Statistical Analysis of Spatial and Temporal Dynamics of a Bacterial Layer in an Aquatic Ecosystem

J.GERSS, S.SCHACH, W.URFER
Department of Statistics, University of Dortmund

14th June 1999

Abstract

Do active vertical mass movements occur within a population of phototropic bacteria in the meromictic Lake Cadagno? An experiment was conducted *in vivo* to record vertical profiles of the parameters turbidity and temperature in a spatial resolution of 30cm repeatedly over time. After eliminating the temporal dependencies within both the space-time data of turbidity measurements and temperatures, the respective spatial correlation structure can be estimated. Spatial prediction (Kriging) then offers a tool to enhance the observed spatial resolution of both processes.
By means of the (temporally repeated) turbidity profiles the vertical position of the bacterial layer can be estimated at each time point. Obviously its vertical displacements in course of the observational time occur not only due to active bacterial swimming; additionally the bacteria are dragged along passively by internal waves in the lake. Eliminating this latter disturbing effect the estimated *temperature* (instead of depth) at the bounds of the layer in course of time allows to draw conclusions on the *active* component of bacterial movements. Such phenomena can be found especially at the lower bound of the bacterial layer with amplitudes up to more than 30cm.

Key Words: Active Bacterial Movement, Space-Time Process, Variogram, Ordinary Kriging, Spatial Resolution

1. Introduction

In the meromictic Alpine Lake Cadagno[1], because of specific physical and chemical conditions a stable population of phototropic sulfur bacteria has colonized. During the summer months it shapes a distinct horizontal layer in about 11m depth (see e.g. Egli(1997)). Laboratory experiments with the predominant species *Chromatium okenii* showed motorial reactions of these microorganisms on external stimuli: Isolated bacteria moved actively depending on the surrounding light intensity (Vaituzis/Doetsch(1969)). Moreover "chemotactically" caused movements are expected due to the bacteria's need for sulfid. Those phenomena taking place on a microscopic scale of *single bacteria* are supposed to add up to synchronic mass movements of the *whole bacterial layer*, that can be registered with *macroscopic* devices. The present paper deals with the question, whether vertical displacements of the bacterial layer in Lake Cadagno occur due to active bacterial movements. Especially the *active* origin of the investigated movements is to be emphasized here, as displacements of the layer additionally show a *passive* component. This effect is caused by large internal waves in the lake that drag along the bacteria passively. It has to be "subtracted" from the observed total vertical displacements of the layer so that the remaining part can be awarded to an active source entirely.

During the past ten years several research projects have been conducted by the Institute of Plant Biology (University of Zürich) in order to investigate vertical movements of the bacterial layer in Lake Cadagno. Data collection of the subsequent projects has been improved due to more sophisticated technical equipment every time, the latest of these succeeding to localize the bacterial layer in high spatial and temporal resolution (see Egli(1997)). Egli et al.(1998) describe the methods of data collection and the results of an exploratory analysis of their extensive experiments. Dynamic changes in the form and structure of the bacterial layer can be identified, which are independent of the physical displacement of water masses. A summary of the current state of biological research concerning this problem is given in Peduzzi et al.(1998).

[1] Lake Cadagno is situated in the southern part of the Swiss Alps in Canton Tessin

Notably until now all results of biological research have been derived by means of exploratory data analysis carried out by biologists. Additionally to exploratory approaches much more efficient methods of data analysis are provided by the statistical theory. By means of inductive analyses based on stochastic model assumptions usually more substantial insight can be gathered from observed data. In the present case spatiotemporal statistical methods appear to be promising. These methods have been developed to analyze environmental phenomena varying in space and time. Tools are provided to model variability in both domains and to interpolate observed data. Moreover empirical results derived from a certain project yield recommendations for future similar projects, in particular concerning the optimal localization of measurement sites.

Data from the above mentioned research project Egli(1997) will be analyzed in the present paper by means of spatiotemporal methods. Concretely, experiment No.6ii, conducted August $26^{\underline{\text{th}}}$ to $29^{\underline{\text{th}}}$ 1996, yielded reliable measurements and no technical problems occurred. Fitting a statistical model to the corresponding data a certain kind of "temporal homogeneity" is required as an elementary condition. This assumption is found not to hold for the complete observational time *on the whole*: Plotting the data an abrupt structural change is noticed on August $28^{\underline{\text{th}}}$, $8^{\underline{00}}$, due to a change in weather. The observational time should therefore be divided into two subperiods in the analysis step. In the present paper a statistical model is fit to only the first of these subperiods, concretely this is the period from August $26^{\underline{\text{th}}}$, $16^{\underline{00}}$ until August $28^{\underline{\text{th}}}$, $8^{\underline{00}}$.
The collection of data was realized with the following arrangement of devices: 9 sensors were placed vertically in distances of 30cm in a depth to cover the bacterial layer. Each sensor was equipped to record the turbidity [FTU[2]] and temperature [°C] of the surrounding water. Data was sampled in 1 min intervals.

[2]FTU = Formazine Turbidity Units; in the statistical analysis the measured turbidity values were transformed on a log-scale

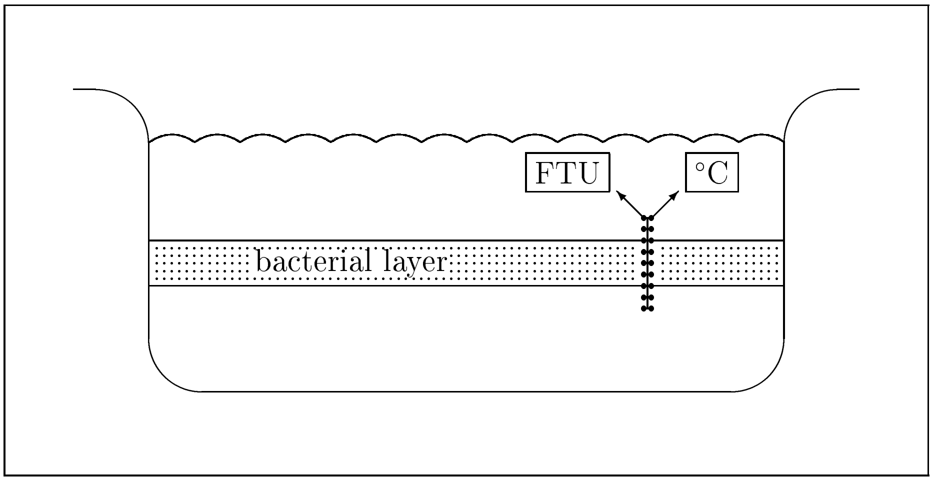

Figure 1: Scheme of the Lake with Vertical Positioning of the Sensors

Thus the vertical distribution of the two parameters turbidity and temperature is observed in a spatial resolution of 30cm. Aim of the statistical analysis is to enhance this resolution by interpolating the measured values of turbidity and temperature into the space *between* the sensors. This improves the facilities of vertically localizing the bacterial layer. Knowing the exact depths of its upper and lower bound the respective temperature of the surrounding water can be predicted in the next step. Thereby information is available on the span of temperature corresponding to the bacterial layer in its vertical expansion. Temporal changes in this span of temperature finally allow to draw conclusions on the *active* component of bacterial movements in vertical direction.

2. The Data

According to the data collection the measured values of turbidity and temperature correspond to an array with a two-dimensional index. Let $Z_t(s)$ represent a measurement of either turbidity or temperature in depth s at time t. The whole data set thereby is represented by the expression $[Z_t(s_i)]_{\substack{i=1,\dots,n \\ t=1,\dots,T}}$: $s_1, \dots, s_n$ are the n depths, where the sensors are placed one beneath the other: 11.9m, 12.2m, ..., 14.3m; $\{t = 1, \ldots, T\}$ is the set of time points of the collected data. Measurements of turbidity and temperature are at hand in each of the n depths at each of the T points of time.

As a basis for the later proceeding note that this data set can be viewed either in "spatial" or in "temporal direction" primarily:

- On the one hand

$$[Z_t(s_i)]_{\substack{i=1,\ldots,n \\ t=1,\ldots,T}} = \left\{ \begin{array}{ccc} Z_1(s_1) & \cdots & Z_T(s_1) \\ \vdots & & \vdots \\ Z_1(s_n) & \cdots & Z_T(s_n) \end{array} \right\}$$

$$= \left\{ [Z_1(s_i)]_{i=1,\ldots,n}, \quad [Z_2(s_i)]_{i=1,\ldots,n}, \ldots, [Z_T(s_i)]_{i=1,\ldots,n} \right\}$$

  represents a set of temporally consecutive spatial processes ("spatial point of view");

- on the other hand, the same set of values

$$[Z_t(s_i)]_{\substack{i=1,\ldots,n \\ t=1,\ldots,T}} = \left\{ \begin{array}{c} [Z_t(s_1)]_{t=1,\ldots,T} \\ [Z_t(s_2)]_{t=1,\ldots,T} \\ \vdots \\ [Z_t(s_n)]_{t=1,\ldots,T} \end{array} \right\}$$

  represents a set of spatial located time series ("temporal point of view").

In statistical terminology, observations of a random variable are taken at spatially discrete arranged sample locations $s_1, \ldots, s_n$ repeatedly over time. This data situation is to be analyzed by methods within the framework of *space-time processes*[3]. Space-time models represent the appropriate tool to interpolate the collected data into the space *between* sample locations (spatial prediction): What is the value of the measured random variable in depths, where no sensors have been placed? Strictly speaking, here *two different* random variables are concerned: turbidity

[3]Compared to the general situation, the present data structure shows several simplifications: The localization of the sensors is done on a one-dimensional spatial scale (depth), the coordinates of the sample locations are equidistant and do not change over time.

and temperature. So two different space-time processes will be analyzed. In the following, each of them is analyzed separately but in the same way. A common space-time model was found to match both of the observed processes.

In general, a model fitted to space-time data should have one important property. It has to be able to explain existing spatial and temporal dependencies, a specific aspect of space-time data due to the fact that spatially or temporally linked measurements are related to each other. Judging a proposed model for an observed data set is done by investigating its ability to explain the special nature of spatial and temporal correlations sufficiently. The strategy of the model presented here is then as follows: At first temporal dependencies within the data are eliminated in order to be able to estimate the spatial dependencies unbiasedly. Knowing about the spatial correlation structure subsequently the problem of spatial prediction can be solved in an optimal way (see e.g. Cressie(1993)).

3. A Spatiotemporal Model

The following space-time model assumes the spatial and the temporal dependence structure to be represented in form of two separate components that are linked additively. Specifically, the temporal structure of a given set of space-time data is modelled by means of an ARIMA[4]-approach, that proved to be a convenient tool in many time series applications (see e.g. Schlittgen/Streitberg(1997)). After eliminating the temporal correlations within the data set, the remaining spatial structure is assumed to be constant over time and is analyzed by means of common Geostatistical procedures.

$$(1-B)^d Z_t(\boldsymbol{s}) = \alpha_1(1-B)^d Z_{t-1}(\boldsymbol{s}) + \ldots + \alpha_p(1-B)^d Z_{t-p}(\boldsymbol{s}) + \delta_t(\boldsymbol{s}) \qquad (1)$$

where $[Z_t(\boldsymbol{s})]$ is the observed space-time process,

B denotes the temporal Backshift-operator, i.e. $BZ_t(\boldsymbol{s}) = Z_{t-1}(\boldsymbol{s})$ and

$\delta_t(\boldsymbol{s})$ represents a component of *purely spatial* correlations and measurement error.

[4] AutoRegressive Moving-Average process

By means of a d-fold application of the difference filter $(1-B)$ temporal trend is eliminated within the set of data. The remaining temporal structure is assumed to be captured by an autoregressive relationship of the random variables $Z_t(\boldsymbol{s})$.

Note that the autoregressive coefficients $\alpha_1, \ldots, \alpha_p$ do not carry a spatial index. This spatial invariance of the temporal structure represents a crucial assumption for the later proceeding. At the stage of model specification it arises the big problem of estimating *common* coefficients of a set of (correlated) time series. A mathematically justified solution of this problem can be found by generalizing the Least-Squares approach used to estimate AR-parameters of an observed univariate time series (see Schlittgen/Streitberg(1997) pg. 165ff). In the present case this procedure proved not to be necessary as will be seen below.

An initial analysis of the present data set showed far reaching simplifications of the general approach (1) to be appropriate. The autoregressive temporal structure was found to show specific characteristics in case of the observed turbidity process as well as the temperature measurements. Concretely the temporal dependencies could be extracted from the data $[Z_t(s)]$ applying the temporal difference filter $(1-B)$ only once. Repeated application of the difference filter as well as the additional autoregressive relationship in model (1) proved not to be necessary in order to characterize/eliminate temporal correlations. Computing

$$(1-B)\Big[Z_t(s_i)\Big]_{\substack{i\,=\,1,\ldots,n \\ t\,=\,1,\ldots,T}} = (1-B)\left\{\begin{array}{c} [Z_t(s_1)]_{t=1,\ldots,T} \\ [Z_t(s_2)]_{t=1,\ldots,T} \\ \vdots \\ [Z_t(s_n)]_{t=1,\ldots,T} \end{array}\right\}$$

$$= \left\{\begin{array}{c} [Z_t(s_1) - Z_{t-1}(s_1)]_{t=1,\ldots,T} \\ [Z_t(s_2) - Z_{t-1}(s_2)]_{t=1,\ldots,T} \\ \vdots \\ [Z_t(s_n) - Z_{t-1}(s_n)]_{t=1,\ldots,T} \end{array}\right\}$$

(where $Z_0(s_i)$ is defined to be equal to zero for all $i = 1, \ldots, n$) and plotting the autocorrelation function of each of the time series

$$\Big[\delta_t(s_i)\Big]_{t=1,\ldots,T} := \Big[Z_t(s_i) - Z_{t-1}(s_i)\Big]_{t=1,\ldots,T} \qquad (i = 1, \ldots, n)$$

showed low dependencies within these sets of values.

Thus model (1) can be simplified in case of the present data set. Its following special case was found to catch the characteristics of the observed data adequately:

$$\begin{aligned} & (1-B)Z_t(s) = \delta_t(s) \\ \Leftrightarrow \quad & \boxed{Z_t(s) = (1-B)^{-1}\delta_t(s)} \quad . \end{aligned}$$

$[\delta_t(s)]$ represents a version of the observed space-time process $[Z_t(s)]$, within that the temporal dependencies have been eliminated:

$$\begin{aligned} (1-B)\Big[Z_t(s_i)\Big]_{\substack{i=1,\ldots,n \\ t=1,\ldots,T}} \\ &= \Big[\delta_t(s_i)\Big]_{\substack{i=1,\ldots,n \\ t=1,\ldots,T}} \\ &= \Big\{[\delta_1(s_i)]_{i=1,\ldots,n} \quad , \quad [\delta_2(s_i)]_{i=1,\ldots,n} \quad , \quad \ldots \quad , \quad [\delta_T(s_i)]_{i=1,\ldots,n}\Big\} \end{aligned}$$

represents a set of temporally consecutive but *uncorrelated* spatial processes. All of these processes $\delta_t(\cdot)$ shall be assumed to have identical stochastical characteristics. So it seems reasonable to work under the model that the spatial processes $\delta_t(\cdot) \quad (t = 1, \ldots, T)$ represent i.i.d. repetitions of one fictitious process $\delta(\cdot)$. This spatial "mother process" itself is to be imputed a stochastical model. In many Geostatistical applications a simple additive decomposition into independent components has proved to be sensible (see Berke(1999)). Concretely the nonstochastical mean structure is added by a (zero-mean) expression representing the contribution of spatial dependencies on the process value at a certain location. A third component capturing all residual influences is assumed to represent a spatially independent term of measurement error:

$$\boxed{[\delta(s)]_{s\in D} = [\mu(s)]_{s\in D} + [\eta(s)]_{s\in D} + [\epsilon(s)]_{s\in D}} \quad ,$$

with mean function (large-scale variation) $\mu(s)$,

small-scale variation term $\eta(s)$ (containing the spatial dependencies) and

White-Noise process $\epsilon(s)$ of measurement errors.

In the present problem the mean function can be assumed to be constant:

$$\boxed{\mu(s) = \mu \qquad \forall \quad s \in D} \quad .$$

As a next step, the correlation structure within the spatial process $\delta(\cdot)$ is to be analyzed. According to the usual proceeding in Geostatistical applications this is done by estimating[5] the variogram

$$2\gamma_\delta(h) = Var\Big[\delta(s) - \delta(s-h)\Big]$$

as a function of the spatial distance h. The (estimated) variogram $2\hat{\gamma}_\delta(\cdot)$ of the (fictitious) $\delta(\cdot)$-process then also represents the common spatial correlation structure within each of the processes $\delta_t(\cdot) \quad (t = 1, \ldots, T)$.
Knowing their variogram optimal spatial prediction can be carried out within each of these processes subsequently.

4. Variogram Estimation and Model Fitting

The array $\Big[\delta_t(s_i)\Big]_{\substack{i=1,\ldots,n \\ t=1,\ldots,T}}$ represents data from a space-time process that does not contain any temporal dependencies. In such situation Sampson/Guttorp(1992) propose to estimate the (temporally constant) spatial variogram $2\gamma_\delta(\cdot)$ of this process as follows[6].

$$2\hat{\gamma}_\delta(h) = \frac{1}{|N(h)|} \sum_{N(h)} \Big[s_{ii} + s_{jj} - 2s_{ij}\Big]$$

[5]Necessary assumptions for the statistical justification of the practical estimation procedure for the variogram $2\gamma_\delta$ are the intrinsical stationarity of the underlying $\delta(\cdot)$-process as well as the isotropy of the variogram function (see e.g. Cressie(1993, pg. 60/61)).

[6]The turbidity values of sensor 3 (depth 12.5m) and the temperatures of sensor 1 (depth 11.9m) were excluded from variogram estimation. If i or j equals 3 within the turbidity process or 1 within the temperature process respectively, the term $[s_{ii} + s_{jj} - 2s_{ij}]$ could clearly be identified as outliers due to disproportionate size.

$$
\begin{aligned}
\text{with } s_{ij} &= \frac{1}{T}\sum_{t=1}^{T}\Big[\delta_t(s_i)-\overline{\delta.(s_i)}\Big]\cdot\Big[\delta_t(s_j)-\overline{\delta.(s_j)}\Big],\\
\overline{\delta.(s_i)} &= \frac{1}{T}\sum_{\tau=1}^{T}\delta_\tau(s_i),\\
N(h) &= \Big\{(s_i,s_j):|s_i-s_j|=h;\quad i,j=1,\ldots,n\Big\} \text{ and}\\
|N(h)| &= \text{ number of distinct elements } (s_i,s_j) \text{ of } N(h).
\end{aligned}
$$

Thus for values of h corresponding to multiples of the sensor distance 0.3m, a set of estimated variogram values

$$
2\hat{\gamma}_\delta(0.3) \quad , \quad 2\hat{\gamma}_\delta(0.6) \quad , \quad \ldots \quad , \quad 2\hat{\gamma}_\delta(2.1)
$$

can be obtained.

The so called Nugget effect (measurement error)[7] NE $:= \lim_{h\to 0} 2\gamma_\delta(h)$ (see Cressie(1993, pg. 59)) was possible to be estimated only for the process of turbidity values: Variations within the set of values $[\delta_t(s_n)]_{t=1,\ldots,T}$ corresponding to the lowest sensor (in depth 14.0m) can be utilized to quantify the precision of turbidity measurement,

$$
\widehat{\mathrm{NE}} = \lim_{h\to 0}\widehat{2\gamma_\delta}(h) = 2\cdot\mathrm{Var}\Big\{[\delta_t(s_n)]_{t=1,\ldots,T}\Big\}
$$

The other sensors' values vary *not only due to measurement error.* Additionally, internal waves in the lake caused systematic fluctuations in those sets of values. Such effects add to the pure measurement error process and cause the respective sensors' variability to *exceed* the value of measurement error quantitatively:

$$
2\cdot\mathrm{Var}\Big\{[\delta_t(s_i)]_{t=1,\ldots,T}\Big\} \quad \hat{=} \quad \mathrm{NE}+c_W \quad > \quad \mathrm{NE} \qquad \forall i=1,\ldots,n-1.
$$

In case of the process of temperature values such systematic fluctuations enhance the variability of *all sensors'* sets of values. Therefore within this process it was not possible to quantify the measurement error (Nugget effect) at all .

With knowledge of the variogram of a spatial process at a set of *discrete* distance values $(\widehat{\mathrm{NE}},)2\hat{\gamma}_\delta(0.3), 2\hat{\gamma}_\delta(0.6),\ldots,2\hat{\gamma}_\delta(2.1)$ only the spatial prediction cannot be

[7] "Microscale-variation" (see Cressie(1993), pg. 59) was assumed to be non-existent, so the Nugget effect is identical to the measurement error.

carried out directly. Optimal spatial prediction requires knowledge of the variogram as a *continuous* function of the spatial distance h. The set of values $(\widehat{\text{NE}},)2\hat{\gamma}_\delta(0.3), 2\hat{\gamma}_\delta(0.6), \ldots, 2\hat{\gamma}_\delta(2.1)$ has to be interpolated to arbitrarily close values of h. Technically this is done by fitting a parametric model $2\gamma_\delta(\cdot|\boldsymbol{\theta})$ to the set of discrete variogram values. Journel/Huijtbregts(1978, pg. 161-195) propose several variogram models, from which the "spherical model" (with *linear* increment at the origin) is most common in Geostatistical applications:

$$\gamma(h|\boldsymbol{\theta}) = \gamma(h|c_0, c_s, a_s) = \begin{cases} 0 & \text{if} \quad h = 0 \\ c_0 + c_s\left[\frac{3}{2}\frac{h}{a_s} - \frac{1}{2}(\frac{h}{a_s})^3\right] & \text{if} \quad 0 < h \leq a_s \\ c_0 + c_s & \text{if} \quad h \geq a_s \end{cases}$$

with parameter space $c_0 \geq 0, c_s \geq 0, a_s \geq 0$.

The estimated parameters $\hat{\boldsymbol{\theta}}$ then characterize the shape of the variogram as a continuous function of the spatial distance h. Difficulties in developing reliable estimation procedures for these parameters in particular arise from the fact that there exist correlations and different variances within the "row data" $(\widehat{\text{NE}},)2\hat{\gamma}_\delta(0.3), 2\hat{\gamma}_\delta(0.6), \ldots, 2\hat{\gamma}_\delta(2.1)$. An optimal way of fitting a (nonlinear) model $2\gamma_\delta(\cdot|\boldsymbol{\theta})$ in this case would be to apply a General-Least-Squares approach. Instead of this computer intensive procedure, Cressie(1985) suggests a compromise between optimality and simplicity by applying a Weighted-Least-Squares approach. By doing this the correlations between the set of values $(\widehat{\text{NE}},)2\hat{\gamma}_\delta(0.3), 2\hat{\gamma}_\delta(0.6), \ldots, 2\hat{\gamma}_\delta(2.1)$ are ignored. Only their different variances are included into the fitting equation, which finally leads to a weighting of the agreement of "data" and model in favour of the agreement at small distances h. Figures 2 and 3 show the result of this variogram fitting procedure in case of the observed turbidity process as well as the temperature process, respectively, computed with the statistical package S-Plus 4.0.

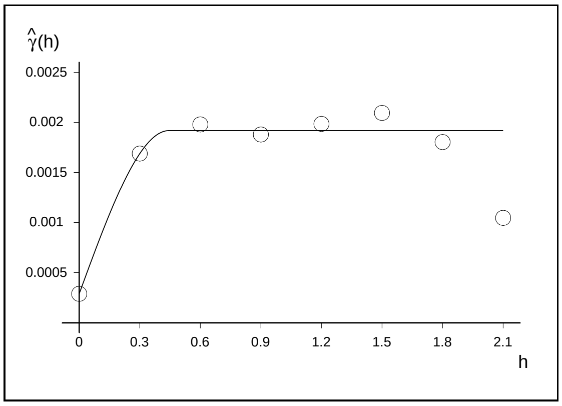

Figure 2: Variogram of the Turbidity Process

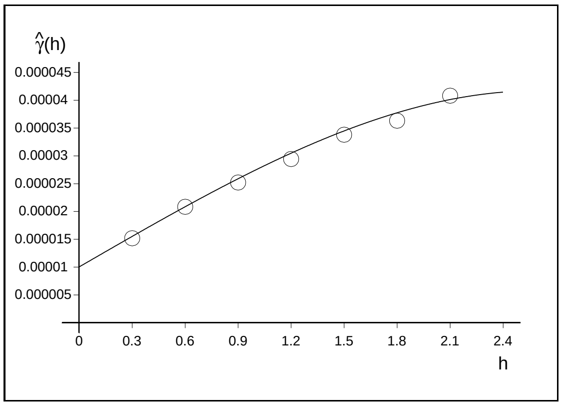

Figure 3: Variogram of the Temperature Process

As can be seen from these figures, additionally to the above mentioned continuous interpolation the fitted variogram model guarantees another aspect: It shows a smoother course than the set of values $(\widehat{\mathrm{NE}},)2\hat{\gamma}_\delta(0.3), 2\hat{\gamma}_\delta(0.6), \ldots, 2\hat{\gamma}_\delta(2.1)$, which goes along with a compensation of estimation errors in the discrete variogram estimators.

5. Spatial Prediction

Based on the fitted variogram function of a spatial process spatial prediction can be carried out. Concretely knowledge of the common variogram $2\gamma_\delta(\cdot|\hat{\boldsymbol{\theta}})$ of each of the processes

$$\Big[\delta_1(s_i)\Big]_{i=1,\ldots,n} \quad , \quad \Big[\delta_2(s_i)\Big]_{i=1,\ldots,n} \quad , \quad \ldots \quad , \quad \Big[\delta_T(s_i)\Big]_{i=1,\ldots,n}$$

allows spatial prediction within each of them. The realization of this procedure in mathematical terms mainly depends on the structure of the deterministic mean function (large-scale variation) of the underlying "mother process" $[\delta.(s)]$. As this function was assumed to be constant for all locations (depths), "Ordinary Kriging" satisfies an optimality criterion: It represents the "Estimated Best Linear Unbiased Estimator" (EBLUP, see Cressie(1988)). Cressie(1988) also presents a way of modifying the spatial prediction procedure in order to compensate measurement errors that occurred at data collection. The resulting Ordinary Kriging Equations yield (for fixed t) the prediction function $\hat{\delta}_t(s_0)^{\text{noiseless}}$ within the $\delta_t(s)$-process at an unobserved location $s_0 \notin \{s_1, \ldots, s_n\}$ (see Gerß(1998)):

$$\hat{\delta}_{\mathrm{t}}(\mathrm{s}_0)^{\text{noiseless}} = \quad \boldsymbol{\lambda}'\boldsymbol{\delta_t} \quad .$$

The vector $\boldsymbol{\lambda}$ of Kriging weights result from the variogram matrices Γ and $\boldsymbol{\Gamma_0}$ as follows:

$$\begin{aligned} \boldsymbol{\lambda} &= \Gamma^{-1}\boldsymbol{\Gamma_0} - \Gamma^{-1}\mathbf{1}(\mathbf{1}'\Gamma^{-1}\mathbf{1})^{-1}(\boldsymbol{\Gamma_0}'\Gamma^{-1}\mathbf{1} - 1) \in \mathrm{I\!R}^{\mathrm{n}} \\ \text{with} \quad \Gamma &= \Big[\gamma_\delta(|s_i - s_j|; \hat{\boldsymbol{\theta}})\Big]_{i,j=1,\ldots,n} \in \mathrm{I\!R}^{\mathrm{n\times n}} \end{aligned}$$

$$
\mathbf{\Gamma_0} = \begin{cases} \left[\gamma_\delta(|s_0 - s_i|;\hat{\boldsymbol{\theta}})\right]_{i=1,\dots,n} \in \mathrm{I\!R}^{\mathrm{n}} & \text{if} \quad s_0 \neq s_i \quad \forall i = 1, \dots, n \\[2ex] \begin{bmatrix} \gamma_\delta(|s_0 - s_1|;\hat{\boldsymbol{\theta}}) \\ \vdots \\ \gamma_\delta(|s_0 - s_{i^\star-1}|;\hat{\boldsymbol{\theta}}) \\ \lim\limits_{h\to 0} 2\hat{\gamma}_\delta(h;\hat{\boldsymbol{\theta}}) \\ \gamma_\delta(|s_0 - s_{i^\star+1}|;\hat{\boldsymbol{\theta}}) \\ \vdots \\ \gamma_\delta(|s_0 - s_n|;\hat{\boldsymbol{\theta}}) \end{bmatrix} \in \mathrm{I\!R}^{\mathrm{n}} & \text{if} \quad \exists i^\star \in \{1, \dots, n\} : s_0 = s_{i^\star} \end{cases}
$$

$$
\begin{aligned}
\mathbf{1} &= (1, \dots, 1)' \in \mathrm{I\!R}^{\mathrm{n}} \\
\boldsymbol{\delta_t} &= \left[\delta_t(s_1), \dots, \delta_t(s_n)\right]' \in \mathrm{I\!R}^{\mathrm{n}}
\end{aligned}
$$

In this way the value of the spatial process $\left[\delta_t(s_i)\right]_{i=1,\dots,n}$ (for fixed t) can be predicted in arbitrary close depths $s_j \quad (j = 1, \dots, m)$ with $m \gg n$, only limited by the capacity of the available computer. Finally a set of values is at hand which represents a noiseless version of the process $[\delta_t(\cdot)]$ in high spatial resolution. Joining all these consecutive spatial processes $[\delta_t(\cdot)]$ a highly resolved version of the observed space-time process $[\delta_.(\cdot)]$ is present:

$$
\left\{[\delta_1(s_j)]_{j=1,\dots,m} \quad , \quad [\delta_2(s_j)]_{j=1,\dots,m}, \dots, [\delta_T(s_j)]_{j=1,\dots,m}\right\} \quad = \quad \left[\delta_t(s_j)\right]_{\substack{j\,=\,1,\,\dots,\,m \\ t\,=\,1,\,\dots,\,T}} .
$$

As a next step the course back to the $Z_t(s)$-process is to be followed. Therefore a “temporal point of view“ is applied to the values of the $\delta_t(s)$-process:

$$
\left[\delta_t(s_j)\right]_{\substack{j\,=\,1,\,\dots,\,m \\ t\,=\,1,\,\dots,\,T}} = \left\{ \begin{array}{c} [\delta_t(s_1)]_{t=1,\dots,T} \\ [\delta_t(s_2)]_{t=1,\dots,T} \\ \vdots \\ [\delta_t(s_m)]_{t=1,\dots,T} \end{array} \right\} .
$$

Application of the inverse difference filter $(1 - B)^{-1}$ to each of these time series yields the corresponding values of the $Z_t(s)$-processes; the formerly eliminated temporal dependencies are brought back into the data:

$$(1-B)^{-1}\Big[\delta_t(s_j)\Big]_{\substack{j=1,\dots,m\\ t=1,\dots,T}} = \left\{\begin{array}{c} (1-B)^{-1}[\delta_t(s_1)]_{t=1,\dots,T} \\ (1-B)^{-1}[\delta_t(s_2)]_{t=1,\dots,T} \\ \vdots \\ (1-B)^{-1}[\delta_t(s_m)]_{t=1,\dots,T} \end{array}\right\} = \Big[Z_t(s_j)\Big]_{\substack{j=1,\dots,m\\ t=1,\dots,T}}$$

$$\text{with } (1-B)^{-1}\Big[\delta_t(s_j)\Big]_{t=1,\dots,T} = \sum_{u=0}^{\infty} B^u\Big[\delta_t(s_j)\Big]_{t=1,\dots,T} = \Big[\sum_{u=1}^{t}\delta_u(s_j)\Big]_{t=1,\dots,T}$$

$$(\quad \delta_u(s_j) := 0 \text{ for } u \leq 0 \quad) \, .$$

Finally a noiseless version of the observed space-time process $[Z_t(s)]$ has been derived in high spatial resolution ($m \gg n$):

$$\Big[Z_t(s_j)\Big]_{\substack{j=1,\dots,m\\ t=1,\dots,T}} = \left\{\begin{array}{c} [Z_t(s_1)]_{t=1,\dots,T} \\ [Z_t(s_2)]_{t=1,\dots,T} \\ \vdots \\ [Z_t(s_m)]_{t=1,\dots,T} \end{array}\right\}$$

$$= \Big\{[Z_1(s_j)]_{j=1,\dots,m}, \quad [Z_2(s_j)]_{j=1,\dots,m}, \dots, [Z_T(s_j)]_{j=1,\dots,m}\Big\}.$$

6. Results

The above described procedure of variogram estimation, variogram model fitting and Kriging was applied to the turbidity process on the one hand as well as to the temperature process on the other[8]. Concretely in each case the spatial resolution could be enhanced from 30cm (sensor distance) to 3.33cm. Thus the basis is set up for estimation of active vertical movements of the bacterial layer in Lake Cadagno.

[8]with the statistical software package S-Plus (version 4.0)

By means of the turbidity profile the vertical expansion of the bacterial layer can be estimated exactly at a certain time point. Localisation is done by defining its bounds to be those depths where the turbidity transgresses a value of 8 FTU[9] (see Egli(1997)); locations with turbidities below that value are regarded not to be covered by the bacterial layer. After estimating the depths of its bounds[10] for each time point those sets of values can be connected in "temporal direction". This yields a time series of the depth of the upper and the lower bound, respectively, during the whole observation period of the experiment:

$$\Big[D_{up}(t)\Big]_{t=1,\ldots,T} \quad \text{and} \quad \Big[D_{low}(t)\Big]_{t=1,\ldots,T} \quad .$$

The corresponding graphs (see Gerß(1998, pg. 89)) show vertical displacements of the bacterial layer of large amplitudes (up to 50cm). But as described above, these movements are due to active bacterial swimming only in part. The main reason for the movement are internal waves in the lake. These passive effects are eliminated by means of spatial prediction within the temperature process: What is the temperature in those depths, where upper and lower bound of the bacterial layer could be made out at a certain time point?
Let $[T_t(s)]$ represent the temperature process. Hence mathematically the values $T_t\Big[D_{up}(t)\Big]$ and $T_t\Big[D_{low}(t)\Big]$ are predicted for each $t = 1, \ldots, T$ yielding time series of the temperature at the layer's upper and lower bound, respectively.

[9] Non-zero turbidity values beyond the bacterial layer are explained by floating particles in the surrounding "clear" sea water.

[10] see Appendix for details on the applied algorithm

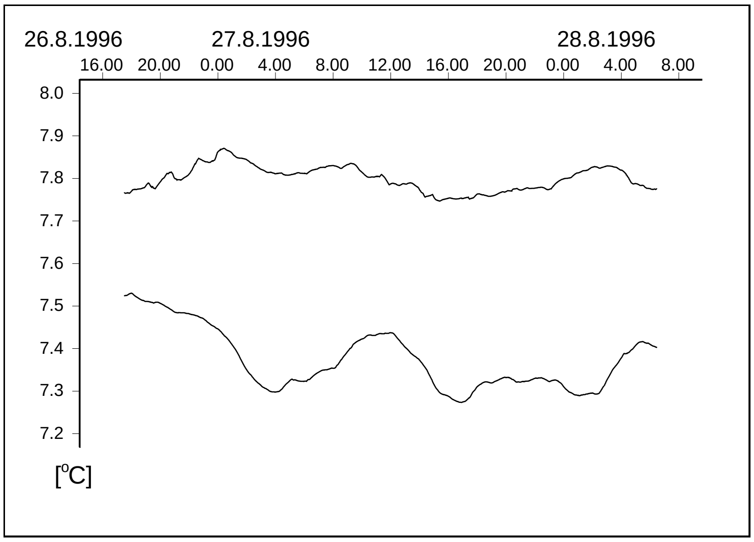

Figure 4: Temperature at the Upper and Lower Bound of the Bacterial Layer in Course of Time.
Smoothed version of the time series $\left\{T_t\left[D_{up}(t)\right]\right\}_{t=1,\ldots,T}$ and $\left\{T_t\left[D_{low}(t)\right]\right\}_{t=1,\ldots,T}$, derived from the original series by Moving-Average filtration.

Figure 4 allows drawing conclusions on active vertical movements of the bacterial layer.
Suppose the temperature at its bounds does not change substantially in the course of time. Consequently the bacteria do not withdraw from their surrounding water masses during a certain interval of time. This does *not* mean that the bacteria did not shift vertically at all. But any vertical displacements occurred synchronously to internal waves in the lake. Thus a causal connection of both phenomena would seem to be proved: Vertical shifts of the layer are realized because the bacteria are dragged along *passively* by the water masses.
On the other hand the temperature at the bounds of the bacterial layer may change during certain intervals of time. In this case its vertical displacements seem to have originated - at least partially - in *active* bacterial self-movements. *"Relatively to the water masses"* the bacteria swam up or down into warmer or colder zones of the lake.

Viewing figure 4 these latter effects can be seen. Especially at the lower bound of the layer active vertical movements can be made out: Here the bacteria pass

through a span of temperature of approximately 0.2°C periodically. This corresponds to a vertical distance of more than 30cm, that is covered within about 8 hours repeatedly. At the upper bound of the layer highly marked effects like that cannot be made out. Here the bacteria stay within water masses of about 7.8°C during the whole observational time.
A recurring pattern of active vertical movements can be found neither at the upper nor at the lower bound of the bacterial layer. Possible diurnal effects can not be derived from figure 4 because the observation time is too short (38 hours).

7. Alternative Approaches and Future Projects

One of the most important properties of Kriging as a means of spatial prediction is the additional ability of quantifying the *accuracy* of the Kriging predictor. In contrast to other means of spatial prediction the Kriging equations yield not only a point-predictor of an unobserved variable but also the Mean Squared Prediction Error (MSPE) associated with it. Further research on the present problem might aim at quantifying the precision of the estimated bacterial movements and calculating some form of confidence intervals. Finally one can carry out a statistical test of the hypothesis that the calculated phenomena are due to *random effects* exclusively. Rejection of this hypothesis would represent a statistical proof of "significant" bacterial activity.

A big disadvantage of the presently applied procedures is the fact that both observed space-time processes - turbidity and temperature - are analyzed *separately.* An obvious extension of this approach is to model a bivariate space-time process instead and *benefit from the correlations* between both individual ("component") processes[11]. Cokriging (see Cressie (1993, pg. 138ff)) then represents a way of predicting within one of the component processes using the additional information provided by the other component processes optimally. This approach has not been applied to the present data set for the following reasons: Firstly the

[11] For vector-valued stochastic processes Cressie (1993, pg. 140) mentions the concept of "cross-variograms": (Co)Variogram-functions that describe the dependence structure within as well as between individual component processes.

existing statistical software packages do not provide procedures for implementing Cokriging in a space-time setting by now. Moreover in the present case the relationship between both component processes turbidity and temperature was found to be *nonlinear*. Thereby it is not recognized by the common measures of stochastic dependence: Variogram and Covariogram only describe *linear* relationships between two random variables. The theory of Cokriging is based upon these *linear* concepts and is not prepared to include additional information of a *nonlinear* kind. In predicting one component process the additional information provided by a *nonlinearly related* second process cannot be exploited adequately by the concept of Cokriging by now.

The inclusion of weather data into the space-time model might be promising as well. Certain meteorological measurements might prove to represent significant explanatory variables for the observed phenomena of bacterial movement. The detection of such causal connections is highly interesting from a biological point of view. For example it possibly proves changing light intensity to be the source of bacterial activity (phototactical movement).

From a theoretical statistical point of view the underlying general spatiotemporal model (1) is worth further research. Future projects might aim at developing formulae for the space-time correlation structure of stochastic processes that fit in the scope of model (1). One of the most restrictive model assumptions is that of space-time separability. Spatial and temporal structure are assumed to be represented by two separate components; the model might be extended in a way, that interactive space-time effects are incorporated. Another point of criticism concerns the general strategy of dealing with temporal correlations within an observed data set: Temporal dependencies are *eliminated* before the spatial analysis is carried out. A promising alternative to this strategy is to *benefit* from the temporal structure as additional information that might improve the efficiency of the spatial analysis. Finally optimal empirical estimation of the model parameters is another unsolved problem of model (1).

As an alternative to the present procedure another completely different approach might serve to identify phenomena of active bacterial movement as well. Markus et al.(1999) analyze hydrogeological space-time data. They state that fluctuations within groundwater level measurements result from different cumulative effects that are impossible to be observed directly (e.g. water infiltration from precipitation). Mathematical identification of such “latent effects“ is performed by means of dynamic factor analysis. In the present case the application of dynamic factor analysis possibly yields bacterial activity as a latent effect governing fluctuations in the observed data. Estimated factor loadings then serve to quantify the intensity of this phenomenon.

Future limnological research projects might benefit from the results of the present study in different ways. Firstly the presented spatiotemporal model can be applied once again if similar problems are to be investigated with data of similar kind. Moreover the efficiency of future projects especially at Lake Cadagno can be enhanced taking into account the present results: The spatial variability of the random variables turbidity and temperature has been characterized via the respective variogram. Knowing the (co)variation structure of a spatial process one can determine an optimal localization of measurement sites if the process is to be observed once again in the future. For example at regions with large variation in the process values a relatively dense network of measurement sites would be recommended. For future projects at Lake Cadagno such considerations of optimal network design can be done in advance. One can determine the localization of turbidity and temperature sensors so that the stochastic characterictics of the respective process are captured optimally.

A few recommendations for future projects at Lake Cadagno can be done across the board. In order to investigate periodical and aperiodical changes of the bacterial layer values of turbidity and temperature should be observed in higher spatial resolution than it has been done in the present study. As far as the temporal resolution is concerned, longer temporal distances between successive measurements surely suffice to observe active movements of the bacterial layer. On the other hand the total length of coherent intervals of observational time has to be enhanced widely. Observing periodic effects with low frequencies like in the present study, the analyzed time span should extent over several days or even weeks.

Acknowledgements

This work was supported by the German Research Foundation (DFG). The authors wish to thank Reinhard Bachofen and Fritz Schwarzenbach for introducing us into the research projects of the Institute of Plant Biology (University of Zürich) at Lake Cadagno. Our special thank goes to Konrad Egli for providing the data of his experiments in 1995 and 1996. Moreover we like to thank Olaf Berke for helpful discussions on the theory of space-time processes.

References

Berke,O. (1999): Estimation and prediction in the spatial linear model. *Water, Air, and Soil Pollution,* 110, 215-237.

Cressie, N.A.C. (1985): Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology,* 17, 563-586.

Cressie, N.A.C. (1988): Spatial prediction and ordinary kriging. *Mathematical Geology,* 20, 405-421.

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

Egli, K. (1997): *Bewegungen einer Population phototropher Schwefelpurpurbakterien im meromiktischen Lago di Cadagno (TI).* Diplomarbeit, Universität Zürich.

Egli, K., Wiggli, M., Klug, J. and Bachofen, R. (1998): Spatial and Temporal Dynamics of the Cell Density in a Plume of Phototropic Microorganisms in their Natural Environment. In Peduzzi, R., Bachofen, R. and Tonolla, M. (eds): *Lake Cadagno: A Meromictic Alpine Lake.* Documenta Ist. ital. Idrobiol., 63, 121-126.

Gerß, J. (1998): *Statistische Analyse von Raum-Zeit-Daten aus der Limnologie - Untersuchung aktiver Vertikalbewegungen von Schwefelpurpurbakterien in ihrer natürlichen Umgebung.* Diplomarbeit, Universität Dortmund.

Journel, A.G. and Huijtbregts, C.J. (1978): *Mining Geostatistics.* Academic Press, London.

Markus, L., Berke, O., Kovacs, J. and Urfer, W. (1999): Spatial prediction of the intensity of latent effects governing hydrogeological phenomena. Accepted by *Environmetrics.*

Peduzzi, R., Bachofen, R. and Tonolla, M. (eds.) (1998): Lake Cadagno: A meromictic Alpine Lake. *Documenta Ist. ital. Idrobiol.,* 63.

Sampson, P.D. and Guttorp, P. (1992): Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association,* 87, 108-119.

Schlittgen, R. and Streitberg, B.H.J. (1997): *Zeitreihenanalyse,* 7.Auflage. Oldenbourg, München.

Vaituzis, Z. and Doetsch, R.N. (1969): Motility tracks: Technique for quantitative study of bacterial movement. *Applied Microbiology,* 17, 584-588.

Appendix

Vertical localization of the bounds of the bacterial layer:

The applied procedure can be characterized as inverse prediction. In the usual prediction problem the value of the observed process, $[Z_t(s)]$ say, *at a fixed location* s_0 is asked for and predicted by means of a function $\hat{Z}_t(s_0)$. In the inverse case the question is the other way round: At what location s does the value of the observed process amount to a given constant c ? Mathematically, s is asked for, so that $Z_t(s) = c$, i.e. $\quad s = Z_t^{-1}(c)$.
Practically, usual ("non-inverse") prediction is applied for locations (depths) $s_j \quad (j = 1, \ldots, m)$ that are as close together as possible. The two neighbouring depths, between those the turbidity value of $c = 8$ FTU is transgressed, are extracted. Linear interpolation of their respective turbidity values yields a unique depth that corresponds to the turbidity of (exactly) 8 FTU. This procedure surely is suboptimal. Errors occur because of the simplifying application of *linear* interpolation. But these errors are expected to be relatively small taking into account the high spatial resolution (3.33cm) to which the observed process is present after the preceding steps of calculation.