Trends in Forest Soil Acidity: A GAM Based Approach with Application on Swedish Forest Soil Inventory Data

By Staffan Betnér

Department of Statistics
Uppsala University

Supervisor: Shaobo Jin

2018
Abstract

The acidification of soils has been a continuous process since at least the beginning of the 20th century. Therefore, an inquiry of how and when the soil pH levels have changed is relevant to gain better understanding of this process. The aim of this thesis is to study the average national soil pH level over time in Sweden and the local spatial differences within Sweden over time. With data from the Swedish National Forest Inventory, soil pH surfaces are estimated for each surveyed year together with the national average soil pH using a generalized additive modeling approach with one model for each pair of consecutive years. A decreasing trend in average national level soil pH was found together with some very weak evidence of year-to-year differences in the spatial structure of soil pH.

Keywords

soil pH, soap film smoothing, generalized additive models, spatial modeling
Table of Contents

1. Introduction 4

2. Data 6
   2.1 Descriptive Statistics 6

3. Methodology 8
   3.1 Generalized Additive Models 8
   3.2 Model 9
      3.2.1 Soap film smoothing 9
      3.2.2 Testing Procedure 10
   3.3 Smooth Functions 11
      3.3.1 Basis expansion 11
      3.3.2 Basis functions and knots 11
   3.4 Estimation 12
      3.4.1 Smoothing parameter 12

4. Results 13
   4.1 Regional Temporal Differences 13
   4.2 National Temporal Differences 16

5. Discussion 17

References 19

Appendix 22
1. Introduction

During the 20th century the forest soils in Sweden have gotten more acid (Falkengren-Grerup, Linnermark, & Tyler, 1987). Soil pH is considered to be the master variable of soils and is of great importance to investigate since it affects numerous chemical processes in the soil. The availability of plant nutrients and microorganisms are affected by soil pH and at low pH levels aluminum, iron and manganese become more soluble which could have a toxic effect on the plants which will, e.g. result in worse growth of trees and crops (Sparks, 2003). To investigate which areas of Sweden where the acidification increases, or decreases is therefore of big importance.

Spatio-temporal models provide useful tools for environmental analysis. Within the previous research of soil pH, modeling or testing only with respect to time has been the main approach for assessing changes in soil pH, typically with measurements at approximately the same locations repeated typically just a few times, often only two times. In some cases, the samples have been taken from different locations, for which the samples have been considered independent and relevant methods assuming independence have been utilized. Since independence, both with respect to time and to space, is an unrealistic assumption due to the spatio-temporal nature of measures of soil pH, this thesis aims to extend this toolbox and showing its relevance.

The only study with a modeling approach of the soil acidity in Sweden over time is from the Department of Soil and Environment at SLU, published at their web site “MarkInfo”¹. The data they used is between 1967 and 1999. Their approach was to divide the mainland into a grid, where each square consists of spatially weighted measures. The change in each square was determined with a linear regression, and its significance with separate F-tests. Several squares showed significant decreases, mainly in northeastern Sweden and in western middle Sweden, but a few significant changes are scattered throughout southern Sweden as well.

Several other studies have investigated the temporal change of acidity in different types of soils and depositions in different countries with different methods, e.g. Curtis & Simpson (2014) used generalized additive models (GAM) with estimated smooth function for time trend and cyclic smooth function for within year variability for each sampled location. Hallbäcken & Tamm (1986) collected previously measured (in 1929) soil pH from 90

¹ http://www-markinfo.slu.se/
locations in south-west Sweden and conducted repeated measured at the same locations between 1982 and 1984.

They found a decrease in average soil pH in all soil depths between 0.3 and 0.9 pH units. Falkengren-Grerup et al. (1987) compared the pH in the surface layer at 104 sites, where the earlier samples were collected in a time period between 1949 and 1970, and the later samples collected in 1983 and 1985. At most of the compared sites, the measured pH had declined, a majority by larger than 0.5 units. Ilvesniemi (1991) collected data from a few locations in southwest Finland for three time periods in the 80s. They conducted unpaired t-tests and found significant decrease in the two sampled upper soil depths. Lapenis et al. (2004) compared measures from 1893 from three sites in western Russia with measures at the same locations in 1998. They found a relationship between depth and the decrease in soil pH, where the lower depths showed smaller differences. Bailey, Horsley, & Long (2005) compared data from 1967 and 1997 at four sites at the Allegheny Plateau in Pennsylvania, US. They found significant pH decreases in all depths using an ANOVA test. Hédl, Petřík, & Boublík (2011) used data from 20 sites in Eastern Sudetes Mountains in 1941 and 2003. They used an ANOVA model with altitude, time, depth and interactions with time to model the pH measures. They found a significant decrease in average soil pH. Akselsson et al. (2013) used data over 18 to 22 years from the time period 1986 to 2008 for nine sites in southern Sweden. They used seasonal Kendall tests and found some decreasing trends, but also some evidence of an increase in soil pH at some sites. Zhu et al. (2016) compared data from six subregions of China with data from two time periods, 1981-1985 and 2006-2010. With unpaired t-tests, they found significant decreases in all subregions except northwest China.

The goal of the thesis is to investigate local trend change of acidity in the humus layer throughout mainland Sweden using the GAM framework, which provides a considerable flexibility for spatio-temporal data. For example, it is possible to account for the spatial structure of the sample values, which previous methods fail to do. By the GAM framework, the regional spatial differences between each consecutive surveyed year and the national yearly average are modeled for the whole surveyed period.

This introduction is followed first by a section about the data, then a section describing the methodology used for achieving the objectives of this thesis. The thesis is ending with a result section and a discussion section.
2. Data

Data are made available from the Department of Soil and Environment at Swedish University of Agricultural Sciences and covers 1983 to 2012, with a gap from 1988 to 1992. 1993 and 1994 were surveyed, but only a different half of the country each time, and therefore those years are excluded from further analysis.

The data consists of spatial coordinates for the sampling locations (in meters from the prime meridian in Greenwich and from the equator, further denoted by easting and northing), the year of the conducted survey, and the pH in the humus layer in the sampled locations.

2.1 Descriptive Statistics

Figure 1 shows maps for each surveyed year and the value of the relevant variable, soil pH in the humus layer, for each sampling location. As can be seen, the sampling locations differ from time to time. From 1995 and onwards, the number of sampling locations is reduced by 50% from the previous period resulting in a lower sampling density. It should be noted that all measured pH values are considered acid, as a pH of 7 is considered neutral. Since sampling locations are not fixed over time, the yearly national averages are not reported here. For details of the sampling design regarding the sampling locations the reader is referred to Fridman et al. (2014).
Figure 1. Plot of sampling locations for each surveyed year and the corresponding soil acidity for that location.
3. Methodology

Previous studies have mostly used unpaired t-tests or ANOVA tests to test for trends in soil pH levels. These methods assume independence of the relevant variable, which is violated through its spatial dependency. In some cases, the approach with repeated sampling is used, where the paired t-test is suitable, but only for testing the average pH level. A more suitable approach is to incorporate spatial structure together with the average pH level, which is possible with the GAM framework. Due to the nature of the available dataset, one of few methods that come in question is to estimate and compare surfaces for each time point, which is possible by GAM. The testing will thus be analogous to paired t-testing for the average pH level. The main reason to use a more flexible modeling framework such as generalized additive models is to incorporate spatial structures.

3.1 Generalized Additive Models

GAM, introduced by Hastie & Tibshirani (1986), is a modeling framework which is an extension to generalized linear models (GLM). A GLM model is structured as

$$ g(E(Y_i | X_i = x)) = \beta_0 + \mathbf{\beta}^T \mathbf{X}_i, $$

in which $g(\cdot)$ is a response link function, $Y_i$ is a response variable, $X_i$ is the matrix of covariates, $\beta_0$ is an intercept, and $\mathbf{\beta}$ is the vector of coefficients for linear functions of covariates. $Y_i$ conditioned on $\mathbf{X}$ is assumed to follow an exponential family distribution. The extension consists in replacing $\mathbf{\beta}^T \mathbf{X}$ with a sum of unknown smooth functions of the covariates ($\sum_{j=1}^{p} f_j(x_i)$), resulting in the model structure:

$$ g(E(Y_i | X_i)) = \beta_0 + \sum_{j=1}^{p} f_j(x_i) $$

(Hastie & Tibshirani, 2006; Wood, 2017), where every $f_j(x_i)$ needs to be estimated.

Several generalizations of this setup of GAM is available, most notably generalized additive models for location, scale, and shape (GAMLSS), which allows modeling of several parameters of the response's distribution e.g. such as the scale parameter if heteroscedasticity is present in the data (Rigby & Stasinopoulos, 2005).

---

2 A function that is continuously differentiable up to a desired order over a domain is defined as a smooth function.
3.2 Model

For each pair of consecutive observed years, a model with the following structure is estimated:

\[ g(Y_i) = \alpha_1 + \gamma_1 I[\text{Second year}] + \sum_{t=1}^{2} (f(\text{Easting}, \text{Northing})_t - \mu_t) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2). \]

The response variable is a Box-Cox transformed pH level at a certain point, modeled as a function of an intercept (corresponding to the average transformed pH level in the reference, i.e. the first year in the pair), a dummy variable for the difference between the first and second year in the pair, and two separate temporal mean-corrected splines for the regional structure in the two years. \( i \) is the index for the observations. This structure of the model allows for separation of the testing of the average national level each year and the testing for the difference in local structure of the pH levels. It is due to computational restrictions that the differences will be estimated by a set of models instead of one model.

The Box-Cox transformation parameter is estimated by a GAMLSS model with the same structure, where the response variable is assumed to follow a truncated skewed normal distribution, where the estimated skewness parameter is equal to the Box-Cox transformation parameter. The reason for not using models of this structure is that the calculations of simultaneously critical values is impractical due to the estimated smooth function being separated into two: the at-boundary smooth and the within boundary smooth. When trying to estimate those parts together, the computational resources available was insufficient for estimation and thus a simpler Box-Cox transformation is used with the transformation parameter from the model for the first two consecutive years.

3.2.1 Soap film smoothing

Since many smoothing methods are available for two-dimensional data, e.g. thin plate smoothing, it is needed to decide on one type. In this thesis soap film smoothing will be used, since it is appropriate when modeling data over bounded domains (Wood, Bravington, & Hedley, 2008), such as in this thesis, the area of mainland Sweden.

The soap film smoother is defined as a solution to the partial differential equation (PDE):

\[ \frac{\delta^2 f}{\delta x^2} + \frac{\delta^2 f}{\delta y^2} = \rho, \] where \( f = s \) on the boundary of the domain. \( s \) is typically a cyclic smoother. \( \rho \) is defined as the solution to the PDE \[ \frac{\delta^2 \rho}{\delta x^2} + \frac{\delta^2 \rho}{\delta y^2} = 0, \] which is 0 at the boundary.
The soap film smoother can be formulated in the following way:

\[
    f(x, y) = s + (f - s) = \sum_{j} \beta_j a_j(x, y) + \sum_{k} \gamma_k g_k(x, y),
\]

where \( \beta_j \) and \( \gamma_k \) are the coefficients for the two parts, and \( a_j(x, y) \) and \( g_k(x, y) \) are the basis functions. For the used dataset, \( x \) corresponds to the easting, and \( y \) to the northing. The first term is the at boundary smoother with its coefficients \( \beta \) and index \( j \), and the second term corresponds to the part which is estimated with numerically solutions to the PDEs. The at boundary basis is a cyclic cubic spline. For details of the penalties for the soap film smoothing, the reader is directed to the appendix.

3.2.2 Testing Procedure

For the differences in the local structure, a simulation procedure is performed to simulate the uncertainty of the whole spline to get an approximate critical value (for a significance level of 0.05) so one rejected null hypothesis would implicate a structural difference of the pH level distribution. If the pointwise tests are not adjusted for multiplicity, the probability of making Type I errors will not correspond to a family-wise error rate of 5%. When this correction is done, the significance level will correspond to the family-wise error rate, so one rejected null hypothesis will indicate a significant difference for the whole function and in the evaluated point, instead of just in the evaluated point.

First, a grid of locations where the splines will be evaluated is chosen. The locations are a number of 746 evenly spread locations throughout Sweden. Then evaluation of both splines’ basis functions is done at all these locations followed by a calculation of the differences. To obtain the standard errors of the differences \( X_d \tilde{V_c} X_d^T \) is calculated, where \( X_d \) is a matrix of said differences, and \( \tilde{V_c} \) is the corrected estimator of the covariance matrix of the estimated coefficients.

The calculation of the critical values for the simultaneously testing is following the approach outlined by Ruppert, Wand & Carroll (2003), see their page 142 for a theoretical outline. In practice the procedure consists of taking many samples from a multivariate normal distribution \( \left( N(0, \tilde{V_c}) \right) \), calculating \( \hat{f}_g - f_g \) by multiplying the samples from the MVND with the evaluations of the estimated function, taking the absolute values of the differences divided by the standard error, taking the maximum for each simulation and picking the corresponding \( 1 - \alpha \) quantile as critical value. \( \hat{f}_g \) is the vector of estimated functions and \( f_g \).
is the vector of the true functions. p-values are calculated by determining a scaling coefficient by how much larger the critical value for the simultaneously confidence intervals are and scaling each standard error with that coefficient. For each location in the grid, the null hypothesis of $X_{year_1} - X_{year_2} = 0$ versus the alternative hypothesis of $X_{year_1} - X_{year_2} \neq 0$ are tested with z-tests, using the estimated differences as observed mean, and the scaled standard error as standard error.

3.3 Smooth Functions

In this section, the theory of smooth functions and their estimation will be outlined. In this subsection (3.3) and in the following subsection (3.4) the reader is directed to Wood (2017) for details.

3.3.1 Basis expansion

The smooth function is approximated by $f_i(x) = \sum_{j=1}^{k} b_j(x) \beta_j$, where each $b_j$ constitutes a basis function and $\beta_j$ its weight coefficient. The number of basis functions, i.e. the basis dimension (also known as the number of knots), determines the smoothness of the estimated function. Since a large basis dimension will overfit the function, a penalty is introduced in the estimation. For example, given a continuous response variable, instead of fitting by minimizing the sum of squares ($||y - X\beta||^2$), a function of the form $||y - X\beta||^2 + \lambda J$ is minimized, where the second term consists of two parts. The first multiplier, $\lambda$, is the smoothing parameter, and the second, $J$, is a measure of wiggliness, which will differ due to the used basis but involve the integral or sum of the square of the second derivative of the function(s).

3.3.2 Basis functions and knots

Several types of basis functions, typically different types of splines, are commonly used in the literature, such as cubic splines, B-splines, P-splines, etc.

The knots’ number together with their location does not have a considerable influence on the fitted function, but it is the smoothing parameter that have the largest influence on the shape of the estimated function. For soap film smoothing it is needed to set the number of knots together with their locations a priori and in this case, the knot’s locations are a random

3 For this context, smoothness refers to the wiggliness, and not the order of differentiability
sample of approximately 500 sampling locations drawn from the dataset. The reason is that the knots will be spread throughout the bounded domain, which is of greater importance than the exact number of knots.

3.4 Estimation

In this setting, a penalized iteratively re-weighted least squares (PIRLS) procedure will be used to estimate the model coefficients, where every basis will have each own smoothing parameter. In every iteration of PIRLS, \( \hat{\beta} \) minimizes the function \( ||z - X\beta||_W^2 + \sum_j \lambda_j \beta^T S_j \beta \), where \( z \) is a function of the response variable and the estimated expectation. See Wood (2017) (p. 251) for details.

All calculations are done in the statistical programming environment R (R Core Team, 2017) with the packages mgcv (Wood, 2018) and GAMLSS (Stasinopoulos et al., 2018).

3.4.1 Smoothing parameter

The choice of estimation method is important when utilizing a smoothing approach. Since it is theoretically possible to more or less perfectly describe the covariates’ relationship with the response variable with enough number of splines and since modeling of noise in the relationship is undesirable, the two most common choices for estimation of smoothing parameter is either a cross-validation approach or a restricted maximum likelihood approach (REML). Since utilizing cross validation for selection of the smoothness parameter is prone to overfitting (Wood, 2011), REML will be used for estimation.

The smoothing penalty can be shown to have a Bayesian interpretation, which is equivalent to a mixed model structure, such that if multivariate normal priors is imposed on the model coefficients \( \beta \sim N \left( 0, \frac{\sigma^2 S^-}{\Lambda} \right) \), the posterior distribution for the parameters follows a normal distribution: \( \beta | y \sim N \left( \hat{\beta}, (X^T X + \lambda S)^{-1} \sigma^2 \right) \). This allows for estimating the smoothing parameters and the variance with restricted maximum likelihood.

The smoothing parameters are found by maximizing the Bayesian log marginal likelihood \( V_r(\lambda) = \log \int f(\mathbf{y}|\beta)f(\beta)d\beta \), where a Laplace approximation is used to approximate the integral. The PIRLS procedure for the estimation of the model coefficients

\[ S^- \text{ is the pseudo-inverse of } S, \text{ based on its eigen-decomposition } S = U \Lambda U^T, \text{ where } \Lambda^{-1} \text{ is the inverse of the diagonal matrix of the non-zero eigenvalues (where the inverse of any zero eigenvalues is replaced with a zero), } S^- = U \Lambda^{-1} U^T \]
are run for a set of candidate smoothing parameters and the selection of smoothing parameter is done after each PIRLS run has converged. The estimation of the smoothing parameter is thus outside the estimation of the model coefficients. For details, see Wood, Pya, & Säfken, (2016).

4. Results

In this section, results are demonstrated. Only figures showing the results are placed in the text. More results can be found in the appendix.

The model presented in Section 3.2 is fitted. The semivariograms of the residuals (Figure C1) show reasonable patterns for spatially independent residuals for all yearly surfaces. The QQ plots (Figure C2) hints at slight deviations from the assumption of normally distributed residuals, which might influence the nominal size of the tests.

4.1 Regional Temporal Differences

In Figure 2, the estimated year-to-year differences, with national yearly average difference removed, are reported. Figure 3 provides the simultaneous p-values (within each year-to-year comparison) for the two-sided test of null hypothesis of no differences. A significance level of 5% is used.

Just four of the comparisons of the consecutive years show significant differences. The first two pairs are 1995-1996 and 1996-1997. For the first pair a decrease in a region of southern Sweden is significant, and the following year an increase is significant in the same region, which is probably a year-to-year local mean-reverting shock. The next significant differences, the difference between 2003 and 2004, is a decrease in pH near the common border of Dalarna and Hälsingland, in upper-southern Sweden. The last pair with significant differences is 2007-2008 where a decrease is estimated. In the following pair of years, 2008-2009, an increase is estimated, but that difference is not significant. There are many spots of differences that is not significant.
Figure 2. Estimated differences (back transformed to original pH scale) between pair of consecutive years.
Figure 3. Plot of simultaneous p-values (within each comparison) for the two-sided null hypothesis of no difference between the two compared years. For four pairs (1995-1996, 1996-1997, 2003-2004, and 2007-2008) simultaneous p-values lower than 0.05 is obtained.
4.2 National Temporal Differences

For the differences at the national level, the estimated intercept for each model is used as the estimate of the average national level for the first year in each pair. Confidence intervals of 95% confidence are calculated with the inverse z-test and those are adjusted with Bonferroni correction for the number of all possible pairwise permutations (i.e. 253 comparisons), such that instead of using 1.96 as critical value for ordinary two-sided confidence intervals with 95% confidence, 3.722 is used.

In Figure 4 below, the estimated national average pH levels are reported. The estimated average national soil pH levels show a decreasing trend during the surveyed period with pH levels around 3.9 during the beginning of the 1980s. Even though the Bonferroni correction of the confidence intervals is conservative, the trend is persistent to the correction. During the end of the period the decreasing trend seem to level off around 3.75 and thus the difference between the beginning and the end of the surveyed period is around 0.15 pH units. The year-to-year volatility for the estimates is quite high, particularly for the years after 2000.

![Figure 4](image_url)

Figure 4. Plot of estimated average pH level in the humus layer with confidence intervals corrected with Bonferroni procedure for all possible contrasts and (in bold dark gray) a locally polynomial scatterplot smoothing line for ease of interpretation. See appendix for a table with estimated coefficients and corresponding numbers.
5. Discussion

In this thesis, soil pH surfaces in Sweden together with the national average soil pH over time is simultaneously estimated using models from the GAM framework. With a GAM approach, several significant differences are found, both between years at the aggregate national level and within the country over time. Large variations are observed, but only a few are found to be significant, when multiple comparisons within each surface is taken into consideration with simultaneous testing.

As in most of the previous mentioned studies, this thesis finds significant average decrease of soil pH over time. The most interesting comparison to do is with Hallbäcken & Tamm (1986), since they compared soil pH levels in Sweden from 1929 with data from 1982-1984, which is the beginning of this thesis’ time of interest. As they found differences on a larger magnitude, around 0.3-0.4 pH units, this might indicate that the major acidification occurred earlier during the 20th century. For a majority of studied sites, Falkengren-Grerup et al. (1987) found the majority of estimated decreases to be larger than 0.5 pH units.

The MarkInfo (2007) study also modelled the local variations of Swedish soil pH over time. However, they only considered the total difference over time, without considering the national trend. The results in this thesis are not always in line with their findings. There are spots in both this thesis’ maps and the MarkInfo study’s map that overlap in direction of change, but also different spots where there is no correspondence in direction of change. The lack of correspondence might be due to different temporal windows of this study and their study and different models used.

As previously mentioned, only a few significant changes are found in the current study. The simultaneously p-values, obtained by simulation, would asymptotically provide nominal size to the simultaneous tests within each pair of surfaces. Although, since the family-wise error rate is not corrected between the different surfaces, and it is likely with up to four rejected null hypotheses for 22 comparisons, given they all are true, there is only weak evidence for structural year-to-year differences over time. In addition, as the power of these tests is not studied thoroughly in the literature, the power might also be potential issue for this approach. There are a couple of other model specifications that might give more power to the tests of differences.

In this study, the surfaces for each pair of years are estimated, instead of e.g. estimating a difference surface for each pair of years, or even differences for all years compared to a reference year within the same model. When estimating the surfaces, the
national average is allowed to be freely estimated in the sense that models for consecutive surveyed years produce similar but not exactly the same average. Restricting the national averages for overlapping years from different models, or combining them after estimation, to be the same would increase the power, but it is not feasible within the R package mgcv. Modelling the difference directly would however allow for testing the hypothesis whether the estimated surface as a whole is different from a flat surface at zero, i.e. no difference at all between the two years. One additional way would be to model the spatial coordinates and time simultaneously and evaluate a first derivative (with respect to time) surface at each surveyed year, and in a similar manner to this thesis’ approach test whether there the first derivative is different from zero for a grid of locations. It would also be possible to model the national average pH level explicitly as a separated margin of the three-dimensional function of easting, northing and year. All of the suggested models could also incorporate other covariates measured together with the soil pH. These approaches are suggested as future directions for continued research in this area.

Above all, a decreasing trend in the Swedish soil pH during the surveyed time period and weak evidence for year-to-year structural differences is found in this thesis. According to the classification of United States Department of Agriculture Soil Survey Division, the levels of soil pH during the surveyed time period all belong to the class of extremely acid (Soil Survey Division Staff, 1993). This is indicative of the severeness of the current situation, but hopefully the pH levels will trend upwards, which is hinted in Figure 4.
References


Appendix

A. The Penalty of Soap Film Smoothing

The wiggleness penalty for the within boundary smooth is $J_Ω(f) = \int_Ω \left( \frac{\delta^2 f}{\delta x^2} + \frac{\delta^2 f}{\delta y^2} \right)^2 \, dx \, dy$, and the function being minimized is $||z - f||^2 + \lambda_s J_B(s) + \lambda_f J_Ω(f)$, where $z$ are noisy observations, $f$ estimated functions, $\lambda_s$ smoothing parameters for the cyclic at boundary smoother, and $J_B(s)$ the wiggleness penalty for the at boundary smoother.
B. Theory of Simultaneous Testing

Consider the true function $f(Easting,Northing)$ at $M$ locations in the national coordinate space. Denote the locations by $g = (g_1, g_2, \ldots, g_M)$, where each $g_i$ is a vector of the locations coordinates (easting and northing). The true functions evaluated at $g$, is defined as

$$f_g \equiv \begin{bmatrix} f(g_1) \\ f(g_2) \\ \vdots \\ f(g_M) \end{bmatrix},$$

and the corresponding estimated functions’ evaluations as

$$\hat{f}_g \equiv \begin{bmatrix} \hat{f}(g_1) \\ \hat{f}(g_2) \\ \vdots \\ \hat{f}(g_M) \end{bmatrix}.$$ 

The differences between the true and the estimated functions is $c_g \left[ \hat{\beta} - \beta \right] - u$, where $c_g$ is the evaluations of the basis functions at $g$, and the second multiplier is the variation in the estimated coefficients of the splines and is approximately distributed as a multivariate normal distribution (MVND) with a vector of zeroes for means, and the estimated Bayesian covariance matrix of the estimated coefficients corrected for uncertainty.

Thus the simultaneously confidence intervals are $\hat{f}_{g_i} \pm m_{1-\alpha} \hat{\sigma}_f(g_i) - f(g_i)$, where $m_{1-\alpha}$ is the $1-\alpha$ quantile of $\sup_{x \in X} \left| \frac{f(x) - \hat{f}(x)}{\hat{\sigma}_f(x)} \right|$. The standard error for the estimated function evaluated at $g_i$. 

$$\hat{\sigma}_f(g_i) - f(g_i) \approx \max_{1 \leq i \leq M} \left| \frac{c_g \left[ \hat{\beta} - \beta \right] - u}{\hat{\sigma}_f(g_i) - f(g_i)} \right|,$$ 

and $\hat{\sigma}_f(g_i) - f(g_i)$ is the standard error for the estimated function evaluated at $g_i$. 


C. Model Diagnostics

Figure C1. Semivariograms for each estimated yearly surface.
Figure C2. QQ plots for the residuals for each model.
Figure C3. Plot of predicted values vs. residuals for each model
D. Estimated coefficients

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimated average</th>
<th>Confidence interval (95%)</th>
<th>Estimated average</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower limit</td>
<td>Upper limit</td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td>3.930210</td>
<td>3.895031</td>
<td>3.966896</td>
<td>0.276427</td>
</tr>
<tr>
<td>1984</td>
<td>3.969004</td>
<td>3.935778</td>
<td>4.003557</td>
<td>0.276498</td>
</tr>
<tr>
<td>1985</td>
<td>3.873522</td>
<td>3.841723</td>
<td>3.906567</td>
<td>0.276318</td>
</tr>
<tr>
<td>1986</td>
<td>3.855909</td>
<td>3.825000</td>
<td>3.887999</td>
<td>0.276283</td>
</tr>
<tr>
<td>1987</td>
<td>3.835782</td>
<td>3.804606</td>
<td>3.868167</td>
<td>0.276241</td>
</tr>
<tr>
<td>1995</td>
<td>3.817440</td>
<td>3.774416</td>
<td>3.862813</td>
<td>0.276202</td>
</tr>
<tr>
<td>1996</td>
<td>3.837224</td>
<td>3.795213</td>
<td>3.881458</td>
<td>0.276244</td>
</tr>
<tr>
<td>1997</td>
<td>3.901808</td>
<td>3.855190</td>
<td>3.951133</td>
<td>0.276374</td>
</tr>
<tr>
<td>1998</td>
<td>3.822949</td>
<td>3.778246</td>
<td>3.870188</td>
<td>0.276214</td>
</tr>
<tr>
<td>1999</td>
<td>3.794279</td>
<td>3.751776</td>
<td>3.839086</td>
<td>0.276152</td>
</tr>
<tr>
<td>2000</td>
<td>3.759862</td>
<td>3.714366</td>
<td>3.808035</td>
<td>0.276075</td>
</tr>
<tr>
<td>2001</td>
<td>3.823260</td>
<td>3.770556</td>
<td>3.879525</td>
<td>0.276215</td>
</tr>
<tr>
<td>2002</td>
<td>3.742923</td>
<td>3.696039</td>
<td>3.792667</td>
<td>0.276036</td>
</tr>
<tr>
<td>2003</td>
<td>3.848612</td>
<td>3.794835</td>
<td>3.906076</td>
<td>0.276268</td>
</tr>
<tr>
<td>2004</td>
<td>3.748171</td>
<td>3.698591</td>
<td>3.800959</td>
<td>0.276048</td>
</tr>
<tr>
<td>2005</td>
<td>3.693442</td>
<td>3.646464</td>
<td>3.743333</td>
<td>0.275917</td>
</tr>
<tr>
<td>2006</td>
<td>3.746819</td>
<td>3.695880</td>
<td>3.801149</td>
<td>0.276045</td>
</tr>
<tr>
<td>2007</td>
<td>3.776703</td>
<td>3.722752</td>
<td>3.834441</td>
<td>0.276113</td>
</tr>
<tr>
<td>2008</td>
<td>3.776024</td>
<td>3.726451</td>
<td>3.828776</td>
<td>0.276112</td>
</tr>
<tr>
<td>2009</td>
<td>3.761609</td>
<td>3.714538</td>
<td>3.811549</td>
<td>0.276079</td>
</tr>
<tr>
<td>2010</td>
<td>3.731116</td>
<td>3.683086</td>
<td>3.782164</td>
<td>0.276008</td>
</tr>
<tr>
<td>2011</td>
<td>3.742041</td>
<td>3.694880</td>
<td>3.792099</td>
<td>0.276034</td>
</tr>
<tr>
<td>2012</td>
<td>3.785527</td>
<td>3.700506</td>
<td>3.880328</td>
<td>0.276133</td>
</tr>
</tbody>
</table>

Table D1. Table of estimated average pH levels with Bonferroni corrected confidence intervals, with corresponding average and standard error for the transformed scale, in the humus layer for each observed year. (The values for 2012 are calculated with a convolution of two normal distributions with the estimated average for 2011 and the difference between 2012.)