algorithm is deduced assuming the errors to be Gaussian. As a result of the data analysis we specify chemical substances which possibly affect the process of humus disintegration. In particular, we find evidence that the presence of aluminium ions is important, but because of the high correlations among the regressors this might be due to confounding with iron.

Key Words: Autoregressive model; Maximum likelihood estimation; Missing data; Mixed effects; Recursive linear model; Spatial-temporal correlations.

1 INTRODUCTION

The hypothesis of humus disintegration developed by Ulrich (1981) states that in the recent decades the decomposition of humus is increasing in central European forests disturbing the natural balance of decomposition and reconstruction. According to this theory the presence of aluminium ions in the soil solution leads to an inhibition of the repolymerisation of the substances in the humus cycle, while the breakdown is not affected. This process results in an increase of dissolved organic carbon (DOC) in the soil solution and an excess of organic nitrogen. However, science is far from completely understanding the mechanisms of humus dynamics. The chemical composition of the seepage water in the forest soil is of basic interest (Eichhorn and Hüttermann 1999).

The case study 'Zierenberg' was developed as a part of the research program 'Stability conditions of forest ecosystems', operated by the University of Göttingen and the Hessian Agency of Forest Management Planning, Forest Research and Forest Ecology to inspect new kinds of forest damage such as humus disintegration in field experiments (Eichhorn

1995). In these experiments, the concentrations of DOC and of other chemical substances have been measured in the seepage water at twenty sites in a forest in central Germany. We use DOC as an indicator of the disintegration and regress it on possible influences.

Independence is not an appropriate assumption for the errors in a regression model when repeated measurements are taken at nearby sites since naive standard errors and tests based on least squares are incorrect then (Cressie 1993, Buonaccorsi and Elkinton 2002). In Fried (2001) a regression model is applied to the DOC-concentrations measured between 1989 and 1994 taking correlations between the observations into account. Because of many missing observations a parsimonious model structure was searched for in an exploratory data analysis. The spatial structure of the data is 3-dimensional as measurements are taken in three depths at sites located in an irregular grid. We neglect the lowest depth in the following since the previous analysis confirms existing knowledge that biochemical processes predominantly happen in the upper layers. The data were modelled as repeated measurements from a recursive linear mixed model (e.g. Wermuth 1980) with random effects describing almost uniform spatial correlations. While the hypothesis of humus disintegration relates DOC mainly to aluminium compounds, nitrate and the pH measured by the H_3O-concentration, the analysis in Fried (2001) points at further ions which possibly influence the DOC-concentration:

> Aluminium Al, calcium Ca, chlorid Cl, iron Fe, magnesium Mg, natrium Na, oxonium H_3O, potassium K, sulphate SO_4.

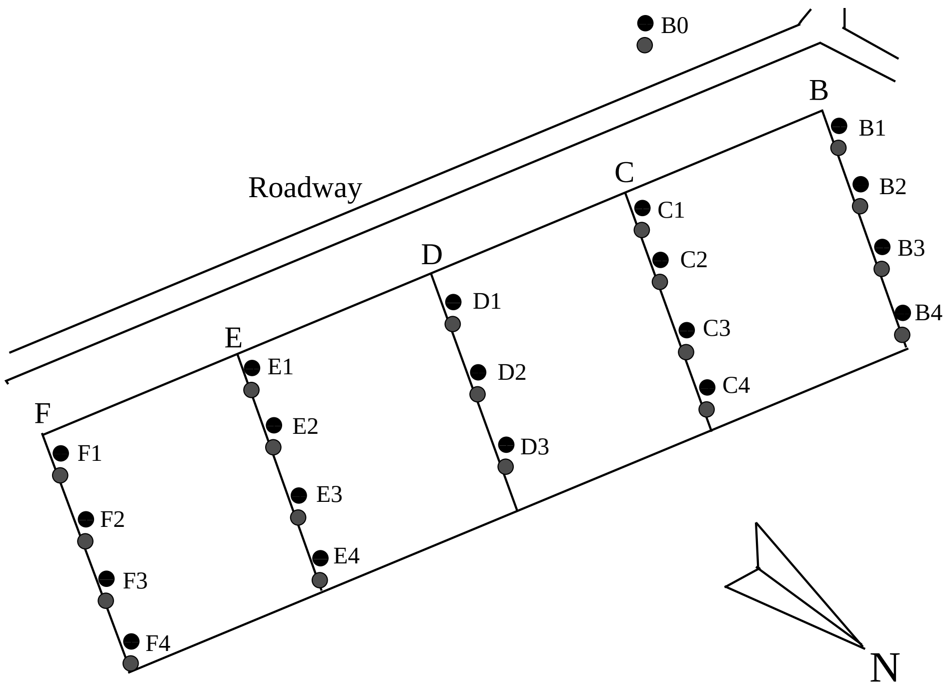
In this paper, we predict new data measured between 1995 and 1997 using the model developed in Fried (2001) in order to check its assumptions. After adding further sources of correlations we fit the extended model to the whole data set and reexamine the effects of the above ions on the DOC-concentrations using the BIC and the mean square prediction error. In Section 2 the data set is described and the model is explained in detail. In Section 3 an EM algorithm for calculation of maximum likelihood estimates is presented. This section can be skipped by a reader who is mainly interested in the results of the data analysis. In Section 4 the extended model is fitted and the important regressors selected. Section 5 discusses the results.

2 A SPATIAL-TEMPORAL REGRESSION MODEL

2.1 THE DATA

Since October 1989 the seepage water has been monitored in several soil depths at $n = 20$ sites in a forest near Zierenberg, a village in the region of Kassel, Germany. This location, which is situated on a hill, was chosen since the criteria for humus disintegration to occur are fulfilled (Eichhorn and Hüttermann 1999). The letters in Figure 1 denote the vertical lines of the gradient and the numbers count the sites along each line. Distances between adjacent lines are 40 meters. Thus, in the horizontal direction the sites are equidistant, whereas in the vertical direction the spacings vary between 5 and 11 meters. The slope in direction of the vertical lines is about 40%. We concentrate on the observations measured in the upper two soil layers in the following. The upper depth of 20 cm is located at the border of the first layer, while the lower depth of 60 cm is right in the center of the second layer. The vegetation in this region is homogeneous consisting mainly of 150-year-old beech trees. Unfortunately, no temperature data were available for this region during the period used for the analysis (October 1989 to December 1997).

Figure 1: Location Zierenberg. Measurements are taken at the 20 marked sites in two soil depths.



Measurements are taken by lysimeters, which collect the seepage water between the observation times. The time lag between subsequent rain fall events varies between a few days and some weeks. As soil humidity follows this pattern, the time delay between subsequent measurements varies between two and six weeks. Hence, the observations are irregular aggregations depending on the rainfall. In the analysis we use the shortest lag of two weeks for the time scale, i.e. we have $T = 196$ observation times. Further problems are caused by the variability of precipitation at Zierenberg. Since even rainy days did not cause seepage water at all sites, there are many values missing individually for each site and depth. As most biochemical processes predominantly happen in the upper layer, the lower depth was only measured when the upper one was observable. Thus, there are more observations in the upper (1480) than in the lower layer (952). If there had been no missing data, 4300 observations in lags of two weeks would have been available for each depth, i.e. about 65.6% and 77.9% of the observations in the upper and the lower layer, respectively, are missing. In the second period between 1995 and 1997 there are more observations in the shortest time lag than in the first period.

2.2 THE MODEL

In Fried (2001) a regression model for DOC was developed including possibly important covariates as well as systematic spatial and temporal effects as regressors. The logarithm of the DOC-concentration was used for stabilizing the variance. We denote the logarithm of DOC at time $t \in \{1, \ldots, T\}$, depth $u \in \{1, 2\}$, and site $s \in S = \{B0, B1, \ldots, F4\}$ by $Y_{t,u,s}$. Similarly, let $x_{t,u,s,j}$ be the measurement of the j-th covariate, $j = 1, \ldots, 9$, and $\beta_{u,j}, j = 1, \ldots, 9$, its regression coefficient. The soil consistency, the environmental variables and the effects of the regressors were assumed to be homogeneous within each layer. Accordingly, the regression coefficients were considered to be identical for all sites in the same layer but possibly different for distinct layers. For modelling an obvious seasonal pattern, a general sinus function $\sin(2\pi t/26)\beta_{u,10} + \cos(2\pi t/26)\beta_{u,11}$ with a period of one year (26 observation times per annum) was included. Significant spatial differences were found after regressing on the explanatory variables and removing seasonal effects. As no simple pattern was found, dummy variables indicating the site were added to the model resulting in a total of $p = 30$ regressors.

For contemporaneous observations almost uniform positive correlations between as well as within the layers were found. Correlations between the depths were modelled by adding an autoregressive term $\alpha Y_{t,1,s}$ to $Y_{t,2,s}$. This term can be interpreted as an unknown fraction of DOC which trickles down in the soil. The almost uniform positive spatial correlations found in the exploratory analysis were assumed to be caused by weather conditions such as rainfall and temperature, which affect all contemporaneous observations in the same layer similarly. Therefore, a random effect $w_{t,u} \sim N(0, \tau_u^2)$ was added to all observations at time t in depth u. Adding a further AR(1)-structure to the observations along the vertical direction of the gradient for modelling stronger correlations between nearby sites did not improve the model fit substantially. Therefore, we neglect this additional structure in the following. Temporal correlations were found to be close to zero for the time lag $\delta = 2$ corresponding to 4 weeks. Since in the first period there were only a few observations for the time lag $\delta = 1$ temporal correlations were assumed to be negligible. The resulting model is

$$\begin{aligned}
Y_{t,u,s} &= \mu_{t,u,s} + w_{t,u} + \varepsilon_{t,u,s} \qquad (1)\\
\text{with} \quad \mu_{t,1,s} &= \sum_{j=1}^{p} x_{t,1,s,j}\beta_{1,j}\\
\mu_{t,2,s} &= \sum_{j=1}^{p} x_{t,2,s,j}\beta_{2,j} + \alpha Y_{t,1,s}\\
\varepsilon_{t,u,s} &\sim N(0, \tilde{\sigma}_u^2)\\
w_{t,u} &\sim N(0, \tau_u^2)
\end{aligned}$$

with all $\varepsilon_{t,u,s}$ and $w_{t,u}$ being independent. Estimates are marked by hats in the following.

Let $\underline{Y}_{t,u}$ be the vector of all observations $Y_{t,u,s}$ at time point t in layer u. Because of missing data, only a subvector $\underline{Z}_{t,u} = \underline{\mathrm{M}}_{t,u}\underline{Y}_{t,u}$ was observed, with $\underline{\mathrm{M}}_{t,u}$ being an $(n_{t,u} \times n)$-matrix having 0/1 entries and $n_{t,u}$ being the number of observations measured at time point t in layer u. We assume that missing data are missing at random and are not informative for the analysis. The values of the explanatory variables corresponding to $\underline{Z}_{t,u}$ are given by the $(n_{t,u} \times p)$-matrix $\underline{\mathrm{X}}_{t,u}$, which also contains the variables describing systematic temporal and spatial effects. Adding as last column those observations of $\underline{Z}_{t,1}$, for which the lower layer has also been measured, to $\underline{\mathrm{X}}_{t,2}$ we get the $(n_{t,2} \times p + 1)$-matrix $\underline{\tilde{X}}_{t,2}$. For simplification of the following formula we set $\underline{\tilde{X}}_{t,1} = \underline{\mathrm{X}}_{t,1}$.

In Fried (1999) and Fried (2001), this basic model was checked using the data in the first time period. The model diagnostics provided evidence that the model captures the variation of the *DOC*-concentration rather well. No problems were noted for the assumptions of normality, of the linearity of the covariates effects, of a constant variance, as well as for the assumptions of spatial stationarity and isotropy. However, scatterplots of residual pairs for fixed short time lags and calculation of empirical space-time autocorrelations raised the question of possible short term autocorrelations. Moreover, positive contemporaneous correlations were found between the residuals in distinct layers without an obvious spatial structure. These findings point at another source of common variation like a random effect $w_{t,0}$ influencing all contemporaneous observations in both layers, and possibly there are temporal correlations at small time lags. In view of the many missing data we model such autocorrelations by adding a simple time series structure to the random effects for getting tractable estimates.

We choose an AR(1)-model for the random effects $\underline{w}_t = (w_{t,0}, w_{t,1}, w_{t,2})'$ with a diagonal coefficient matrix $\underline{G} = diag(\gamma_0, \gamma_1, \gamma_2)$, $\gamma_u \in [0, 1)$, $u = 0, 1, 2$, and a diagonal covariance matrix $\underline{\Delta} = diag(\tilde{\tau}_0^2, \tilde{\tau}_1^2, \tilde{\tau}_2^2)$ for the innovations $\underline{\delta}_t$. Further, let

$$\underline{\mathrm{H}}_{t,1} = \underline{\mathrm{M}}_{t,1} \times \begin{bmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \end{bmatrix}, \qquad \underline{\mathrm{H}}_{t,2} = \underline{\mathrm{M}}_{t,2} \times \begin{bmatrix} 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix},$$

$\underline{\varepsilon}_{t,u}$ be the vector of all errors $\varepsilon_{t,u,s}$, $s \in S$, $\underline{\beta}_u = (\beta_{u,1}, \ldots, \beta_{u,p})'$, $\underline{\tilde{\beta}}_2 = (\underline{\beta}_2', \alpha)'$ and $\underline{\tilde{\beta}}_1 = \underline{\beta}_1$ for the ease of notation. With these settings, an extended model can be written as follows:

$$\begin{aligned} \underline{Z}_{t,u} &= \underline{\tilde{X}}_{t,u}\underline{\tilde{\beta}}_u + \underline{\mathrm{H}}_{t,u}\underline{w}_t + \underline{\mathrm{M}}_{t,u}\underline{\varepsilon}_{t,u} \\ \underline{\varepsilon}_{t,u} &\sim N\left(\underline{0}, \tilde{\sigma}_u^2\underline{\tilde{\Sigma}}_u\right) \\ \underline{w}_t &= \underline{G}\underline{w}_{t-1} + \underline{\delta}_t \\ \underline{\delta}_t &\sim N(\underline{0}, \underline{\Delta}) \\ \underline{w}_0 &\sim N(\underline{\mu}, \underline{\Omega}) \\ \text{with} \quad & \underline{w}_0,\, \underline{\delta}_1,\, \underline{\varepsilon}_{1,1},\, \underline{\varepsilon}_{1,2},\, \underline{\delta}_2,\, \underline{\varepsilon}_{2,1},\, \underline{\varepsilon}_{2,2} \ldots \text{ being independent.} \end{aligned} \tag{2}$$

We set $\underline{\tilde{\Sigma}}_u = \underline{\mathrm{I}}$, the identity matrix. Otherwise, $\underline{\tilde{\Sigma}}_u$ could capture further spatial correlations. Initializing the process of the random effects by $\underline{\Omega} = diag(\tau_0^2, \tau_1^2, \tau_2^2)$, $\tau_u^2 = \dfrac{\tilde{\tau}_u^2}{1-\gamma_u^2}$, and $\underline{\mu} = \underline{0}$, the process of the random effects is stationary with $E(\underline{w}_t) = \underline{0}$ and $Cov(\underline{w}_t) = \underline{\Omega}$. Choosing an autoregressive model with order higher than one would be possible without changing the following approach to maximum likelihood estimation for the expense of higher matrix dimensions (Harrison and Stevens 1976, Jones 1980).

3 MAXIMUM LIKELIHOOD ESTIMATION

The likelihood function of the extended model formulated in the previous section is difficult to treat analytically because of the missing data. Let $\underline{\theta} \in \Theta \subseteq I\!R^q$ be the vector of all unknown model parameters. We set $\underline{Z}_t(\alpha) = (\underline{Z}_{t,1}', \underline{Z}_{t,2}' - \alpha\underline{Y}_{t,1}'\underline{\mathrm{M}}_{t,2}')'$ and

$$\underline{V}_t = \begin{bmatrix} \underline{V}_{t,1} & \tau_0^2\underline{J} \\ \tau_0^2\underline{J} & \underline{V}_{t,2} \end{bmatrix},$$

where $\underline{V}_{t,u} = \underline{\mathrm{M}}_{t,u}[(\tau_u^2+\tau_0^2)\underline{J}+\tilde{\sigma}_u^2\underline{\tilde{\Sigma}}_u]\underline{\mathrm{M}}_{t,u}'$ and $\underline{J}$ is an $(n\times n)$-matrix with all entries equal to 1. For a given set of parameters, knowing $\underline{Z}_t(\alpha)$ is equivalent to knowing $\underline{Z}_t = (\underline{Z}_{t,1}', \underline{Z}_{t,2}')'$. Further, let $\underline{\mathrm{X}}_t = (\underline{\mathrm{X}}_{t,1}', \underline{\mathrm{X}}_{t,2}')'$, $\underline{\beta} = (\underline{\beta}_1', \underline{\beta}_2')'$, N_u be the total number of observations in layer u and $N = N_1 + N_2$. We denote the determinant of a square matrix $\underline{\mathrm{A}}$ by $|\underline{\mathrm{A}}|$ and its trace by $tr(\underline{\mathrm{A}})$ in the following. Then twice the negative log-likelihood L for the observations $\underline{Z} = (\underline{Z}_1, \ldots, \underline{Z}_T)$ can be written in innovations form as

$$L = N\ln(2\pi) + \sum_{t=1}^{T} \ln|\underline{V}_{t|t-1}| + \sum_{t=1}^{T} \left(\underline{Z}_t(\alpha) - \underline{z}_{t|t-1}(\alpha)\right)' \underline{V}_{t|t-1}^{-1} \left(\underline{Z}_t(\alpha) - \underline{z}_{t|t-1}(\alpha)\right)$$

with

$$\begin{aligned} \underline{z}_{t|t-1}(\alpha) &= E\left(\underline{Z}_t(\alpha)|\underline{Z}_1 = \underline{z}_1, \ldots, \underline{Z}_{t-1} = \underline{z}_{t-1}\right) \\ \underline{V}_{t|t-1} &= Cov\left(\underline{Z}_t(\alpha) \mid \underline{Z}_1 = \underline{z}_1, \ldots, \underline{Z}_{t-1} = \underline{z}_{t-1}\right) . \end{aligned}$$

Unknown parameters in the covariance matrices have nonlinear effects on the conditional moments and hence on the likelihood, and missing observations further complicate the calculation of maximum likelihood estimates. For a given set of parameters, the conditional moments can be calculated by Kalman filtering as described in the next subsection.

We follow a suggestion by Shumway and Stoffer (1982) to use an EM algorithm for calculation of maximum likelihood estimates in view of the missing data. We treat the random effects $\underline{w} = (\underline{w}_0, \underline{w}_1, \ldots, \underline{w}_T)$ as missing values in the following. Twice the negative log-likelihood $\tilde{L}$ for the 'complete' data $\tilde{\underline{z}} = (\underline{w}_0, \underline{w}_1, \underline{z}_1, \ldots, \underline{w}_T, \underline{z}_T)$ has a simple form, see the Appendix. In the EM algorithm we condition $\tilde{L}$ on the observed data calculating the parameter values which maximize $Q_i(\underline{\theta}) = E_{\underline{\theta}^{(i-1)}}(\tilde{L}(\underline{\theta})|\underline{Z} = \underline{z})$ in the i-th step. Here, the subscript $\underline{\theta}^{(i-1)}$ indicates that we compute the conditional expectation assuming the parameter set obtained in step $(i-1)$ to be true. Starting with an arbitrary initial set of parameter values $\underline{\theta}^{(0)}$, Dempster, Laird and Rubin (1977) show that iterating this procedure yields non-decreasing likelihoods, with the fixed points being stationary points of the likelihood function (see also Wu 1983). In each iteration we need to perform an expectation E-step and a maximization M-step to get a new set of estimates $\underline{\theta}^{(i)}$, $i = 1, 2, \ldots$, until this sequence converges.
In the E-step we evaluate $Q_i(\underline{\theta}) = E_{\underline{\theta}^{(i-1)}}(\tilde{L}(\underline{\theta})|\underline{Z} = \underline{z})$ applying the Kalman filter to calculate the necessary conditional moments. In the M-step then a new set of parameter estimates $\underline{\theta}^{(i)}$ is calculated by maximization of $Q_i(\underline{\theta})$.

3.1 KALMAN FILTERING

It is shown in the Appendix that in the E-step of the EM algorithm we just need to evaluate conditional expectations and covariance matrices conditioning on the observations $\underline{z}$ and fixing the parameters to the estimates calculated in the previous iteration. We set $\underline{\mathrm{H}}_t = (\underline{\mathrm{H}}'_{t,1}, \underline{\mathrm{H}}'_{t,2})'$, $\underline{S}_{t,u} = \underline{\mathrm{M}}_{t,u}\tilde{\underline{\Sigma}}_u\underline{\mathrm{M}}'_{t,u}$, and $\tilde{\sigma}_1^2\underline{S}_t = diag(\tilde{\sigma}_1^2\underline{S}_{t,1}, \tilde{\sigma}_2^2\underline{S}_{t,2})$. Let $\underline{w}_{t|\tilde{t}}$ be the conditional expectation of $\underline{w}_t$, $\underline{\Omega}_{t|\tilde{t}}$ and $\underline{V}_{t|\tilde{t}}$ be the conditional covariance of $\underline{w}_t$ and $\underline{z}_t(\alpha)$, respectively, and $\underline{\Omega}_{t,t-1|\tilde{t}}$ be the conditional covariance of $\underline{w}_t$ and $\underline{w}_{t-1}$ given all observed data up to time point $\tilde{t}$, $\underline{z}_1, \ldots, \underline{z}_{\tilde{t}}$. These conditional moments can be calculated using Kalman filtering (Kalman 1960). For the reason of simplicity we drop the suffices of the estimates and call them simply α, $\underline{\beta}$ and $\tilde{\sigma}_1^2$. The iterations read as follows.
Initialization: Set $\underline{w}_{0|0} = \underline{\mu}$, $\underline{\Omega}_{0|0} = \underline{\Omega}$.
Forward recursions: For $t = 1, \ldots, T$ iterate the following steps.

Kalman predictor:

$$\begin{aligned}
\underline{w}_{t|t-1} &= \underline{G}\underline{w}_{t-1|t-1} \\
\underline{\Omega}_{t|t-1} &= \underline{G}\underline{\Omega}_{t-1|t-1}\underline{G}' + \underline{\Delta} \\
\underline{z}_{t|t-1}(\alpha) &= \underline{\mathrm{H}}_t\underline{w}_{t|t-1} + \underline{\mathrm{X}}_t\underline{\beta} \\
\underline{V}_{t|t-1} &= \underline{\mathrm{H}}_t\underline{\Omega}_{t|t-1}\underline{\mathrm{H}}_t' + \tilde{\sigma}_1^2\underline{S}_t
\end{aligned}$$

Kalman filter:

$$\begin{aligned}
\underline{K}_t &= \underline{\Omega}_{t|t-1}\underline{\mathrm{H}}_t'\underline{V}_{t|t-1}^{-1} \\
\underline{w}_{t|t} &= \underline{w}_{t|t-1} + \underline{K}_t(\underline{z}_t(\alpha) - \underline{z}_{t|t-1}(\alpha)) \\
\underline{\Omega}_{t|t} &= \underline{\Omega}_{t|t-1} - \underline{K}_t\underline{\mathrm{H}}_t\underline{\Omega}_{t|t-1}
\end{aligned}$$

Backward recursions: Iterate for $t = T, T-1, \dots, 1$ the following steps.
Kalman Smoother:

$$\begin{aligned}
\underline{\mathrm{L}}_{t-1} &= \underline{\Omega}_{t-1|t-1}\underline{G}'\underline{\Omega}_{t|t-1}^{-1} \\
\underline{w}_{t-1|T} &= \underline{w}_{t-1|t-1} + \underline{\mathrm{L}}_{t-1}(\underline{w}_{t|T} - \underline{G}\underline{w}_{t-1|t-1}) \\
\underline{\Omega}_{t-1|T} &= \underline{\Omega}_{t-1|t-1} + \underline{\mathrm{L}}_{t-1}(\underline{\Omega}_{t|T} - \underline{\Omega}_{t|t-1})\underline{\mathrm{L}}_{t-1}' \\
\underline{\Omega}_{t,t-1|T} &= (\underline{\mathrm{I}} - \underline{K}_T\underline{\mathrm{H}}_T)\underline{G}\underline{\Omega}_{T-1|T-1} \qquad \text{for } t = T \\
\underline{\Omega}_{t,t-1|T} &= \underline{\Omega}_{t|t}\underline{\mathrm{L}}_{t-1}' + \underline{\mathrm{L}}_t(\underline{\Omega}_{t+1,t|T} - \underline{G}\underline{\Omega}_{t|t})\underline{\mathrm{L}}_{t-1}' \qquad \text{for } t < T
\end{aligned}$$

The Kalman predictor provides forecasts $\underline{z}_{t+1|t}(\alpha)$ for the next time point. We get predictions $\underline{z}_{t+h|t}(\alpha)$ for a longer time span $h = 2, 3, \dots$ iteratively from

$$\begin{aligned}
\underline{w}_{t+h|t} &= \underline{G}\underline{w}_{t+h-1|t}\,, \\
\underline{z}_{t+h|t}(\alpha) &= \underline{\mathrm{H}}_{t+h}\underline{w}_{t+h|t} + \underline{\mathrm{X}}_{t+h}\underline{\beta}\,,
\end{aligned}$$

see West and Harrison (1989, p. 600f).

3.2 THE MAXIMIZATION STEP

For the maximization of $Q_i(\underline{\theta}) = E_{\underline{\theta}^{(i-1)}}(\tilde{L}(\underline{\theta})|\underline{Z} = \underline{z})$ in the i-th M-step of the EM algorithm the following formula can be used (see the Appendix).

$$\begin{aligned}
\hat{\tilde{\underline{\beta}}}_u &= \sum_{t=1}^{T} \underline{\mathrm{A}}_{t,u}(\underline{z}_{t,u} - \underline{\mathrm{H}}_{t,u}\underline{w}_{t|T}) \\
\hat{\tilde{\sigma}}_u^2 &= \frac{1}{N_u}\left[\sum_{t=1}^{T} \left(\underline{z}_{t,u} - \underline{\mathrm{H}}_{t,u}\underline{w}_{t|T}\right)' \underline{S}_{t,u}^{-1} \left(\underline{z}_{t,u} - \underline{\mathrm{H}}_{t,u}\underline{w}_{t|T} - \underline{\tilde{X}}_{t,u}\hat{\tilde{\underline{\beta}}}_u\right) + tr\left(\underline{S}_{t,u}^{-1}\underline{\mathrm{H}}_{t,u}\underline{\Omega}_{t|T}\underline{\mathrm{H}}_{t,u}'\right)\right],
\end{aligned}$$

$$\hat{\tilde{\tau}}_u^2 = \frac{1}{T+1}(a_u - 2b_u\gamma_u + c_u\gamma_u^2)\,,$$

with

$$\underline{\mathrm{A}}_{t,u} = \left(\sum_{\tilde{t}=1}^{T} \underline{\tilde{X}}'_{\tilde{t},u}\underline{S}^{-1}_{\tilde{t},u}\underline{\tilde{X}}_{\tilde{t},u}\right)^{-1} \underline{\tilde{X}}'_{t,u}\underline{S}^{-1}_{t,u}$$

$$a_u = \sum_{t=1}^{T}(\underline{\Omega}_{t|T} + \underline{w}_{t|T}\underline{w}'_{t|T})_{(u,u)},$$

$$b_u = \sum_{t=1}^{T}(\underline{\Omega}_{t,t-1|T} + \underline{w}_{t|T}\underline{w}'_{t-1|T})_{(u,u)},$$

$$c_u = \sum_{t=0}^{T-1}(\underline{\Omega}_{t|T} + \underline{w}_{t|T}\underline{w}'_{t|T})_{(u,u)},$$

where $\underline{\mathrm{M}}_{(i,j)}$ denotes the entry of a matrix $\underline{\mathrm{M}}$ at position (i,j). It remains to minimize $-\ln(1-\gamma_u^2) + (T+1)\ln(\hat{\tilde{\tau}}_u^2(\gamma_u))$ separately w.r.t. γ_u, see the Appendix.

3.3 STOPPING THE ALGORITHM

For stopping the EM algorithm we used a stopping criterion based on the relative and the absolute changes of γ_u, $\tilde{\tau}_u^2$, $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$, α and L. We denote the values obtained in the $i-$th step by the suffix (i). Following suggestions by Heinz and Spellucci (1994) we stopped the recursions if one of the following conditions is fulfilled in ten subsequent iterations:

$$(i) \qquad \left.\begin{array}{rcl} |\vartheta^{(i+1)} - \vartheta^{(i)}| & \leq & h_1 \quad \text{and} \\ |\vartheta^{(i+1)} - \vartheta^{(i)}| & \leq & h_3(|\vartheta^{(i)}| + h_2) \end{array}\right\} \forall\vartheta \in \{\gamma_1, \gamma_2, \gamma_3, \tilde{\tau}_1^2, \tilde{\tau}_2^2, \tilde{\tau}_3^2, \sigma_1^2, \sigma_2^2, \alpha\}$$

$$(ii) \qquad |L^{(i+1)} - L^{(i)}| \leq h_0$$

Here, we chose $h_0 = h_1 = h_2 = 50\sqrt{MP}$ and $h_3 = 2000\sqrt{MP}$, where MP denotes the machine precision, which for the HP workstation used in the calculations is about 10^{-16}. If condition (i) is fulfilled the changes of the estimates are so small that in view of possible rounding errors we do not expect noteworthy parameter changes any more. If condition (ii) is fulfilled we no longer expect the model fit to improve substantially.

In order to overcome possible problems arising because of local optima we chose six distinct sets of parameters for starting the algorithm. Stopping the algorithm took us about 2000 iterations in each case. The final solutions could be considered to be identical, i.e. we did not note problems with local optima.

4 APPLICATION TO THE *DOC* DATA

In the following we apply the extended model (2) to the data. First we check this model using new data, and then we select the important covariates thereafter.

4.1 MODEL VALIDATION

First we check the extended model using new data observed between 1995 and 1997, which had not been used for model identification by Fried (2001). We estimate the model parameters from the data observed up to 1994 applying the EM algorithm as described in the previous section. The resulting estimates are used to predict the data observed in the second period and to analyze the residuals of the predictions.

Figure 2 depicts histograms of these predicted residuals (PR). Keeping in mind that predicted residuals are typically more variable than ordinary residuals as the corresponding observations are not used for model fitting, a normal distribution suits the residuals in the upper layer rather well. The fit in the lower layer is worse, but this can be explained by the smaller number of observations and higher correlations, see Section 4.2. Moreover, the estimation of α adds further uncertainty. We notice a few outliers in both layers.

Figure 2: Histograms of the predicted residuals for the upper (left) and the lower layer (right)

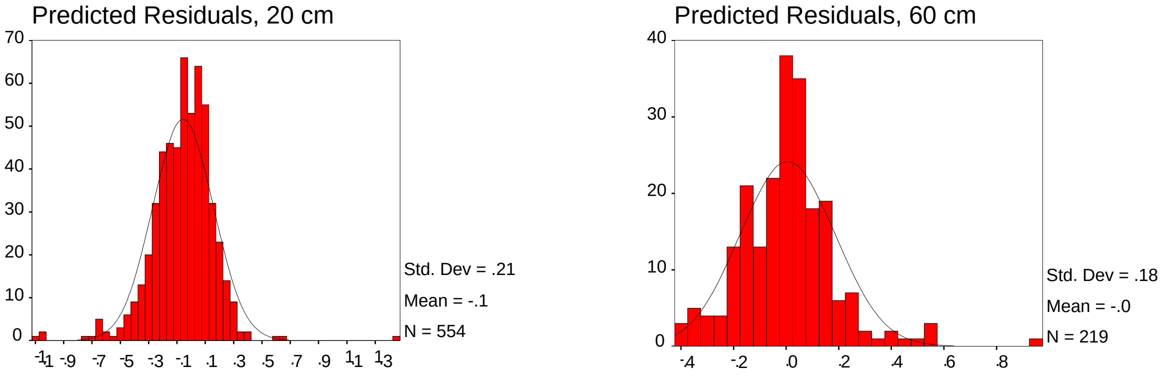


Figure 3 depicts plots of the predicted residuals against the predictions and against the particularly interesting covariates Fe, Al, and NO_3. The latter plots confirm the findings in Fried (2001) that there does not seem to be a relationship between DOC and NO_3 which cannot be explained by other effects. There might be some deviations from linearity as there are more negative than positive residuals for small values of Fe and Al. We applied other transformations like different roots for overcoming these difficulties, but we did not find them to give better results. Alternatively, non-parametric methods like splines could be applied for obtaining a curvilinear fit, or a model with time-varying parameters could be tried out. However, these methods are difficult to apply if there are many regressors. Reducing the dimension of the regressor space first by selecting important covariates by linear regression as is done here is helpful then. We keep in mind that there are some slight deviations from linearity and hence should act with caution when interpreting the following results. Furthermore, we find a few outliers. As the covariates have moderate values for most of these outliers, they do not provide evidence against the model. The outliers are eliminated from the further analysis since we regard them as artifacts caused by effects not considered here.

Boxplots of the residuals against the time points and the sites, respectively, not shown here do not point at a serious lack of model fit w.r.t. the systematic temporal and spatial effects and the assumption of a constant variance.

We report correlations for the vertical direction since dependencies are most likely to be expected along this direction of the gradient. Table 1 displays empirical space-time-autocorrelations

$$\hat{\rho}_{(u,i),(v,j)}(\delta) = \frac{1}{N_{(u,i),(v,j)}(\delta)} \sum_{L \in \{B,C,D,E,F\}} \sum_{t=1}^{T} \left(z_{t,u,Li} - \hat{\mu}_{t,u,Li}\right)\left(z_{t+\delta,v,Lj} - \hat{\mu}_{t+\delta,v,Lj}\right)$$

in the vertical direction between pairs of residuals located at sites $i, j \in \{1, 2, 3, 4\}$ in depth u and v, respectively, along the same line $L \in \{B, C, D, E, F\}$ in a lag of δ time points. Here, correlations are averaged over the horizontal lines for reducing the variability, and $N_{(u,i),(v,j)}(\delta)$ is the number of pairs of observations from which the estimate is calculated. We find the contemporaneous correlations to be mostly positive with spatial distances not being important. In the second period there are a couple of observations lagged by one time period (two weeks) confirming positive temporal correlations. Again, we do not find a spatial structure of these correlations. Since the correlations in the horizontal direction are close to those reported here, we consider the correlation structure used in the extended model to be a viable approximation of the real world phenomena.

Summarizing these impressions we may state that the extra parameters in the extended model are helpful to overcome some deficiencies of the model used in Fried (2001).

4.2 MODEL FITTING

Next we fit the extended model (2) to the whole data set, see Table 2. The estimates for the simpler model (1) and the corresponding t-statistics $t_{u,i}$ are also included for the reason of comparison, as well as estimates of $\sigma_u^2 = \tilde{\sigma}_u^2 + \tau_u^2 + \tau_0^2$, which is the (conditional) variance, and $\rho_u = (\tau_u^2 + \tau_0^2)/\sigma_u^2$, which is the uniform correlation in layer u (conditional on the upper layer for $u = 2$), obtained by inserting the estimates of the individual parameters.

The estimated regression coefficients are similar in both models and show the same tendencies as in Fried (2001). Only the estimate for Mg changes its sign and becomes positive. The autoregressive parameter α looses its importance when introducing a common random effect influencing both layers, just like the random effect $w_{t,1}$ for the upper

Figure 3: Scatterplots of the predicted residuals against the predictions and against some covariates.

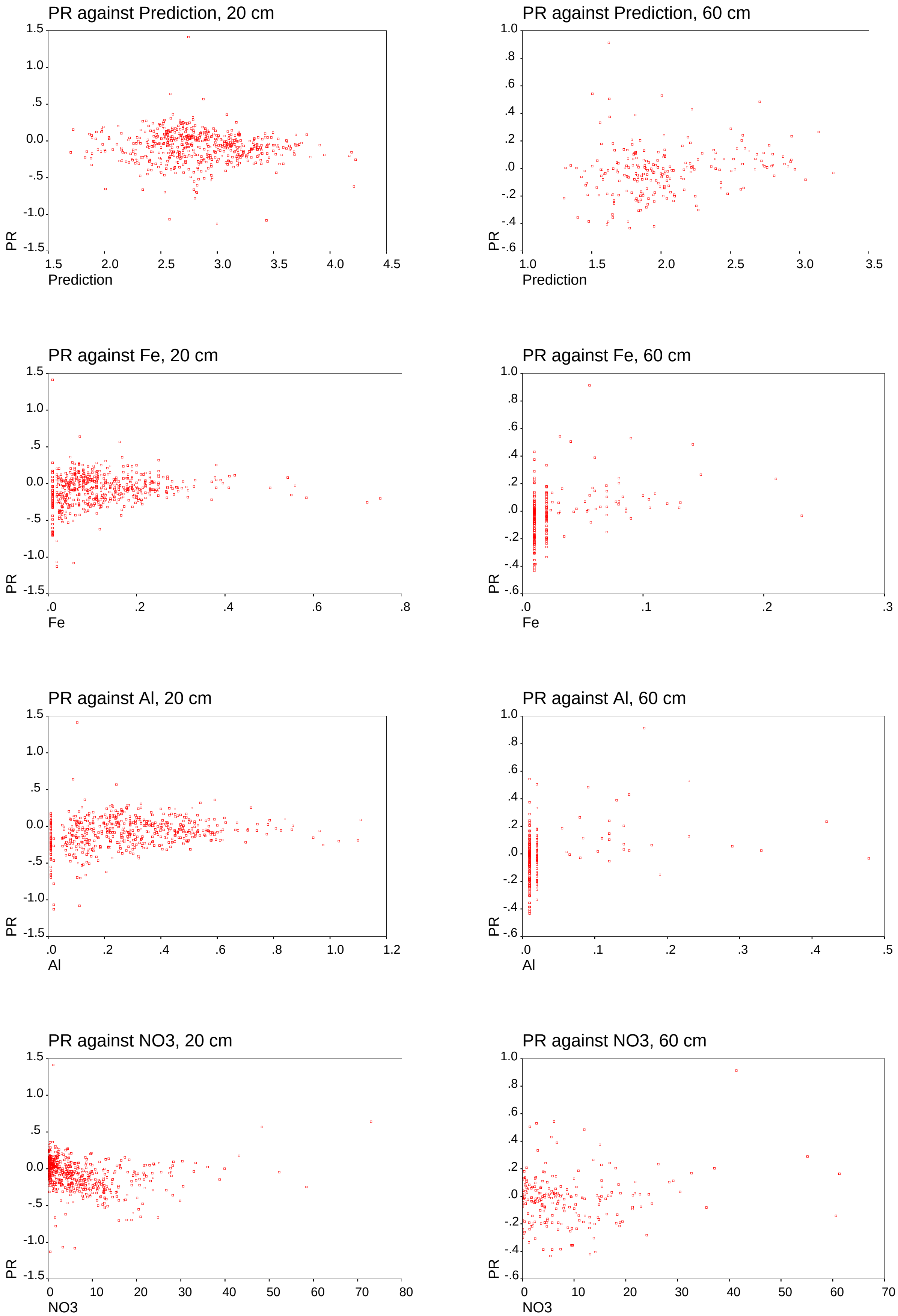


Table 1: Empirical space-time correlations $\hat{\rho}_{(1,i),(1,j)}(\delta)$ for the upper layer (top) and $\hat{\rho}_{(1,i),(2,j)}(\delta)$ between the layers (bottom) in the vertical direction between the horizontal lines i and j. N is the number of pairs of observations.

δ		0				1				2			
i	j	1	2	3	4	1	2	3	4	1	2	3	4
1	$\hat{\rho}$	1.00	.160	.236	.262	.083	.425	.161	-.020	.270	.158	.039	.088
	N	162	92	99	73	76	101	82	53	93	69	77	52
2	$\hat{\rho}$		1.00	.596	.296	.850	.247	.522	.024	.195	.719	.513	.219
	N		132	93	64	71	65	73	43	67	63	74	41
3	$\hat{\rho}$			1.00	.327	.561	.062	.670	.316	.000	.453	.647	.137
	N			144	66	71	75	89	45	74	65	79	43
4	$\hat{\rho}$				1.00	-.020	.163	.146	.195	.099	-.035	.208	.034
	N				99	47	51	52	44	51	45	51	45
1	$\hat{\rho}$	-.079	.093	.154	.148	.005	-.086	.229	.334	.152	.191	.168	.040
	N	55	37	39	31	30	42	36	26	29	21	25	15
2	$\hat{\rho}$	-.043	.163	.216	.065	.330	.420	.261	-.044	.126	.128	.322	.254
	N	40	58	41	32	33	34	38	25	26	25	29	17
3	$\hat{\rho}$	-.087	.145	.331	.228	.296	.010	.462	.278	-.120	.240	.375	-.276
	N	42	44	63	29	30	38	44	22	27	25	27	15
4	$\hat{\rho}$	.307	.099	.093	-.341	.181	.030	-.002	.283	.038	-.282	-.023	-.465
	N	31	34	27	41	20	19	26	21	19	17	17	14

layer, which is almost completely absorbed into the common effect $w_{t,0}$. The estimated variances of the measurement errors $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ in the upper and the lower layer, respectively, are very close to each other in the extended model. The estimated total variance σ_2^2 in the lower layer increases considerably when adding the common random effect.

The primary aim here is to find those covariates which possibly influence the *DOC*-concentration. The t-statistics in Table 2 are strongly significant in both layers except for oxonium H and potassium K mirroring a well-known problem of statistical significance testing in complex situations. A statistical model is only a rough summary description of real world phenomena and all model assumptions are approximations only. Hodges and Lehmann (1954) noted that for a large data set this often results in many significant parameters (overfitting). In the following we apply other criteria for variable selection.

4.3 VARIABLE SELECTION

There are some alternatives to significance testing for variable selection in statistical models. Often information criteria like AIC or BIC are applied. We prefer Schwarz's (1978) BIC since it results in more parsimonious models avoiding overfitting and since it is asymptotically consistent. In case of subsequently sampled data the out-of-sample

Table 2: Model parameters estimated from the whole data set. N_u is the number of observations in layer $u = 1, 2$.

Model		Basic				Extended	
u		1 (20 cm)		2 (60 cm)		1 (20 cm)	2 (60 cm)
N_u		1476		951		1476	951
α		.0246				.0001	
ρ_u		.30030		.39372		.34432	.50153
σ_u^2		.02917		.03575		.02827	.04102
$\tilde{\sigma}_u^2$		.01951		.02181		.01929	.02045
γ_0	$\tilde{\tau}_0^2$	0		0		.34702	.00789
γ_1	γ_2	0		0		7×10^{-7}	.01535
$\tilde{\tau}_1^2$	$\tilde{\tau}_2^2$	.00966		.01394		$.7 \times 10^{-7}$	.01160
i	Var	$\beta_{1,i}$	$t_{1,i}$	$\beta_{2,i}$	$t_{2,i}$	$\beta_{1,i}$	$\beta_{2,i}$
1	SO_4	.0170	7.41	.0242	7.93	.0173	.0225
2	Cl	-.0076	-4.05	-.0076	-3.05	-.0081	-.0067
3	H	-1.3388	-0.25	-83.036	-1.88	-2.3321	-57.422
4	Mg	.0215	5.53	-.0150	-3.79	.0218	-.0147
5	Ca	-.0258	-9.17	-.0192	-7.70	-.0259	-.0187
6	K	.0281	9.66	.0165	1.21	.0273	.0159
7	Na	-.0631	-6.37	.0652	6.04	-.0612	.0644
8	Fe	.8893	10.51	1.4558	3.95	.8758	1.4614
9	Al	.4910	11.54	.4336	2.76	.5055	.4000
10	sin	-.1861	-12.60	-.0924	-4.08	-.2067	-.1434
11	cos	-.0630	-4.31	-.0153	-0.72	-.0763	-.0197

mean square prediction error (MPE) is also frequently applied. We use the weighted average of the MPE's for the upper and the lower layer with weights according to the numbers of observations.

The importance of the seasonal and the spatial effects has already been stressed. These variables are not the main focus here. We only consider the nine regressors representing chemical substances fixing the spatial and the temporal effects in the model. This results in $2^9 = 512$ possible subsets rendering a whole subset selection feasible. Both criteria lead to similar results. Large models with six to nine covariates are preferred to smaller ones, with the MPE resulting in slightly more parsimonious models than the BIC. We find the best models for a fixed number of variables to be identical in most of the cases with some realignments. Cl, H and Mg are the least necessary variables. When choosing a moderately large model with three to five covariates, Na, SO_4, and K are also often excluded, while Ca and Al are almost always included. The best models with two covariates are combinations of either Al or Fe with one of Ca and Mg, with the combination Al, Ca being best. This interchangeability agrees with the results of a correlation analysis reported in Fried, Eichhorn and Paar (2001). If only one variable is chosen, either Fe or Al is selected with minor differences. The model with Fe has a smaller MPE because of the better predictions in the lower layer. The model with all covariates being excluded is much worse than the others. Although it includes 19 spatial and two temporal regressors, its MPE is twice as large as the MPE of the models including either Fe or Al, and it is three times the MPE of the best model.

The choice of the penalty for the number of parameters in criteria like the BIC is arbitrary. The ranking of the models is objective only for models with the same number of variables. Therefore, we select the best model for each dimension, see Table 3. Here, Id is a model identification and NC is the number of covariates. The second and the third row show the estimated regression coefficients for the covariates and for the fixed seasonal effect (*sin* and *cos*) in both layers. Adding a further variable improves both the BIC and the MPE substantially for up to three variables. The MPE for the best model with five covariates is about 10 % and 15% smaller than that for the best model with three covariates, and the gain is small thereafter. Models with Al, Fe, Ca and SO_4 or K fit the DOC-concentrations well and are still parsimonious. The regression coefficients are rather stable across all these models. The largest differences occur when adding Al to Fe or vice versa.

5 DISCUSSION

We have extended a regression model with spatial-temporal correlations for the DOC-concentrations in the seepage water of a German forest and applied it to new data. We

Table 3: Best variable selections according to the MPE, which is multiplied by 10^4, and by the BIC for each number of covariates. The regression coefficients for the upper layer are provided in the first, those for the lower layer in the second row.

Id	NC	BIC	MPE	SO_4	Cl	Ca	K	Na	Fe	Al	sin	cos
0	0	316.2	1261.5	0	0	0	0	0	0	0	-.284	-.167
			818.1	0	0	0	0	0	0	0	-.231	-.020
128	1	1024.3	671.1	0	0	0	0	0	1.875	0	-.239	-.116
			445.2	0	0	0	0	0	2.398	0	-.208	-.013
256	1	1033.8	665.8	0	0	0	0	0	0	.998	-.216	-.104
			498.5	0	0	0	0	0	0	.827	-.195	-.014
144	2	1282.4	508.0	0	0	-.0201	0	0	1.497	0	-.232	-.085
			314.6	0	0	-.0168	0	0	1.961	0	-.195	-.002
272	2	1295.8	462.4	0	0	-.0198	0	0	0	.802	-.211	-.075
			362.4	0	0	-.0176	0	0	0	.690	-.184	-.002
400	3	1401.6	436.7	0	0	-.0174	0	0	.859	.514	-.214	-.075
			339.6	0	0	-.0167	0	0	1.652	.191	-.189	-.003
401	4	1475.8	423.9	.0144	0	-.0195	0	0	.882	.519	-.202	-.076
			291.3	.0216	0	-.0192	0	0	1.708	.291	-.178	-.018
433	5	1511.3	392.7	.0129	0	-.0206	.0175	0	.867	.520	-.188	-.077
			286.9	.0215	0	-.0193	.0166	0	1.757	.264	-.164	-.020
497	6	1547.5	385.4	.0158	0	-.0147	.0248	-.0509	.793	.521	-.201	-.083
			281.1	.0200	0	-.0253	.0085	.0517	1.886	.307	-.144	-.018
499	7	1555.6	396.2	.0173	-.0067	-.0136	.0251	-.0468	.774	.510	-.195	-.073
			252.4	.0214	-.0079	-.0239	.0177	.0540	1.769	.319	-.140	-.016

prefer simple model structures because of the many missing observations in this study. Spatial-temporal correlations have been modelled by random effects and an autoregressive structure of these effects. Maximum likelihood estimates can be calculated in spite of the many missing data using an EM algorithm. As a summary of a model validation, we find that some deficiencies of the basic model applied in Fried (2001) can be overcome by adding further sources of possible correlations.

Selection of the important covariates is difficult here since some of them are highly correlated. We applied the BIC and the out-of-sample mean square prediction error in this analysis. Both criteria confirm most of the findings reported in Fried (2001), where t-statistics and stepwise search strategies were used. There is evidence in favor of the hypothesis of humus disintegration, which predicts a strong association between DOC and aluminium compounds since we find a strong relationship between DOC and aluminium and / or iron ions. Excluding both variables from the model results in a poor model fit and worsens the predictions a lot. The almost interchangeability of aluminium and iron corresponds to the fact that both arise simultaneously from temporary soil acidifications. Although the hypothesis of humus disintegration predicts a relationship between DOC and nitrate, this could not be confirmed by our analysis. The influence of the pH (the oxonium ions) is uncertain as the regression coefficient is very variable, but this is due to its small variability in the data. Instead, we have identified other interesting compounds. Calcium, sulfate and potassium ions should be further investigated since their inclusion improves the model fit. This could not be expected based on a simple correlation analysis reported in Fried et al. (2001). Sulfate is an oxidation product of sulphur dioxide, emissions of which contribute to the acidification of forest soils. The inclusion of calcium agrees with ecological knowledge since it is well known to construct chemical complexes, which might be important in the humus cycle. The same is true for aluminium and iron. Hence, this study confirms some theories about the mechanisms of humus dynamics (Eichhorn and Hüttermann 1999).

The detected seasonality is probably caused by the dependence of biochemical mechanisms on environmental variables like the temperature, data for which we did not have available. Instead, a deterministic seasonal component and random effects allowing for positive correlations were included. The results indicate that the disintegration is increased by higher temperatures occurring in summer. A time trend did not become obvious in the analysis. This had been expected before since humus disintegration is supposed to increase slowly. Therefore the detection of changes needs a longer observation period than the few years analyzed here.

All these conclusions depend on the model identified in an exploratory data analysis, which has been extended here. The uncertainty introduced by this selection is not reflected in the results and can hardly be quantified. As more data becomes available, the predictive performance of the model can be assessed to detect model inadequacies. The possible curvilinearity of some relationships could be further examined using non-parametric regression. This might clarify the influences of a reduced set of covariates on the *DOC*-concentration. Dynamic evolution of the model parameters over time as well as spatial inhomogeneities of the parameters within each layer could also be considered. For the latter extension, random coefficient models may be applied (Longford 1993), where random variation of the regression coefficients about a fixed mean is assumed. In the former approach, dynamic parameters could account for seasonal variations in the relationships. Applying more complex models could be worthwhile since some of the weaker associations found here could be due to nonlinearities. Our analysis can be seen as a preliminary step for more complex approaches since we have achieved a dimension reduction identifying important covariates and finding a spatial-temporal correlation structure.

It is uncertain whether our results for the location Zierenberg also apply to other areas. Humus disintegration may strongly depend on environmental conditions. However, its causes are likely to be the same all over central Europe. Therefore our results may be relevant for other sites, too.

ACKNOWLEDGMENTS

The author is grateful to Johannes Eichhorn and Uwe Paar from the Hessian Agency of Forest Management Planning, Forest Research and Forest Ecology, Hannoversch Münden, Germany, for making the data available and their advice. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

APPENDIX: DERIVATION OF THE EM ALGORITHM

Augmenting the model of Shumway and Stoffer (1982) by covariates we see that twice the negative log-likelihood for $\underline{\tilde{z}} = (\underline{w}_0, \underline{w}_1, \underline{z}_1, \ldots, \underline{w}_T, \underline{z}_T)$ is (dropping constants)

$$\begin{aligned}\tilde{L} &\doteq \ln|\underline{\Omega}| + (\underline{w}_0 - \underline{\mu})'\underline{\Omega}^{-1}(\underline{w}_0 - \underline{\mu}) + T\ln|\underline{\Delta}| \\ &+ \sum_{t=1}^{T}(\underline{w}_t - \underline{G}\underline{w}_{t-1})'\underline{\Delta}^{-1}(\underline{w}_t - \underline{G}\underline{w}_{t-1}) + N\ln(\tilde{\sigma}_1^2) + \sum_{t=1}^{T}\ln|\underline{S}_t| \\ &+ \tilde{\sigma}_1^{-2}\sum_{t=1}^{T}\left(\underline{z}_t(\alpha) - \underline{\mathrm{X}}_t\underline{\beta} - \underline{\mathrm{H}}_t\underline{w}_t\right)'\underline{S}_t^{-1}\left(\underline{z}_t(\alpha) - \underline{\mathrm{X}}_t\underline{\beta} - \underline{\mathrm{H}}_t\underline{w}_t\right) .\end{aligned}$$

The conditional expectation of $\tilde{L}(\underline{\theta})$ given the actually observed values and a set of parameters $\underline{\tilde{\theta}}$ then is

$$\begin{aligned}Q_{\underline{\tilde{\theta}}}(\underline{\theta}) &= term_1 + term_2 \\ term_1 &= \ln|\underline{\Omega}| + tr\left(\underline{\Omega}^{-1}\left[\underline{\Omega}_{0|T} + (\underline{w}_{0|T} - \underline{\mu})(\underline{w}_{0|T} - \underline{\mu})'\right]\right) + T\ln|\underline{\Delta}| \\ &+ tr\Big(\underline{\Delta}^{-1}\sum_{t=1}^{T}\Big[\underline{\Omega}_{t|T} + \underline{w}_{t|T}\underline{w}'_{t|T} - (\underline{\Omega}_{t,t-1|T} + \underline{w}_{t|T}\underline{w}'_{t-1|T})\underline{G}' \\ &- \underline{G}(\underline{\Omega}_{t-1,t|T} + \underline{w}_{t-1|T}\underline{w}'_{t|T}) + \underline{G}(\underline{\Omega}_{t-1|T} + \underline{w}_{t-1|T}\underline{w}'_{t-1|T})\underline{G}\Big]\Big) \\ term_2 &= N\ln(\tilde{\sigma}_1^2) + \sum_{t=1}^{T}\ln|\underline{S}_t| + \tilde{\sigma}_1^{-2}\sum_{t=1}^{T} tr\left(\underline{S}_t^{-1}\underline{\mathrm{H}}_t\underline{\Omega}_{t|T}\underline{\mathrm{H}}'_t\right) \\ &+ \tilde{\sigma}_1^{-2}\sum_{t=1}^{T}(\underline{z}_t(\alpha) - \underline{\mathrm{X}}_t\underline{\beta} - \underline{\mathrm{H}}_t\underline{w}_{t|T})\underline{S}_t^{-1}(\underline{z}_t(\alpha) - \underline{\mathrm{X}}_t\underline{\beta} - \underline{\mathrm{H}}_t\underline{w}_{t|T})'\end{aligned}$$

The first term corresponds to the unobserved states $\underline{w}_t$, while the second corresponds to a recursive linear mixed model for the observed data with covariance matrices $\tilde{\sigma}_1^2\underline{S}_t$ and an additional term $\tilde{\sigma}_1^{-2}\sum_{t=1}^{T} tr(\underline{S}_t^{-1}\underline{\mathrm{H}}_t\underline{\Omega}_{t|T}\underline{\mathrm{H}}'_t)$, which only changes the calculation of the unknown variances and covariances. Therefore we can derive formulae for the maximum of $Q_{\underline{\tilde{\theta}}}$ w.r.t. $\underline{\beta}$ and α in dependence on the other parameters in the same way as in Fried (2001) using some minor modifications. This leads to the formula for $\hat{\tilde{\underline{\beta}}}_u$ stated in Section 3.2. The case of identical regression coefficients $\underline{\beta}_1 = \underline{\beta}_2$ (Fried 2001), which is not further considered here, can be treated similarly. If $\underline{\beta}_1 \neq \underline{\beta}_2$, partitioning $\underline{S}_t$ into the blocks for the layers and inserting $\hat{\tilde{\underline{\beta}}}_1$ and $\hat{\tilde{\underline{\beta}}}_2$ results into separate minimization of

$$\begin{aligned}&N_u\ln(\tilde{\sigma}_u^2) + \sum_{t=1}^{T}\ln|\underline{S}_{t,u}| + \tilde{\sigma}_u^{-2}\sum_{t=1}^{T} tr\left(\underline{S}_{t,u}^{-1}\underline{\mathrm{H}}_{t,u}\underline{\Omega}_{t|T}\underline{\mathrm{H}}'_{t,u}\right) \\ &+ \tilde{\sigma}_u^{-2}\sum_{t=1}^{T}(\underline{z}_{t,u} - \underline{\tilde{X}}_{t,u}\hat{\tilde{\underline{\beta}}}_u - \underline{\mathrm{H}}_{t,u}\underline{w}_{t|T})\underline{S}_{t,u}^{-1}(\underline{z}_{t,u} - \underline{\mathrm{X}}_{t,u}\hat{\tilde{\underline{\beta}}}_u - \underline{\mathrm{H}}_{t,u}\underline{w}_{t|T})' .\end{aligned}$$

This results in the formula for $\tilde{\sigma}_u^2$. Therefore, in the M-step it remains to apply numerical maximization to estimate the unknown parameters in $\underline{\tilde{\Sigma}}_u$, $\underline{\Omega}$, $\underline{\Delta}$, $\underline{G}$ and $\underline{\mu}$. For the special structure of the model used here we can deduce some further simplifications since the matrices $\underline{G}, \underline{\Omega}$ and $\underline{\Delta}$ are diagonal, $\underline{\tilde{\Sigma}}_u = \underline{\mathrm{I}}$ and $\underline{\mu} = \underline{0}$. We get

$$\begin{aligned}
& \tilde{\sigma}_1^{-2} \sum_{t=1}^{T} tr(\underline{S}_{t,1}^{-1} \underline{\mathrm{H}}_{t,1} \underline{\Omega}_{t|T} \underline{\mathrm{H}}_{t,1}') \\
= \; & \tilde{\sigma}_1^{-2} \sum_{t=1}^{T} tr \left[\begin{pmatrix} \underline{1}_{n_{t,1}} & \underline{1}_{n_{t,1}} & \underline{0}_{n_{t,1}} \end{pmatrix} \underline{\Omega}_{t|T} \begin{pmatrix} \underline{1}_{n_{t,1}} & \underline{1}_{n_{t,1}} & \underline{0}_{n_{t,1}} \end{pmatrix}' \right] \\
= \; & \tilde{\sigma}_1^{-2} \sum_{t=1}^{T} tr \left[\begin{pmatrix} n_{t,1} & n_{t,1} & 0 \\ n_{t,1} & n_{t,1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \underline{\Omega}_{t|T} \right] \\
= \; & \tilde{\sigma}_1^{-2} \sum_{t=1}^{T} n_{t,1} \left[(\underline{\Omega}_{t|T})_{(0,0)} + (\underline{\Omega}_{t|T})_{(1,0)} + (\underline{\Omega}_{t|T})_{(0,1)} + (\underline{\Omega}_{t|T})_{(1,1)} \right]
\end{aligned}$$

where $\underline{\mathrm{M}}_{(i,j)}$ denotes the element of a matrix $\underline{\mathrm{M}}$ at position (i, j). Analogously, we get

$$\begin{aligned}
& \tilde{\sigma}_2^{-2} \sum_{t=1}^{T} tr(\underline{S}_{t,2}^{-1} \underline{\mathrm{H}}_{t,2} \underline{\Omega}_{t|T} \underline{\mathrm{H}}_{t,2}') \\
= \; & \tilde{\sigma}_2^{-2} \sum_{t=1}^{T} n_{t,2} \left[(\underline{\Omega}_{t|T})_{(0,0)} + (\underline{\Omega}_{t|T})_{(2,0)} + (\underline{\Omega}_{t|T})_{(0,2)} + (\underline{\Omega}_{t|T})_{(2,2)} \right] ,
\end{aligned}$$

and furthermore

$$\begin{aligned}
term_1 \;=\; & -\sum_{u=0}^{2} \ln(1-\gamma_u^2) + (T+1) \sum_{u=0}^{2} \ln(\tilde{\tau}_u^2) + tr \left[\underline{\Omega}^{-1} (\underline{\Omega}_{0|T} + \underline{w}_{0|T} \underline{w}_{0|T}') \right] \\
& + tr \Big(\underline{\Delta}^{-1} \sum_{t=1}^{T} \Big[\underline{\Omega}_{t|T} + \underline{w}_{t|T} \underline{w}_{t|T}' - (\underline{\Omega}_{t,t-1|T} + \underline{w}_{t|T} \underline{w}_{t-1|T}') \underline{G}' \\
& - \underline{G} (\underline{\Omega}_{t-1,t|T} + \underline{w}_{t-1|T} \underline{w}_{t|T}') + \underline{G} (\underline{\Omega}_{t-1|T} + \underline{w}_{t-1|T} \underline{w}_{t-1|T}') \underline{G} \Big] \Big) \\
=\; & -\sum_{u=0}^{2} \ln(1-\gamma_u^2) + (T+1) \sum_{u=0}^{2} \ln(\tilde{\tau}_u^2) + \sum_{u=0}^{2} \frac{1-\gamma_u^2}{\tilde{\tau}_u^2} (\underline{\Omega}_{1|T} + \underline{w}_{0|T} \underline{w}_{0|T}')_{(u,u)} \\
& + \sum_{u=0}^{2} \tilde{\tau}_u^{-2} \sum_{t=1}^{T} \Big[(\underline{\Omega}_{t|T} + \underline{w}_{t|T} \underline{w}_{t|T}')_{u,u} - (\underline{\Omega}_{t,t-1|T} + \underline{w}_{t|T} \underline{w}_{t-1|T}')_{(u,u)} \gamma_u \\
& - \gamma_u (\underline{\Omega}_{t-1,t|T} + \underline{w}_{t-1|T} \underline{w}_{t|T}')_{u,u} + \gamma_u (\underline{\Omega}_{t-1|T} + \underline{w}_{t-1|T} \underline{w}_{t-1|T}')_{(u,u)} \gamma_u \Big] \\
=\; & -\sum_{u=0}^{2} \ln(1-\gamma_u^2) + \sum_{u=0}^{2} \tilde{\tau}_u^{-2} (a_u - 2 b_u \gamma_u + c_u \gamma_u^2) \\
\text{with } a_u \;=\; & \sum_{t=1}^{T} (\underline{\Omega}_{t|T} + \underline{w}_{t|T} \underline{w}_{t|T}')_{(u,u)},
\end{aligned}$$

$$
\begin{aligned}
b_u &= \sum_{t=1}^{T} (\underline{\Omega}_{t,t-1|T} + \underline{w}_{t|T}\underline{w}'_{t-1|T})_{(u,u)}, \\
c_u &= \sum_{t=0}^{T-1} (\underline{\Omega}_{t|T} + \underline{w}_{t|T}\underline{w}'_{t|T})_{(u,u)}, \quad u = 0, 1, 2\,.
\end{aligned}
$$

Inspection of the partial derivatives results into the formula for $\hat{\hat{\tau}}_u^2$ reported in Section 3.2. Insertion of $\hat{\hat{\tau}}_u^2$ leads to

$$
term_1 = -\sum_{u=0}^{2} \ln(1-\gamma_u^2) + (T+1)\sum_{u=0}^{2} \ln\left(\hat{\hat{\tau}}_u^2(\gamma_u)\right) + 3(T+1)
$$

so that separate minimizations of $-\ln(1-\gamma_u^2) + (T+1)\ln(\hat{\hat{\tau}}_u^2(\gamma_u))$ w.r.t. γ_u remains to be done. This can be done numerically or using the cardanic formulae since setting the partial derivatives to zero results into cubic equations:

$$
\begin{aligned}
\frac{\partial term_1}{\partial \gamma_u} &= \frac{2\gamma_u}{1-\gamma_u^2} + \frac{2(T+1)^2(c_u\gamma_u - b_u)}{a_u - 2b_u\gamma_u + c_u\gamma_u^2} = 0 \\
\Longleftrightarrow \quad 0 &= \gamma_u(c_u\gamma_u^2 - 2b_u\gamma_u + a_u) + (T+1)^2(1-\gamma_u^2)(c_u\gamma_u - b_u) \\
&= -\gamma_u^3 c_u(T^2+2T) + \gamma_u^2 b_u(T^2+2T-1) + \gamma_u(a_u + c_u(T+1)^2) - b_u(T+1)^2
\end{aligned}
$$

REFERENCES

Buonaccorsi, J. P., and Elkinton, J. S. (2002), "Regression Analysis in a Spatial-temporal Context: Least Squares, Generalized Least Squares, and the Use of the Bootstrap," *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 4–20.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm" (With Discussion), *Journal of the Royal Statistical Society*, Series B, 39, 1–38.

Eichhorn, J. (1995), "Stickstoffsättigung und ihre Auswirkungen auf das Buchenwaldökosystem der Fallstudie Zierenberg," *Berichte des Forschungszentrums Waldökosysteme*, Reihe A, 124, Göttingen, Germany.

Eichhorn, J., and Hüttermann, A. (1999), "Mechanisms of Humus Dynamics and Nitrogen Mineralisation," in N. Rastin and J. Bauhus (eds.) *Going Underground – Ecological Studies in Forest Soils*, Trivandrum (India): Research Signpost, 239–277.

Fried, R. (1999), *Räumlich-zeitliche Modellierung der Kohlenstoffkonzentration im Waldbodensickerwasser zur Untersuchung der Hypothese der Humusdisintegration*, Ph.D-thesis, Department of Mathematics, Darmstadt University of Technology, D-64289 Darmstadt, Germany.

Fried, R. (2001), "Autoregressive Linear Mixed Models for Investigating Humus Disintegration," *Biometrical Journal*, 43, 757–766.

Fried, R., Eichhorn, J., and Paar, U. (2001), "Exploratory Analysis and a Stochastic Model for Humus Disintegration," *Environmental Monitoring and Assessment*, 68, 273–295.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (With Discussion), *Journal of the Royal Statistical Society*, Series B, 38, 205–246.

Heinz, J., and Spellucci, P. (1994), "A Successful Implementation of the Pantoja-Mayne SQP Method," *Optimization Methods and Software*, 4, 1–28.

Hodges, J. L., and Lehmann, E. L. (1954), "Testing the Approximate Validity of Statistical Hypotheses," *Journal of the Royal Statistical Society*, Series B, 16, 259–268.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations," *Technometrics*, 22, 389–395.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 34–45.

Longford, N. T. (1993), *Random Coefficient Models,* Oxford: Clarendon Press.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Ulrich, B. (1981), "Theoretische Betrachtung des Ionenkreislaufs in Waldökosystemen," *Zeitschrift für Pflanzenernährung und Bodenkunde*, 144, 647–659.

Wermuth, N. (1980), "Linear Recursive Equations, Covariance Selection, and Path Analysis," *Journal of the American Statistical Association*, 75, 963–972.

West, M., and Harrison, J. (1989), "Bayesian Forecasting and Dynamic Models," New York: Springer.

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11, 95–103.