ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

Differential and integral transformations of parametric functions in biometry

Tõnu Möls and Marek Tuul

Abstract. We present some useful applications of the linear statistical covariance modelling. The common classical model $Y = X\beta + \varepsilon$ is assumed to contain at least one continuous variable in X. Treating the model as a parametric function $X\beta$, and applying certain linear operators on X, makes it possible to get additional information about the dependent variable Y. In particular, it is possible to estimate derivatives, Riemann integrals and Fourier transforms of the dependent variable. The proposed methods are illustrated on real chemical data of Lake Peipsi (Estonia/Russia). Examples cover the estimation of dynamics of changes in the concentration of chemical substances in Lake Peipsi, and the estimation of the total quantity of a substance heterogeneously distributed in the lake. Calculations are carried out with the SAS software.

1. Introduction

During several years large hydrochemical and hydrobiological data sets of Lake Peipsi were analyzed using covariance models containing up to hundreds of parameters (Möls and Starast, 2000), (Starast, Milius, Möls and Lindpere, 2001). These models present a detailed analytical description of how water properties have changed during years, and how they depend on day within the year, geographical coordinates, depth etc. Using parametric functions, it was possible to compare different years or different parts of the lake, to test the seasonality, depth-dependence, etc. But we could not find in literature statistical applications of analytical properties of complicated regression lines. For example, we did not find methods for estimation of changes of concentration in time by using the derivative of the regression

Received December 15, 2003.
2000 *Mathematics Subject Classification.* 62P12.
Key words and phrases. Parametric functions of linear model, derivative of dependent variable, integral of dependent variable.
The authors are thankful for financial support through Estonian Science Foundation Grants 5686 and 6008 and Target Financed Project 0362483s03.

line. Another unsolved problem was the estimation of the total amount of a substance distributed inhomogeneously in space.

In the present paper we propose a solution for these and analogous problems. The main idea is to consider the predicted value of the dependent variable as an estimable parametric function depending on continuous factors as the year, day within the year etc. Applying then appropriate linear operators, for example, the derivative, or integral, to these parametric functions, we get new estimable parametric functions that answer the questions. This approach has been introduced in Möls and Starast (2000).

2. Parametric functions and their confidence limits

We consider statistical models of the form

$$\mathcal{Y}(X) = f_1(X_1, \ldots, X_N)\beta_1 + \ldots + f_k(X_1, \ldots, X_N)\beta_k + \varepsilon \qquad (2.1)$$

where $\mathcal{Y}(X)$ is the dependent scalar variable, $X_1, \ldots, X_N$ are regressors, $\beta_1, \ldots, \beta_k$ are model parameters, $f_1, \ldots, f_k$ are scalar functions of regressors and $\varepsilon \sim N(0, \sigma^2)$ is the error term. Denoting

$$f(X) = (f_1(X_1, \ldots, X_N), \ldots, f_k(X_1, \ldots, X_N)) \qquad (2.2)$$

where $X = (X_1, \ldots, X_N)$, and

$$\beta^T = (\beta_1, \ldots, \beta_k),$$

the formula (2.1) can be written as

$$\mathcal{Y}(X) = f(X)\beta + \varepsilon. \qquad (2.3)$$

Denote also

$$Y(X) = f(X)\beta, \qquad (2.4)$$

so that $Y(X) = \mathbf{E}(\mathcal{Y}(X))$ where $\mathbf{E}$ is the expectation. Note that for each possible value of X, $Y(X)$ is a scalar and therefore the product in the right side of formula (2.4) defines for each possible value of X a linear combination of model parameters. Linear combinations $L\beta$ of model parameters β, where L is a row vector of constants, are called in this paper *parametric functions*. In applications the parameters β_i are unknown and must be estimated from empirical data. Denote by

$$\mathbf{x} = \begin{pmatrix} x_{11} & \ldots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nN} \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

the matrix of design points (regressor values used in an experiment) and the vector of corresponding values (responses) of the dependent variable Y.

Introduce, further, the design matrix

$$F(\mathbf{x}) = \begin{pmatrix} f_1(x_{11} \dots x_{1N}) & \dots & f_k(x_{11} \dots x_{1N}) \\ \vdots & \ddots & \vdots \\ f_1(x_{n1} \dots x_{nN}) & \dots & f_k(x_{n1} \dots x_{nN}) \end{pmatrix}.$$

The common least squares estimate for β is then (Rao, 1965)

$$\hat{\beta} = \left(F^T(\mathbf{x})F(\mathbf{x})\right)^{-} F^T(\mathbf{x})\mathbf{y}, \tag{2.5}$$

where A^- denotes a generalized inverse of A.

For simplification, let us call the following equation

$$\hat{Y}(X) = f(X)\hat{\beta} \tag{2.6}$$

an *estimation of the model* ($f(X)$ is defined in (2.2)). Note that $\hat{Y}(X)$ is a scalar here and therefore the formula (2.6) defines for each value of X a least squares estimate for the parametric function (2.4).

In general, the estimate given by (2.6) depends on the selection of the generalized inverse in (2.5). To avoid complications on this basis, suppose that for each value of X, the parametric function (2.4) is estimable. This means, that $f(X)$ is assumed to be for all values of X a linear combination of rows of $F(\mathbf{x})$. Formally this means that for each X there exists a constant vector $\kappa(X)$ so that $f(X) = \kappa(X)F(\mathbf{x})$.

Let $1-\alpha$ be the confidence level, then the corresponding confidence limits for an arbitrary estimable parametric function $L\beta$ are given by equations

$$\begin{cases} \underline{L\beta} &= L\hat{\beta} - q(L)st_{\alpha/2} \\ \overline{L\beta} &= L\hat{\beta} + q(L)st_{\alpha/2} \end{cases} \tag{2.7}$$

where $t_{\alpha/2}$ stands for the $\alpha/2$ complementary quantile of the t_{n-r} -distribution ($r = rankF(\mathbf{x})$) and $q(L)$ and s are defined by the formulas

$$q(L) = \sqrt{L(F^T(\mathbf{x})F(\mathbf{x}))^{-}L^T}$$

and

$$s = \sqrt{\frac{\mathbf{y}^T(\mathbf{1}_n - F(\mathbf{x})(F^T(\mathbf{x})F(\mathbf{x}))^{-}F^T(\mathbf{x}))\mathbf{y}}{n-r}}. \tag{2.8}$$

Let us consider the covariance model (2.3) and let G be some linear operator which can be applied on the elements of vector $f(X)$, $G(f(X)) = (G(f_1(X)), \dots, G(f_k(X)))$. Then due to the linearity of G we can apply it on the estimable parametric function $f(X)\beta$.

Lemma 2.1. *A linear transformation of an estimable parametric function $f(X)\beta$ is also an estimable parametric function.*

Proof. According to the linearity of G,

$$G(f(X)) = G(\kappa(X)F(\mathbf{x})) = G(\kappa(X))F(\mathbf{x}) = \kappa^*(X)F(\mathbf{x}).$$

□

The confidence limits corresponding to the confidence level $1-\alpha$ are given for a parametric function $G(f(X)\beta)$ by equations

$$\begin{cases} \underline{(Gf(X))\beta} &= (Gf(X))\hat{\beta} - q(X)st_{\alpha/2}, \\ \overline{(Gf(X))\beta} &= (Gf(X))\hat{\beta} + q(X)st_{\alpha/2}, \end{cases} \tag{2.9}$$

where $t_{\alpha/2}$ stands for the $\alpha/2$-complementary quantile of the t_{n-r} -distribution ($r = rankF(\mathbf{x})$), $\hat{\beta}$ and s are given by equations (2.5) and (2.8) respectively, and $q(X)$ is defined by the formula

$$q(X) = \sqrt{G(f(X))(F^T(\mathbf{x})F(\mathbf{x}))^-[G(f(X))]^T}.$$

3. Estimating the derivative of a dependent variable

To obtain the estimate of the derivative $\frac{\partial Y(X)}{\partial X_i}$, and the appropriate confidence limits, we calculate derivative by X_i of the estimate (2.6):

$$\frac{\partial \hat{Y}(X)}{\partial X_i} = \frac{\partial}{\partial X_i} f_1(X_1,\ldots,X_N)\hat{\beta}_1 + \ldots + \frac{\partial}{\partial X_i} f_k(X_1,\ldots,X_N)\hat{\beta}_k.$$

Look at this operation as an application of the differentiation operator D_{X_i}. As a result, we obtain an estimate of a new parametric function estimating the derivative of $Y(X)$ for the given X value:

$$D_{X_i}\hat{Y}(X) = D_{X_i}f(X)\hat{\beta} = \frac{\partial f_1(X_1,\ldots,X_N)}{\partial X_i}\hat{\beta}_1 + \ldots + \frac{\partial f_k(X_1,\ldots,X_N)}{\partial X_i}\hat{\beta}_k.$$

The confidence limits for this parametric function follow from (2.9):

$$\begin{cases} \underline{D_{X_i}Y(X)} &= D_{X_i}\hat{Y}(X) - q(X)st_{\alpha/2} \\ \overline{D_{X_i}Y(X)} &= D_{X_i}\hat{Y}(X) + q(X)st_{\alpha/2}, \end{cases}$$

where

$$q(X) = \sqrt{D_{X_i}f(X)(F^T(\mathbf{x})F(\mathbf{x}))^-[D_{X_i}f(X)]^T}.$$

The obtained confidence limits are related to the variation speed of $\mathcal{Y}(X)$ in direction of X_i.

3.1. Example. In this example we demonstrate the use of derivative in estimating the periods of significant changes of iron content in Lake Peipsi. The used data set is a part of the Lake Peipsi database compiled by the Institute of Zoology and Botany of the Estonian Agricultural University (Möls and Starast, 2000). The dependent variable, denoted by R, is binary logarithm of the content of iron Fe in the water $[mg\ l^{-1}]$. Regressors (factors) are the geographical coordinates (longitude and latitude) of the observation site, year of observation and the day within the observation year. Variables that express influence of the year, the day within a year and the geographical position to the dependent variable, are presented in Table 1. Arguments $X_1, \ldots, X_4$ are the transformed year $a = (year - 1920)/10$, the transformed day within a year $t = (number\ of\ a\ day\ within\ a\ year)/365$, the transformed latitude $\xi = (latitude - 57.83)/1.18$ and the transformed longitude $\psi = (longitude - 26.93)/1.24$. In Table 1, $f(x|\mu;\sigma)$ denotes the density function of the normal distribution with mean μ and standard deviation σ, and the definitions of argument functions $f_i(a, t, \xi, \psi)$ are presented in abridged form f_i. The estimate of the model which describes the variation of variable R is given as follows:

$$\hat{R}(a, t, \xi, \psi) = (f_1(a, t, \xi, \psi), \ldots, f_{68}(a, t, \xi, \psi))\hat{\beta}. \tag{3.1}$$

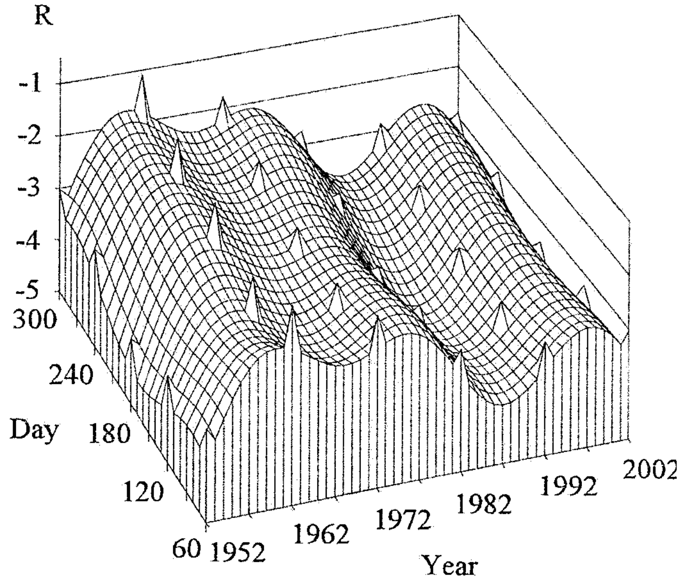

FIGURE 1. Estimate of binary logarithm R of the content of iron Fe in the water $[mg\ l^{-1}]$ at the geographical point O, with 95% confidence limits.

TABLE 1. Terms of the linear model which describes logarithmic content of iron in water (R) of Lake Peipsi. Notation: a - transformed year, t - transformed number of a day within a year, ξ - transformed latitude, ψ - transformed longitude, $f(x \mid \mu; \sigma)$ - density function of normal distribution.

$f_1 = f(a \mid 4.0; 0.95)$	$f_2 = f(a \mid 5.6; 0.95)$
$f_3 = f(a \mid 7.7; 0.95)$	$f_4 = t^2(1-t)^4$
$f_5 = t^3(1-t)^3$	$f_6 = t^4(1-t)^2$
$f_7 = \psi^2(1-\psi)^4$	$f_8 = \psi^3(1-\psi)^3$
$f_9 = \psi^4(1-\psi)^2$	$f_{10} = \xi^2(1-\xi)^4$
$f_{11} = \xi^3(1-\xi)^3$	$f_{12} = \xi^4(1-\xi)^2$
$f_{13} = f(a \mid 4.0; 0.95)\ f(a \mid 5.6; 0.95)$	$f_{14} = f(a \mid 5.6; 0.95)\ f(a \mid 7.7; 0.95)$
$f_{15} = f(a \mid 4.0; 0.95)\ \xi^2(1-\xi)^4$	$f_{16} = f(a \mid 5.6; 0.95)\ \xi^2(1-\xi)^4$
$f_{17} = f(a \mid 7.7; 0.95)\ \xi^2(1-\xi)^4$	$f_{18} = f(a \mid 4.0; 0.95)\ \xi^3(1-\xi)^3$
$f_{19} = f(a \mid 5.6; 0.95)\ \xi^3(1-\xi)^3$	$f_{20} = f(a \mid 7.7; 0.95)\ \xi^3(1-\xi)^3$
$f_{21} = f(a \mid 4.0; 0.95)\ \xi^4(1-\xi)^2$	$f_{22} = f(a \mid 5.6; 0.95)\ \xi^4(1-\xi)^2$
$f_{23} = f(a \mid 7.7; 0.95)\ \xi^4(1-\xi)^2$	$f_{24} = f(a \mid 4.0; 0.95)\ \psi^2(1-\psi)^4$
$f_{25} = f(a \mid 5.6; 0.95)\ \psi^2(1-\psi)^4$	$f_{26} = f(a \mid 7.7; 0.95)\ \psi^2(1-\psi)^4$
$f_{27} = f(a \mid 4.0; 0.95)\ \psi^3(1-\psi)^3$	$f_{28} = f(a \mid 5.6; 0.95)\ \psi^3(1-\psi)^3$
$f_{29} = f(a \mid 7.7; 0.95)\ \psi^3(1-\psi)^3$	$f_{30} = f(a \mid 4.0; 0.95)\ \psi^4(1-\psi)^2$
$f_{31} = f(a \mid 5.6; 0.95)\ \psi^4(1-\psi)^2$	$f_{32} = f(a \mid 7.7; 0.95)\ \psi^4(1-\psi)^2$
$f_{33} = \xi^2(1-\xi)^4\ \psi^2(1-\psi)^4$	$f_{34} = \xi^3(1-\xi)^3\ \psi^2(1-\psi)^4$
$f_{35} = \xi^4(1-\xi)^2\ \psi^2(1-\psi)^4$	$f_{36} = \xi^2(1-\xi)^4\ \psi^3(1-\psi)^3$
$f_{37} = \xi^3(1-\xi)^3\ \psi^3(1-\psi)^3$	$f_{38} = \xi^4(1-\xi)^2\ \psi^3(1-\psi)^3$
$f_{39} = \xi^2(1-\xi)^4\ \psi^4(1-\psi)^2$	$f_{40} = \xi^3(1-\xi)^3\ \psi^4(1-\psi)^2$
$f_{41} = \xi^4(1-\xi)^2\ \psi^4(1-\psi)^2$	$f_{42} = f(a \mid 4.0; 0.95)\ t^2(1-t)^4$
$f_{43} = f(a \mid 5.6; 0.95)\ t^2(1-t)^4$	$f_{44} = f(a \mid 7.7; 0.95)\ t^2(1-t)^4$
$f_{45} = f(a \mid 4.0; 0.95)\ t^3(1-t)^3$	$f_{46} = f(a \mid 5.6; 0.95)\ t^3(1-t)^3$
$f_{47} = f(a \mid 7.7; 0.95)\ t^3(1-t)^3$	$f_{48} = f(a \mid 4.0; 0.95)\ t^4(1-t)^2$
$f_{49} = f(a \mid 5.6; 0.95)\ t^4(1-t)^2$	$f_{50} = f(a \mid 7.7; 0.95)\ t^4(1-t)^2$
$f_{51} = \xi^2(1-\xi)^4\ t^2(1-t)^4$	$f_{52} = \xi^2(1-\xi)^4\ t^3(1-t)^3$
$f_{53} = \xi^2(1-\xi)^4\ t^4(1-t)^2$	$f_{54} = \xi^3(1-\xi)^3\ t^2(1-t)^4$
$f_{55} = \xi^3(1-\xi)^3\ t^3(1-t)^3$	$f_{56} = \xi^3(1-\xi)^3\ t^4(1-t)^2$
$f_{57} = \xi^4(1-\xi)^2\ t^2(1-t)^4$	$f_{58} = \xi^4(1-\xi)^2\ t^3(1-t)^3$
$f_{59} = \xi^4(1-\xi)^2\ t^4(1-t)^2$	$f_{60} = \psi^2(1-\psi)^4\ t^2(1-t)^4$
$f_{61} = \psi^2(1-\psi)^4\ t^3(1-t)^3$	$f_{62} = \psi^2(1-\psi)^4\ t^4(1-t)^2$
$f_{63} = \psi^3(1-\psi)^3\ t^2(1-t)^4$	$f_{64} = \psi^3(1-\psi)^3\ t^3(1-t)^3$
$f_{65} = \psi^3(1-\psi)^3\ t^4(1-t)^2$	$f_{66} = \psi^4(1-\psi)^2\ t^2(1-t)^4$
$f_{67} = \psi^4(1-\psi)^2\ t^3(1-t)^3$	$f_{68} = \psi^4(1-\psi)^2\ t^4(1-t)^2$

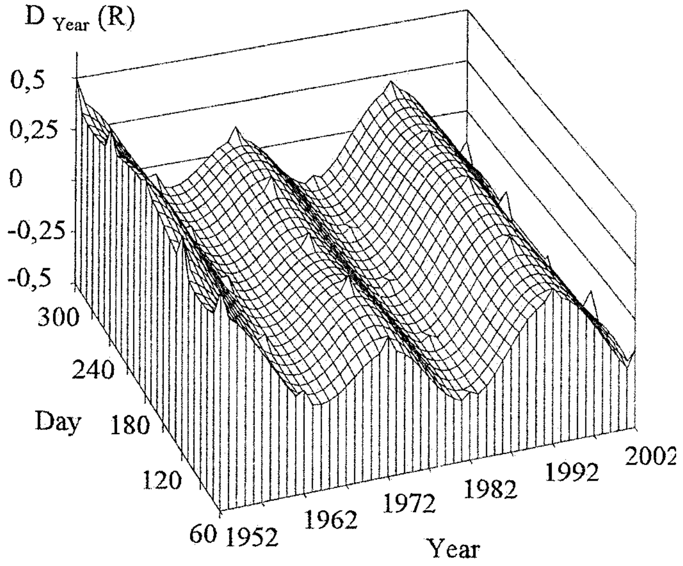

FIGURE 2. Estimate of derivative of logarithmic content R of iron, with its 95% confidence limits (at the geographical point O).

To display the values of $\hat{R}$ predicted by model, we fix an arbitrary geographical reference point in the lake. In the current example we use a point of $58°51'11''$ latitude and $26°58'8''$ longitude and mark it with symbol O. The prediction and confidence limits for variable R are presented in Figure 1. In 3-dimensional figures the upper confidence limits are presented by "nails", the lower limits have the same length underneath the surface of the diagram.

To estimate the speed of annual change of variable R, we have to calculate the derivative of the estimate of the model (3.1) by the argument a and divide it by 10, since $1a = 10$ years:

$$D_{Year}\hat{R}(a,t,\xi,\psi) = \frac{1}{10}D_a\hat{R}(a,t,\xi,\psi),$$

where

$$D_a\hat{R}(a,t,\xi,\psi) = \left(\frac{\partial}{\partial a}f_1(a,t,\xi,\psi),\dots,\frac{\partial}{\partial a}f_{68}(a,t,\xi,\psi)\right)\hat{\beta}.$$

Figure 2 presents the changing speed of variable R and its 0.95 confidence limits at the geographical point O of the lake. Figures 3 and 4 present respectively the cross-sections of Figures 1 and 2 by the 300^{th} day. From Figure 4 it is easy to find the time intervals where the variation speed of iron content differs from zero with the confidence probability 0.95.

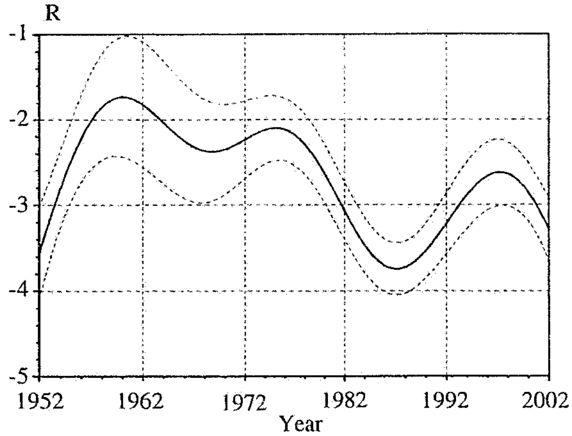

FIGURE 3. Estimate of logarithmic content R of iron, with its 95% confidence limits by the 300^{th} day (at the geographical point O).

4. Estimating the integral of dependent variable

To estimate the integral $\int_{X_{il}}^{X_{iu}} Y(X)dX_i$ and to find its appropriate confidence limits, we integrate (2.6) by X_i over the interval $[X_{il}, X_{iu}]$:

$$\int_{X_{il}}^{X_{iu}} \hat{Y}(X)dX_i = \int_{X_{il}}^{X_{iu}} f_1(X_1, \ldots, X_N)\hat{\beta}_1 dX_i + \ldots$$
$$+ \int_{X_{il}}^{X_{iu}} f_k(X_1, \ldots, X_N)\hat{\beta}_k dX_i. \quad (4.1)$$

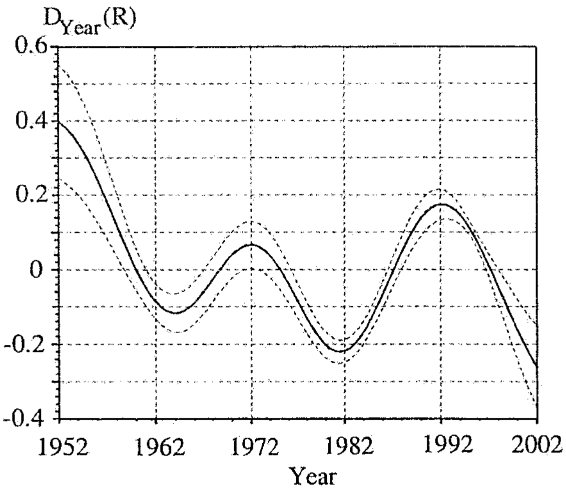

FIGURE 4. Estimate of the derivative of logarithmic content R of iron, with its 95% confidence limits by the 300^{th} day of the observation years (at the geographical point O).

Look at this operation as applying an integral operator $I_{X_i(X_{il},X_{iu})}$. As a result, we obtain an estimate of a parametric function:

$$I_{X_i(X_{il},X_{iu})}\hat{Y}(X) = I_{X_i(X_{il},X_{iu})}f(X)\hat{\beta} = \sum_{j=1}^{k}\int_{X_{il}}^{X_{iu}} f_j(X_1,\dots,X_N)dX_i\hat{\beta}_j.$$

This estimate depends on arguments $X_1,\dots,X_{i-1},X_{i+1},\dots,X_N$. The confidence limits for the related parametric function follow from the formula (2.9):

$$\begin{cases} \underline{I_{X_i(X_{il},X_{iu})}Y(X)} &= I_{X_i(X_{il},X_{iu})}\hat{Y}(X) - q(X)st_{\alpha/2} \\ \overline{I_{X_i(X_{il},X_{iu})}Y(X)} &= I_{X_i(X_{il},X_{iu})}\hat{Y}(X) + q(X)st_{\alpha/2}, \end{cases}$$

where

$$q(X) = \sqrt{I_{X_i(X_{il},X_{iu})}f(X)(F^T(\mathbf{x})F(\mathbf{x}))^{-}[I_{X_i(X_{il},X_{iu})}f(X)]^T}.$$

4.1. Example. In the following example we estimate amount of total nitrogen (denoted by K and measured in tons) in the 1-meter surface layer of Lake Peipsi. We calculate K from the following Riemann integral:

$$K(a,t) = \int_{NL}\int_{EL}\int_{D} N(a,t,\vartheta,\varphi,s)\,\alpha_{EL}\,\alpha_{NL}\,ds\,d\varphi\,d\vartheta, \qquad (4.2)$$

where NL is the integration interval for the latitude ϑ, EL is the integration interval for the longitude φ, D is the integration interval for the depth s (in current example $D = [0,1]$ and we assume that in this 1-meter deep surface layer the content of nitrogen is constant), α_{EL} and α_{NL} are the transformation coefficients from degree scale to meter scale (in the region of Lake Peipsi, $\alpha_{EL} \approx 57800\ m/^o$, $\alpha_{NL} \approx 111250\ m/^o$), and $N(a,t,\vartheta,\varphi,s)$ $[mg\ m^{-3}]$ denotes the content of nitrogen in the surface layer of Lake Peipsi on day t of the observation year a. At first we compose a model which describes distribution of nitrogen in the lake during the observation period (1985 – 2002). Model terms which express influence of a year, day within a year, and geographical position, to the nitrogen content, are presented in Table 2. Arguments$X_1,\dots,X_4$ are the transformed number of year $a = (year - 1920)/10$, the transformed day within a year $t = (number\ of\ a\ day\ within\ a\ year)/365$, the transformed latitude $\vartheta = latitude - 57.83$, and the transformed longitude $\varphi = longitude - 26.93$. In Table 2 $f(x \mid \mu;\sigma)$ denotes the density function of the normal distribution and the argument functions $f_i(a,t,\vartheta,\varphi)$ are presented in abridged form f_i. The estimate of the model which describes the content of nitrogen is given as follows:

$$\hat{N}(a,t,\vartheta,\varphi) = (f_1(a,t,\vartheta,\varphi),\dots,f_{38}(a,t,\vartheta,\varphi))\hat{\beta}.$$

The geographical integration intervals of integral (4.2) are defined by 83 rectangles. The north-south directional side of all 83 rectangles has a length of 50" latitude and the length of the east-west directional side depends on the coastline.

TABLE 2. Terms of the linear model which describes the nitrogen content of water in Lake Peipsi. Notation: a - transformed year, t - transformed number of a day within a year, ϑ - transformed latitude, φ - transformed longitude, $f(x \mid \mu; \sigma)$ - density function of normal distribution.

$f_1 = f(a \mid 6.85; 0.5)$	$f_2 = f(a \mid 7.2; 0.5)$
$f_3 = f(a \mid 7.55; 0.5)$	$f_4 = f(a \mid 7.9; 0.5)$
$f_5 = t^2(1-t)^4$	$f_6 = t^3(1-t)^3$
$f_7 = t^4(1-t)^2$	$f_8 = \vartheta$
$f_9 = \vartheta^2$	$f_{10} = \varphi$
$f_{11} = \varphi\vartheta$	$f_{12} = f(a \mid 6.85; 0.5)\ f(a \mid 7.2; 0.5)$
$f_{13} = f(a \mid 7.2; 0.5)\ f(a \mid 7.55; 0.5)$	$f_{14} = f(a \mid 7.55; 0.5)\ f(a \mid 7.9; 0.5)$
$f_{15} = f(a \mid 6.85; 0.5)\ \vartheta$	$f_{16} = f(a \mid 7.2; 0.5)\ \vartheta$
$f_{17} = f(a \mid 7.55; 0.5)\ \vartheta$	$f_{18} = f(a \mid 7.9; 0.5)\ \vartheta$
$f_{19} = f(a \mid 6.85; 0.5)\ \varphi$	$f_{20} = f(a \mid 7.2; 0.5)\ \varphi$
$f_{21} = f(a \mid 7.55; 0.5)\ \varphi$	$f_{22} = f(a \mid 7.9; 0.5)\ \varphi$
$f_{23} = f(a \mid 6.85; 0.5)\ \vartheta\varphi$	$f_{24} = f(a \mid 7.2; 0.5)\ \vartheta\varphi$
$f_{25} = f(a \mid 7.55; 0.5)\ \vartheta\varphi$	$f_{26} = f(a \mid 7.9; 0.5)\ \vartheta\varphi$
$f_{27} = f(a \mid 6.85; 0.5)\ t^2(1-t)^4$	$f_{28} = f(a \mid 7.2; 0.5)\ t^2(1-t)^4$
$f_{29} = f(a \mid 7.55; 0.5)\ t^2(1-t)^4$	$f_{30} = f(a \mid 7.9; 0.5)\ t^2(1-t)^4$
$f_{31} = f(a \mid 6.85; 0.5)\ t^3(1-t)^3$	$f_{32} = f(a \mid 7.2; 0.5)\ t^3(1-t)^3$
$f_{33} = f(a \mid 7.55; 0.5)\ t^3(1-t)^3$	$f_{34} = f(a \mid 7.9; 0.5)\ t^3(1-t)^3$
$f_{35} = f(a \mid 6.85; 0.5)\ t^4(1-t)^2$	$f_{36} = f(a \mid 7.2; 0.5)\ t^4(1-t)^2$
$f_{37} = f(a \mid 7.55; 0.5)\ t^4(1-t)^2$	$f_{38} = f(a \mid 7.9; 0.5)\ t^4(1-t)^2$

The estimate for the mass of nitrogen in the one meter deep surface layer of Lake Peipsi is given by the estimate of the parametric function as follows:

$$\hat{K}(a,t) = \left(\sum_{i=1}^{83} \int_{NL_i} \int_{EL_i} \int_0^1 f_1(a,t,\vartheta,\varphi)\, \alpha_{EL}\, \alpha_{NL}\, ds\, d\varphi\, d\vartheta, \dots, \right.$$
$$\left. \sum_{i=1}^{83} \int_{NL_i} \int_{EL_i} \int_0^1 f_{38}(a,t,\vartheta,\varphi)\, \alpha_{EL}\, \alpha_{NL}\, ds\, d\varphi\, d\vartheta \right) \hat{\beta}.$$

The estimate of variable K with its 0.95 confidence limits over the whole observation period is presented in Figure 5. Figure 6 shows the cross-section of Figure 5 by the 185^{th} day of the observation years.

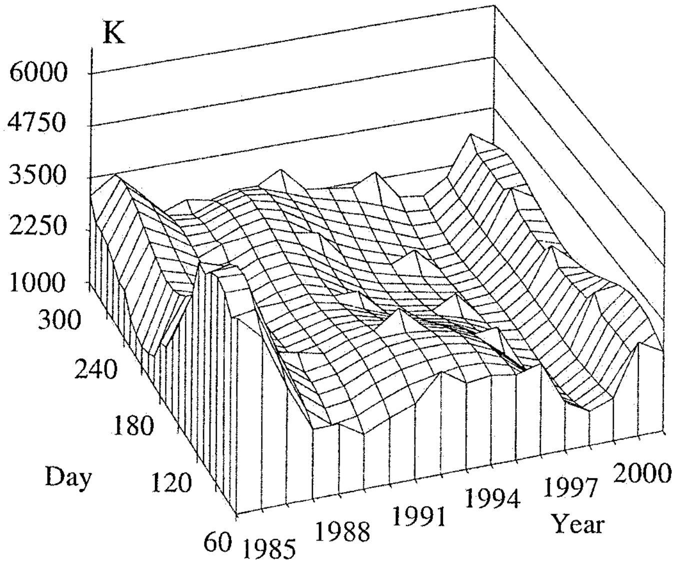

FIGURE 5. Estimate of total amount K of nitrogen (in tons) in the 1-meter deep surface water in Lake Peipsi, and its 95% confidence limits.

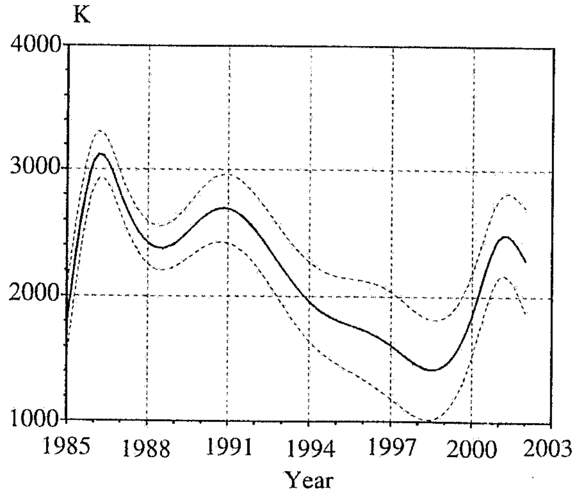

FIGURE 6. Estimate of total amount K of nitrogen (in tons) in 1-meter deep surface layer of Lake Peipsi on the 185^{th} day of the observation years, and the corresponding 95% confidence limits.

References

Rao C. R. (1968). *Linear Statistical Inference and Its Applications.* Wiley, New York.

Starast H., Milius A., Möls T. and Lindpere A. (2001). Hydrochemistry of Lake Peipsi. In: *Lake Peipsi. Meteorology, Hydrology, Hydrochemistry (Ed. T. Nõges).* Sulemees Publishers, Tartu, 97–131.

Möls T. and Starast H. (2000). Modelling of the water composition of Lake Peipsi. In: *Yearbook of the Estonian Naturalists' Society.* Vol. 79. Estonian Academy Publishers, Tallinn, 48–86. (In Estonian)

Institute of Mathematical Statistics, University of Tartu, J. Liivi 2,, 50409 Tartu, Estonia

E-mail address: tonu@zbi.ee

E-mail address: tuulm@ut.ee