Math Geol (2007) 39: 289–305
DOI 10.1007/s11004-007-9085-9

Extending Multivariate Space-Time Geostatistics for Environmental Data Analysis

Chunxue Liu · Katsuaki Koike

Received: 31 August 2005 / Accepted: 5 January 2006 /
Published online: 8 June 2007

Abstract Environmental studies require multivariate data such as chemical concentrations with space-time coordinates. There are two general conditions related to such data: the existence of correlations among the coregionalized variables and the differences in numbers of data which occur because of insufficient data caused by measurement error or bad weather conditions. This study proposes geostatistical techniques for space-time multivariate modeling that take into consideration these correlations and data absences. These techniques consist of suitable modeling of semivariograms and cross-semivariograms for quantifying correlation structures among multivariables and of extending standardized ordinary cokriging. The tensor product cubic smoothing surface method is used for space-time semivariogram modeling. These methods are applied to the chemical component data of the Ariake Sea, a typical closed sea in southwest Japan. In order to clarify environmental changes in the Ariake Sea, the concentration data of four nutritive salts (NO_2–N, NO_3–N, NH_4–N, and PO_4–P) at 38 stations over 25 years are used as environmental indicators. For each of the kinds of data, there are spaces and times for which there is no data available. The effectiveness of the modeling of space-time semivariograms and the high estimation capability of the extended cokriging are demonstrated by cross-validation. Compared with ordinary kriging for a single variable, multivariate space-time standardized ordinary cokriging can provide a more detailed concentration map of nutritive salts and while elucidating their temporal changes over sparsely spaced data areas. In the space-time models by ordinary kriging, on the other hand, smooth trends are obvious.

C. Liu
Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan
e-mail: 047d9904@gsst.stud.kumamoto-u.ac.jp

K. Koike (✉)
Department of Civil and Environmental Engineering, Faculty of Engineering, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan
e-mail: koike@gpo.kumamoto-u.ac.jp

Keywords Space-time semivariogram modeling · Tensor product cubic spline · Standardized ordinary cokriging · Nutritive salt · Sea environment

Introduction

In geochemical, geophysical, and environmental monitoring research on specific targets such as geothermal reservoirs and water pollution, several kinds of data are sampled in space-time regions. In addition to the existence of the space-time correlation of one variable, coregionalized variables are generally correlated among each other. In most cases, all of the variables have not been obtained equally at all of the space-time coordinates, as bad weather conditions and errors in measurement can cause some data to be lacking. For example, if there are two variables, v_1 and v_2, at location u_1, there may only be one variable, v_1, at location u_2.

The purpose of this study is to develop geostatistical techniques that are suitable for space-time multivariate modeling by considering correlation structures and irregular absences of data. Multivariate space-time geostatistics is an extension of spatial geostatistics and has been formulated in different forms by Rouhani and Wackernagel (1990), Goovaerts and Sonnet (1993), and Christakos and Hristopulos (1998). Its useful applications can be seen in the analyses of groundwater quality (Koike et al. 2002) and atmospheric pollution (De Iaco et al. 2002a, 2002b). However, these multivariate space-time kriging methods are just extended forms of traditional ordinary cokriging, in which all the variables must be present at the same space-time sample points. Thus, they are not applicable to cases such as those described above with incomplete data. In this study, a multivariate space-time kriging method is extended from standardized ordinary cokriging and made applicable to case studies in which the variable numbers are not necessarily the same.

Variogram modeling is also an indispensable problem in space-time geostatistics. For a single variable, an integrate model (Cressie and Huang 1999) and a product-sum model (De Cesare et al. 2001) have been proposed to represent the nonseparable space-time covariance, while the Linear Coregionalization Method (LCM) (Goovaerts 1993) has been generally accepted for multivariate modeling. However, a more complicated model is needed for space-time analysis. De Iaco et al. (2003) have proposed an extended LCM in which the space-time basic semivariograms are modeled by the product-sum method. Koike et al. (2002) have shown the applicability of a polynomial method to multivariate semivariograms that fluctuate largely in the space-time region. This paper improves on this polynomial method by a tensor product cubic smoothing surface for better fitting of the model to experimental semivariograms.

The proposed techniques are applied to a case study of a closed sea environment using the monthly concentration data of nutritive salts. The effectiveness of space-time concentration models are checked by comparisons with single variable models by ordinary kriging and a sea-color image produced from a color composite of the visible wavelength band data of the Landsat–5 Thematic Mapper.

Multivariate Space-Time Semivariogram Matrix

Let $\{\mathbf{Z}(\mathbf{u}); \mathbf{u} \in R^{d+1}\}$ be a multivariate space-time random field which has the following form of matrix equation

$$\mathbf{Z}(\mathbf{u}) = [Z_1(\mathbf{u}), Z_2(\mathbf{u}), \ldots, Z_p(\mathbf{u})]^{\mathrm{T}} \tag{1}$$

where $\mathbf{u} = (\mathbf{s}, t)$, $\mathbf{s} = (s_s, s_2, \ldots, s_d) \in D \subseteq R^d$ (generally $d = 3$) denotes spatial coordinates, $t \in T \subseteq R$ is the temporal coordinate, p is the number of variables (generally $p \geq 2$), and $D \times T \subseteq R^{d+1}$.

If $\mathbf{Z}(\mathbf{u})$ satisfies the second-order stationarity, then its mean vector, $\mathbf{M}$, is independent of $(\mathbf{s}, t)$

$$\mathrm{E}[\mathbf{Z}(\mathbf{u})] = \mathbf{M} = [M_1, M_2, \ldots, M_p]^{\mathrm{T}} \tag{2}$$

In addition, covariance and semivariogram matrices ($p \times p$) can be defined for a pair of multivariates, $\mathbf{Z}(\mathbf{u})$ and $\mathbf{Z}(\mathbf{u}+\mathbf{h})$, which only depend on the space-time separation vector, $\mathbf{h} = (\mathbf{h}_s, h_t)$.

$$\mathbf{C}[\mathbf{Z}(\mathbf{u}), \mathbf{Z}(\mathbf{u}+\mathbf{h})] = \mathrm{E}\{[\mathbf{Z}(\mathbf{u}) - \mathbf{M}][\mathbf{Z}(\mathbf{u}+\mathbf{h}) - \mathbf{M}]^{\mathrm{T}}\} = \mathbf{C}(\mathbf{h}) = [C^{\alpha\beta}(\mathbf{h})] \tag{3}$$

$$\begin{aligned}\boldsymbol{\Gamma}[\mathbf{Z}(\mathbf{u}), \mathbf{Z}(\mathbf{u}+\mathbf{h})] &= \mathrm{E}\{[\mathbf{Z}(\mathbf{u}) - \mathbf{Z}(\mathbf{u}+\mathbf{h})][\mathbf{Z}(\mathbf{u}) - \mathbf{Z}(\mathbf{u}+\mathbf{h})]^{\mathrm{T}}\} = \boldsymbol{\Gamma}(\mathbf{h}) \\ &= [\gamma^{\alpha\beta}(\mathbf{h})]\end{aligned} \tag{4}$$

where α and β denote the kind of variable ($\alpha, \beta = 1, 2, \ldots, p$). $C^{\alpha\beta}(\mathbf{h})$ and $\gamma^{\alpha\beta}(\mathbf{h})$ are the cross-covariance and cross-semivariogram, respectively, between a pair of variables, $Z^{\alpha}(\mathbf{u})$ and $Z^{\beta}(\mathbf{u}+\mathbf{h})$, if $\alpha \neq \beta$, and are the direct covariance and semivariogram if $\alpha = \beta$.

In order to model the experimental multivariate space-time semivariogram, De Iaco et al. (2002a, 2002b) have proposed the following two integrated product and product-sum models by extending the models of Cressie and Huang (1999) and De Cesare et al. (2001).

1. The integrated product model is formulated under the assumption that $C_s(\mathbf{h}_s; a)C_t(h_t; a)$ is integrable with respect to a measure μ over V for each $\mathbf{h}_s$ and h_t with an arbitrary coefficient $k > 0$.

$$\begin{aligned}\gamma_{s,t}(\mathbf{h}_s, h_t) = \int_V & k[C_t(0; a)\gamma_s(\mathbf{h}_s; a) + C_s(0; a)\gamma_t(h_t; a) \\ & - \gamma_s(\mathbf{h}_s; a)\gamma_t(h_t; a)]d\mu(a)\end{aligned} \tag{5}$$

   where $\mu(a)$ is a positive measure over $U \subseteq R$ for each choice of $a \in V \subseteq U$, $\gamma_s(\mathbf{h}_s; a)$ and $\gamma_t(h_t; a)$ are valid spatial and temporal semivariograms, respectively, in $D \subset R^n$ and $T \subset R_+$, and $C_s(0; a)$ and $C_t(0; a)$ are the corresponding sill values.
2. The integrated product-sum model assumes that $k_1 C_s(\mathbf{h}_s; a)C_t(h_t; a) + k_2 C_s(\mathbf{h}_s; a) + k_3 C_t(h_t; a)$ is integrable with respect to the measure μ over V for each $\mathbf{h}_s$ and h_t with coefficients $k_1 > 0, k_2 \geq 0$, and $k_3 \geq 0$.

$$\gamma_{s,t}(\mathbf{h}_s, h_t) = \int_V [(k_2 + k_1 C_t(0; a))\gamma_s(\mathbf{h}_s; a) + (k_3 + k_1 C_s(0; a))\gamma_t(h_t; a) - k_1 \gamma_s(\mathbf{h}_s; a)\gamma_t(h_t; a)] d\mu(a) \tag{6}$$

The superiority of the product-sum model for the product model was proved by De Iaco et al. (2002a, 2002b). Alternative modeling has been based on a polynomial expression for nested structures (Koike et al. 2002), in which the space-time model is expressed directly by the sum of a series of functions, $f_i^{\alpha\beta}$, with the coefficient, $\tau_i^{\alpha\beta}$, for $(\gamma_s^{\alpha\beta}, \gamma_t^{\alpha\beta})$ at different locations $(\mathbf{h}_s, h_t)$ by

$$\gamma_{st}^{\alpha\beta}(\mathbf{h}_s, h_t) = \sum_i \tau_i^{\alpha\beta} f_i^{\alpha\beta}(\mathbf{h}_s, h_t) \tag{7}$$

In practice, this polynomial expression holds the same characteristics as the summation of different types of elementary semivariograms.

We improve the polynomial form by applying a tensor product cubic smoothing surface to the model's experimental space-time semivariogram. The tensor product function, $S(\mathbf{h}_s, h_t)$, is the weighted sum of the products of two elementary independent functions, $f(\mathbf{h}_s)$ multiplied by $g(h_t)$ with coefficients b_{kl}

$$S(\mathbf{h}_s, h_t) = \sum_{k=0}^{K} \sum_{l=0}^{L} b_{kl} f_k(\mathbf{h}_s) g_l(h_t) \tag{8}$$

where K and L are the orders of f and g, respectively. In the case of a cubic spline function, K and L are 3, with 16 unknown b_{kl}. Considering that the first and second derivatives of S must be continuous at all the lags $(\mathbf{h}_s, h_t)$, 15 independent conditions are necessary to determine the function. The remaining condition is the balance of the two criteria between (1) the closeness of the function to the experimental semivariogram data, and (2) its smoothness as defined by small curvature. For simplification, we shall let the weights for this balance be equal to all of the $(\mathbf{h}_s, h_t)$.

1. The closeness is measured by the summation of squared residuals

$$\mathrm{E}(S) = \sum_{i=1}^{m} \sum_{j=1}^{n} |S(\mathbf{h}_{si}, h_{tj}) - \gamma_{ij}|^2 \tag{9}$$

   where m and n are the numbers of $\mathbf{h}_s$ and h_t, respectively, and the amount of experiment semivariogram data is $m \times n$.
2. The smoothness is defined by the summation of the squared second derivative of S, S''

$$F(S'') = \int_{h_{s1}}^{h_{sm}} \int_{h_{t1}}^{h_{tn}} \lambda(\mathbf{h}_s, h_t) |S''(\mathbf{h}_s, h_t)|^2 d\mathbf{h}_s dh_t \tag{10}$$

   where S'' is the second derivative of S.

The two criteria can be balanced with a parameter, $\rho \in [0, 1]$, that determines which criterion is considered important

$$\rho E(f) + (1 - \rho) F(S'') \tag{11}$$

In this study, ρ is defined as 0.9 in order to moderately fit the model to the experimental semivariogram data.

Standardized Ordinary Cokriging

Multivariate space-time ordinary cokriging modeling can be accomplished by extending ordinary cokriging. Let $\hat{z}^\alpha(\mathbf{u}^\alpha)$ be the kriged estimate of variable $\alpha(\alpha = 1, \ldots, p)$ for the true value $z^\alpha(\mathbf{u}^\alpha)$ at location $\mathbf{u}^\alpha$. The $\mathbf{u}$ is changeable depending on the kind of variable and should therefore be rearranged because patterns of insufficient data differ. The general form of ordinary cokriging, which is a weighted mean method performed by assigning suitable weights to sample data around $\mathbf{u}^\alpha$, is

$$\hat{z}^\alpha(\mathbf{u}^\alpha) = \sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta z^\beta(\mathbf{u}_w^\beta) \quad (\alpha = 1, 2, \ldots, p) \tag{12}$$

where $z^\beta(\mathbf{u}_w^\beta)$, λ_w^β, and N_β denote the sample data of variable β at location $\mathbf{u}_w^\beta$, the kriging weight, and the data number of variable β, respectively.

The kriging weights in multivariate space-time ordinary cokriging can be derived through the following two conditions that should be satisfied in cokriging.

1. The unbiased condition

$$\mathrm{E}[\hat{z}^\alpha(\mathbf{u}^\alpha) - z^\alpha(\mathbf{u}^\alpha)] = 0 \tag{13}$$

which derives the following two different equations

$$\sum_{w=1}^{N_\alpha} \lambda_w^\alpha \mathrm{E}[z^\alpha(\mathbf{u}_w^\alpha)] + \sum_{\beta=1,\beta\neq\alpha}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta \mathrm{E}[z^\beta(\mathbf{u}_w^\beta)] - \mathrm{E}[z^\alpha(\mathbf{u}^\alpha)] = 0 \quad \text{and} \tag{14}$$

$$\sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta \mathrm{E}[z^\beta(\mathbf{u}_w^\beta)] - \mathrm{E}[z^\alpha(\mathbf{u}^\alpha)] = 0 \tag{15}$$

Equation (14) is the same form as that of traditional ordinary cokriging based on the following unbiased condition

$$\sum_{w=1}^{N_\alpha} \lambda_w^\alpha = 1 \quad \text{and} \quad \sum_{w=1}^{N_\beta} \lambda_w^\beta = 0 \quad (\beta = 1, 2, \ldots, p, \beta \neq \alpha)$$

This restricts severely the influence of the other variables (Deutsch and Journel 1998), and allows the formation of the kriging weight matrix of (19) only when all the variables are present at each sample point.

On the other hand, (15) expresses the standardized ordinary cokriging that replaces the sample value $z^\beta(\mathbf{u}_w^\beta)$ with $z^\beta(\mathbf{u}_w^\beta) + M_\alpha - M_\beta$ under the condition of $\sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta = 1$. M_α and M_β are the means of z^α and z^β, respectively. This paper adopts the second unbiased condition for utilizing the influence from the other variables and considering the above absences of data.

2. The least estimation variance condition

$$\begin{aligned}[\sigma^2]^\alpha &= \mathrm{Var}[\hat{z}^\alpha(\mathbf{u}^\alpha) - z^\alpha(\mathbf{u}^\alpha)] \\ &= \mathrm{E}\{[z^\alpha(\mathbf{u}^\alpha)]^2\} - 2\mathrm{E}[\hat{z}^\alpha(\mathbf{u}^\alpha) z^\alpha(\mathbf{u}^\alpha)] + \mathrm{E}\{[\hat{z}^\alpha(\mathbf{u}^\alpha)]^2\} \\ &= C(\mathbf{u}^\alpha, \mathbf{u}^\alpha) - 2\sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta C(\mathbf{u}_w^\beta, \mathbf{u}^\alpha) \\ &\quad + \sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta \sum_{\theta=1}^{p}\sum_{v=1}^{N_\theta} \lambda_v^\theta C(\mathbf{u}_w^\beta, \mathbf{u}_v^\theta)\end{aligned} \tag{16}$$

The minimization of objective function $F = [\sigma^2]^\alpha - 2\mu(\sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta}\lambda_w^\beta - 1)$ with Lagrange multiplier μ is accomplished by

$$\begin{cases} \dfrac{\partial F}{\partial \lambda_w^\beta} = -2C(\mathbf{u}_w^\beta, \mathbf{u}^\alpha) + 2\displaystyle\sum_{\vartheta=1}^{p}\sum_{v=1}^{N_\theta} \lambda_v^\theta C(\mathbf{u}_w^\beta, \mathbf{u}_v^\theta) - 2\mu = 0 \\ \dfrac{\partial F}{\partial \mu} = -2\left(\displaystyle\sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta - 1\right) = 0 \end{cases} \tag{17}$$

which derives the linear equation system

$$\begin{cases} \displaystyle\sum_{\theta=1}^{p}\sum_{v=1}^{N_\theta} \lambda_v^\theta C(\mathbf{u}_w^\beta, \mathbf{u}_v^\theta) - \mu = C(\mathbf{u}_w^\beta, \mathbf{u}^\alpha) \\ \displaystyle\sum_{\beta=1}^{p}\sum_{w=1}^{N_\beta} \lambda_w^\beta = 1 \end{cases} \tag{18}$$

Consequently, the kriging weights can be determined by solving the matrix

$$\begin{bmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I}^\mathrm{T} & 0 \end{bmatrix} \times \begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{C}^\alpha \\ 1 \end{bmatrix} \tag{19}$$

$$\mathbf{C} = \begin{bmatrix}
C^{11}(\mathbf{h}_{11}) & C^{11}(\mathbf{h}_{12}) & \cdots & C^{11}(\mathbf{h}_{1N_1}) & C^{12}(\mathbf{h}_{11}) & C^{12}(\mathbf{h}_{12}) & \cdots & C^{12}(\mathbf{h}_{1N_2}) \\
C^{11}(\mathbf{h}_{21}) & C^{11}(\mathbf{h}_{22}) & \cdots & C^{11}(\mathbf{h}_{2N_1}) & C^{12}(\mathbf{h}_{21}) & C^{12}(\mathbf{h}_{22}) & \cdots & C^{12}(\mathbf{h}_{2N_2}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
C^{11}(\mathbf{h}_{N_1 1}) & C^{11}(\mathbf{h}_{N_1 2}) & \cdots & C^{11}(\mathbf{h}_{N_1 N_1}) & C^{12}(\mathbf{h}_{N_1 1}) & C^{12}(\mathbf{h}_{N_1 2}) & \cdots & C^{12}(\mathbf{h}_{N_1 N_2}) \\
C^{21}(\mathbf{h}_{11}) & C^{21}(\mathbf{h}_{12}) & \cdots & C^{21}(\mathbf{h}_{1N_1}) & C^{22}(\mathbf{h}_{11}) & C^{22}(\mathbf{h}_{12}) & \cdots & C^{22}(\mathbf{h}_{1N_2}) \\
C^{21}(\mathbf{h}_{21}) & C^{21}(\mathbf{h}_{22}) & \cdots & C^{21}(\mathbf{h}_{2N_1}) & C^{22}(\mathbf{h}_{21}) & C^{22}(\mathbf{h}_{22}) & \cdots & C^{22}(\mathbf{h}_{2N_2}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
C^{21}(\mathbf{h}_{N_1 1}) & C^{21}(\mathbf{h}_{N_1 2}) & \cdots & C^{21}(\mathbf{h}_{N_1 N_1}) & C^{22}(\mathbf{h}_{N_1 1}) & C^{22}(\mathbf{h}_{N_1 2}) & \cdots & C^{22}(\mathbf{h}_{N_1 N_2}) \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
C^{p1}(\mathbf{h}_{11}) & C^{p1}(\mathbf{h}_{12}) & \cdots & C^{p1}(\mathbf{h}_{1N_1}) & C^{p2}(\mathbf{h}_{11}) & C^{p2}(\mathbf{h}_{12}) & \cdots & C^{p2}(\mathbf{h}_{1N_2}) \\
C^{p1}(\mathbf{h}_{21}) & C^{p1}(\mathbf{h}_{22}) & \cdots & C^{p1}(\mathbf{h}_{2N_1}) & C^{p2}(\mathbf{h}_{21}) & C^{p2}(\mathbf{h}_{22}) & \cdots & C^{p2}(\mathbf{h}_{2N_2}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
C^{p1}(\mathbf{h}_{N_1 1}) & C^{p1}(\mathbf{h}_{N_1 2}) & \cdots & C^{p1}(\mathbf{h}_{N_1 N_1}) & C^{p2}(\mathbf{h}_{N_1 1}) & C^{p2}(\mathbf{h}_{N_1 2}) & \cdots & C^{p2}(\mathbf{h}_{N_1 N_2})
\end{bmatrix}$$

$$
\left.\begin{matrix}
 & C^{1p}(\mathbf{h}_{11}) & C^{1p}(\mathbf{h}_{12}) & \cdots & C^{1p}(\mathbf{h}_{1N_p}) \\
 & C^{1p}(\mathbf{h}_{21}) & C^{1p}(\mathbf{h}_{22}) & \cdots & C^{1p}(\mathbf{h}_{2N_p}) \\
\cdots & \vdots & \vdots & \ddots & \vdots \\
 & C^{1p}(\mathbf{h}_{N_1 1}) & C^{1p}(\mathbf{h}_{N_1 2}) & \cdots & C^{1p}(\mathbf{h}_{N_1 N_p}) \\
 & C^{2p}(\mathbf{h}_{11}) & C^{2p}(\mathbf{h}_{12}) & \cdots & C^{2p}(\mathbf{h}_{1N_p}) \\
 & C^{2p}(\mathbf{h}_{21}) & C^{2p}(\mathbf{h}_{22}) & \cdots & C^{2p}(\mathbf{h}_{2N_p}) \\
\cdots & \vdots & \vdots & \ddots & \vdots \\
 & C^{2p}(\mathbf{h}_{N_1 1}) & C^{2p}(\mathbf{h}_{N_1 2}) & \cdots & C^{2p}(\mathbf{h}_{N_1 N_p}) \\
 & \vdots & \vdots & & \vdots \\
 & C^{pp}(\mathbf{h}_{11}) & C^{pp}(\mathbf{h}_{12}) & \cdots & C^{pp}(\mathbf{h}_{1N_p}) \\
\cdots & C^{pp}(\mathbf{h}_{21}) & C^{pp}(\mathbf{h}_{22}) & \cdots & C^{pp}(\mathbf{h}_{2N_p}) \\
 & \vdots & \vdots & \ddots & \vdots \\
 & C^{pp}(\mathbf{h}_{N_1 1}) & C^{pp}(\mathbf{h}_{N_1 2}) & \cdots & C^{pp}(\mathbf{h}_{N_1 N_p})
\end{matrix}\right]
$$

$$
\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1^1 \\ \lambda_2^1 \\ \vdots \\ \lambda_{N_1}^1 \\ \lambda_1^2 \\ \lambda_2^2 \\ \vdots \\ \lambda_{N_2}^2 \\ \vdots \\ \lambda_1^p \\ \lambda_2^p \\ \vdots \\ \lambda_{N_p}^p \end{bmatrix}, \qquad \mathbf{C}^{\alpha} = \begin{bmatrix} C^{1\alpha}(\mathbf{h}_{1\mathbf{u}}) \\ C^{1\alpha}(\mathbf{h}_{2\mathbf{u}}) \\ \vdots \\ C^{1\alpha}(\mathbf{h}_{N_1\mathbf{u}}) \\ C^{2\alpha}(\mathbf{h}_{1\mathbf{u}}) \\ C^{2\alpha}(\mathbf{h}_{2\mathbf{u}}) \\ \vdots \\ C^{2\alpha}(\mathbf{h}_{N_2\mathbf{u}}) \\ \vdots \\ C^{p\alpha}(\mathbf{h}_{1\mathbf{u}}) \\ C^{p\alpha}(\mathbf{h}_{2\mathbf{u}}) \\ \vdots \\ C^{p\alpha}(\mathbf{h}_{N_p\mathbf{u}}) \end{bmatrix}
$$

where $C^{\alpha\beta}(\mathbf{h}_{wv})$ denotes the cross-covariance of variables α and β between two space-time locations, $\mathbf{u}_w^\alpha$ and $\mathbf{u}_v^\beta$, and $\mathbf{I}$ is the unit column vector with the same dimension of $\mathbf{C}$ as $\sum_{\beta=1}^{p} N_\beta$. For the cross-semivariogram form, $C^{\alpha\beta}(\mathbf{h}_{wv})$ can be replaced by $\gamma^{\alpha\beta}(\mathbf{h}_{wv})$.

A Case Study

Data Sets and Variogram Modeling

The proposed methods are tested for a sea environmental change problem on Ariake Sea in central Kyusyu, southwest Japan, which is typical closed sea with only one narrow channel to the outer sea in the west (Fig. 1). This is shallow sea with an average water depth of 20 m, and especially shallow in the northern part under

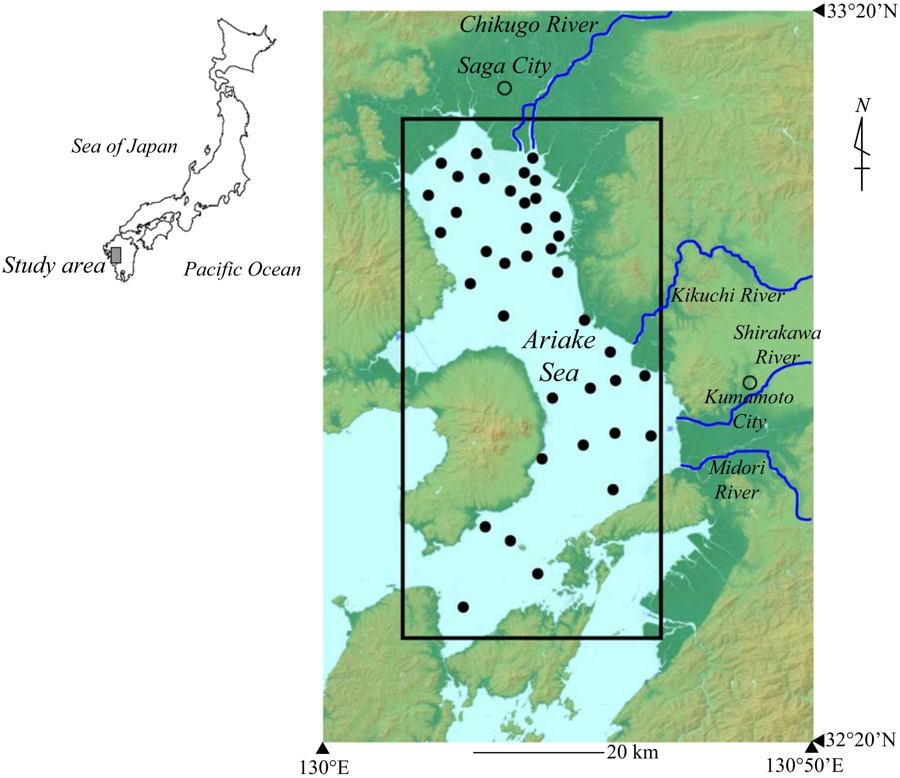

Fig. 1 Location of study area and 38 stations for monthly measuring physical and chemical properties of sea water. The rectangle is the calculation area for kriging shown in Figs. 6 and 7

5 m depth. The exchange cycle of sea water is about two months generally, but it is faster in the channel and river mouths (Takikawa et al. 2003). Its environment is known to have been worse recently due to many factors such as urban developments near the sea. There are 38 stations for periodically measuring (at intervals of nearly one month) the physical and chemical properties of the sea water in the study area with about 40 km × 80 km extent. The stations were set by three marine resources laboratories in the three prefectures (Kumamoto, Saga, and Fukuoka) surrounding the sea to monitor the water quality. Investigation items include: vertical profiles of water temperature and salinity, water clarity, amount of plankton sediments, pH, dissolved oxygen (DO), chemical oxygen demand (COD); and concentration of NO_2–N, NO_3–N, NH_4–N, PO_4–P, SiO_2–Si, and dissolved inorganic nitride (DIN) (Takikawa et al. 2003). Among the data, four nutritive salts (NO_2–N, NO_3–N, NH_4–N, and PO_4–P), which had been sampled monthly near the sea surface from 1975 to 2000, were selected for the analysis by multivariate space-time geostatistics. This is because of relatively high data quality and their importance as environmental indicator.

Because the four nutritive salts are compositional volume data, the methods proposed by Aitchison et al. (2000), Pawlowsky-Glahn and Egozcue (2002), and Martín-Fernández et al. (2003), which consider the specific characters of this type of data, may be necessary. However, the four nutritive salts are minor elements and the con-

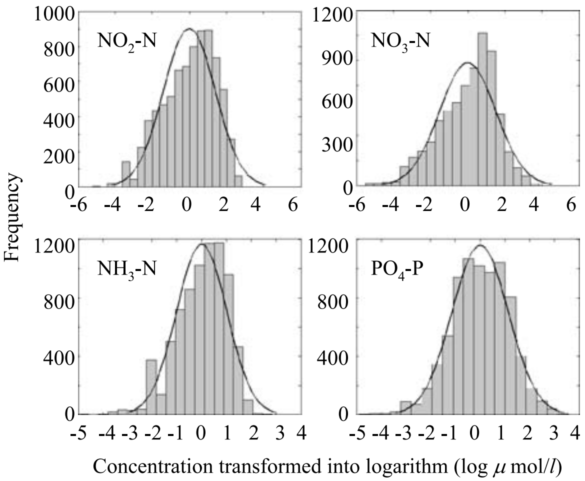

Fig. 2 Histograms of the concentration data of four nutritive salts, NO_2–N, NO_3–N, NH_4–N, and PO_4–P, that are transformed into logarithm and zero means by subtracting their logarithmic averages. Gaussian curves are overlaid with the histograms to examine lognormality

centrations are much smaller than the data examined in those references. This requires no constraint conditions that relate the nutritive salts concentrations. Thus, the present data are applicable to cokriging as usual multivariate.

Figure 2 shows the histograms of the concentration data of the four nutritive salts transformed into logarithm, whereby they can be approximated by a normal distribution that is suitable to geostatistical application. The average values are removed from the log-transformed data to make zero-mean data sets. The following analyses use these data sets, and the results are transformed inversely into the original scale.

The experimental space-time semivariograms, cross-semivariograms, and the tensor product cubic smoothing surface models of the nutritive salt concentrations are drawn in Fig. 3, in which the lag along space is 2 km with 1 km tolerance and the lag along time is one month with half a month tolerance. Considering the shape of the Ariake Sea and the arrangement of the 38 stations (Fig. 1), the anisotropic behaviors of variograms should be checked along the two directions of the long axis, NE–SW, and the short axis, NW–SE, of the arrangement. However, their directional variograms are almost the same, and for this reason we used omni-directional variograms only.

Cross-Validation

The effectiveness of the variogram models and the multivariate space-time standardized cokriging was evaluated by cross-validation that clarified the relationships between the observation and estimation values. The estimations were calculated by the sample data around the sample point but excluding the data at the point. Kriging estimation requires at least 3 and 1 data along the space and time, and in this evaluation these values at most grid points were 4 and 2, respectively, taking into consideration the search ranges and minimizing estimation errors. Accordingly, in order to estimate the value at each space-time grid point, we used 3 to 8 sample data generally, consisting of 1 to 4 variables around the grid point. The largest spatial search range of the data was 15 km, and the longest temporal search range was 4 months for kriging.

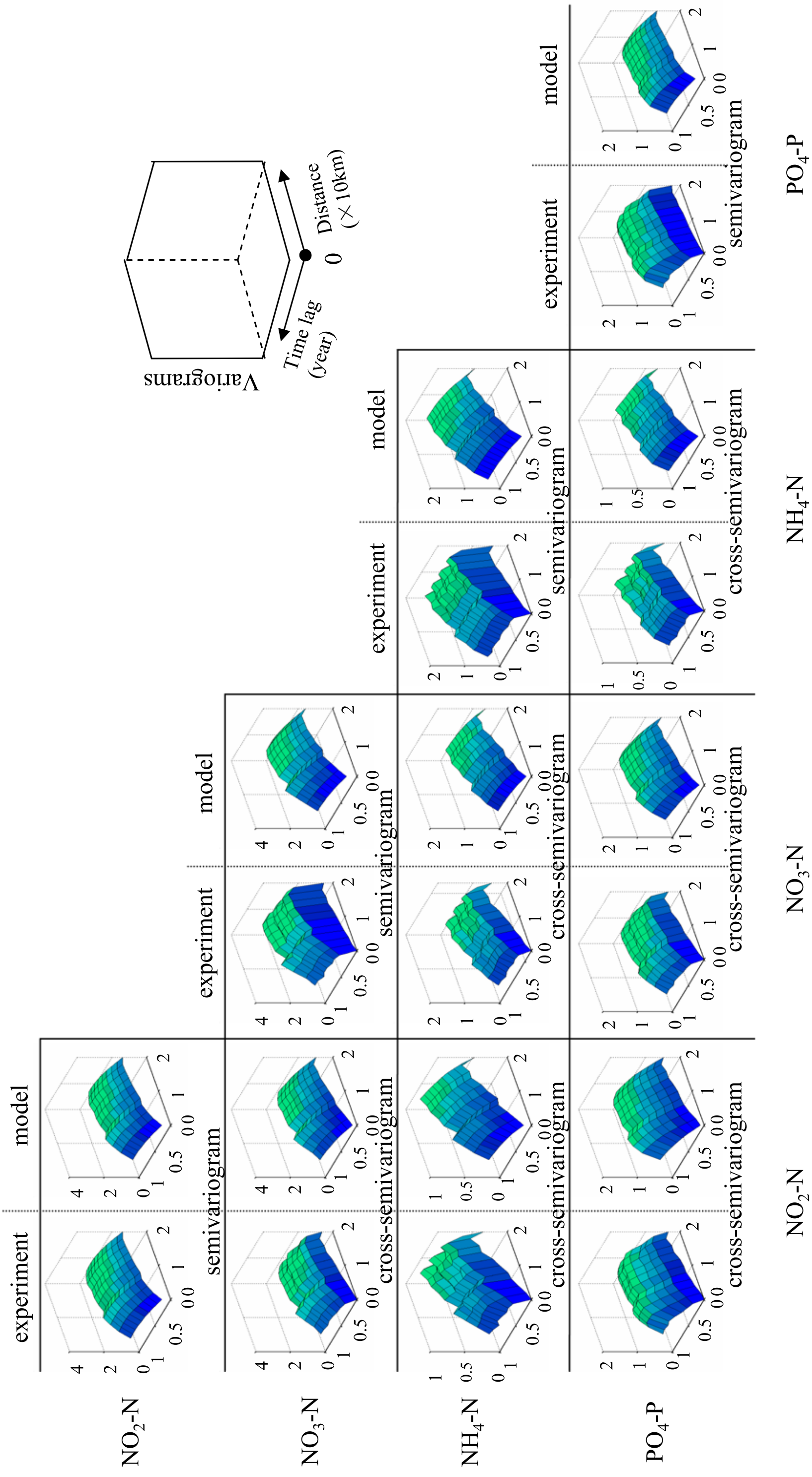

Fig. 3 Matrix showing experimental and modeled space-time semivariograms of the four nutritive salts and cross-semivariograms between two components. The tensor product cubic smoothing surface method was used for the modeling

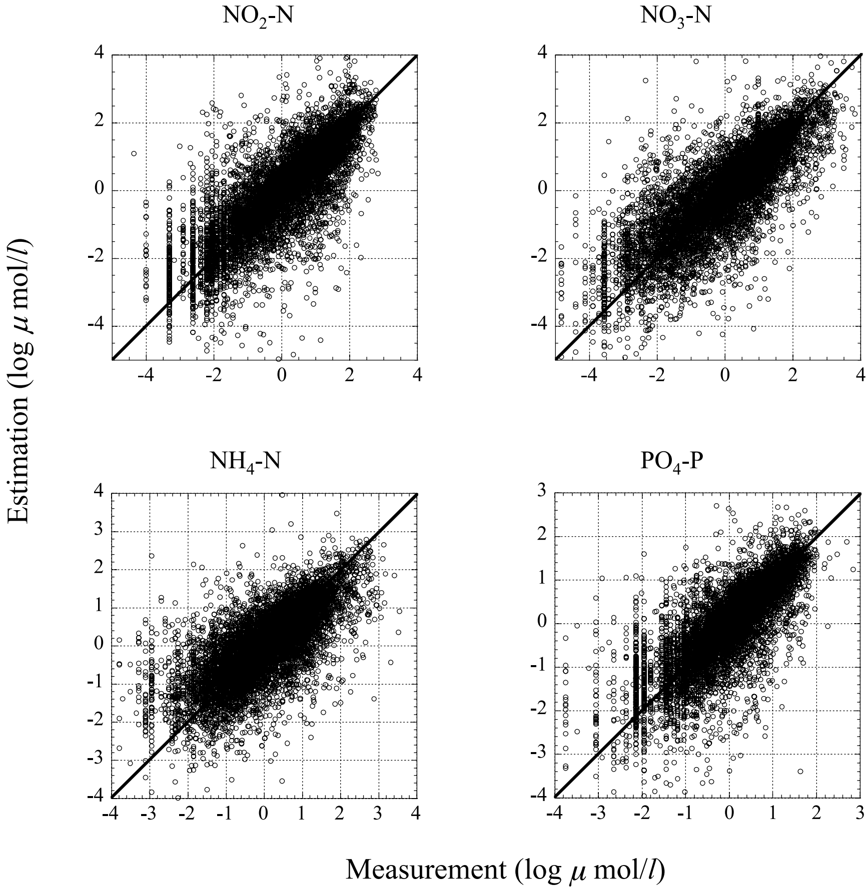

Fig. 4 Scattergrams for the cross-validation between the observation and estimation values by multivariate standardized space-time ordinary cokriging. The 45 degree lines are included to check the degree of the smoothing effect

The correlation coefficients obtained in the cross-validation were: NO_2–N = 0.79, NO_3–N = 0.80, NH_4–N = 0.70, and PO_4–P = 0.77. The 45 degree lines are overlaid with the scattergrams of Fig. 4 to understand the degree of the smoothing effect. In all the nutritive salts, the clusters are found to be elongated along the lines. Furthermore, two space-time semivariogram models of NH_4–N and two cross semivariogram models of NO_3–N and PO_4–P, which were produced by the above product-sum and tensor product cubic smoothing surface methods, are compared in Fig. 5. Obviously, the product-sum models are almost flat, and the nugget effects are too large for the sills, signifying weaknesses in the correlation structures. The tensor product cubic smoothing surface models fit the complicated changes of the experimental variograms much better. The flat variogram models obtained by the product-sum method resulted in a decrease of estimation accuracy, shown by lower correlation coefficients for the cross-validation: NO_2–N = 0.74, NO_3–N = 0.55, NH_4–N = 0.62, and

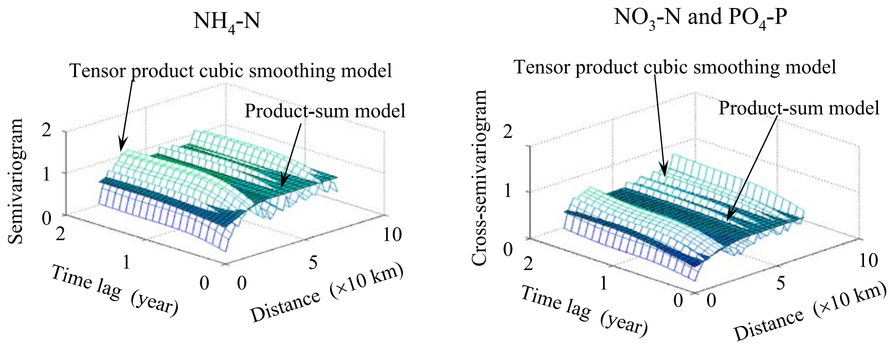

Fig. 5 Comparison of two space-time semivariogram models for NH_4–N and cross-semivariogram models for NO_3–N and PO_4–P by the product-sum method and the tensor product cubic smoothing surface method

PO_4–P = 0.77. Therefore, it could be confirmed that the tensor product cubic smoothing method is more suitable.

Space-time Concentration Model

Space-time concentrations were estimated at a space interval of 1 km × 1 km over the study area and at a time interval of 3 months from 1975 to 2000. Therefore, the 40 × 80 × 104 space-time points in total were targeted for the calculation. Figure 6 illustrates examples of the estimation results of NO_3–N and NH_4–N in June and NO_2–N and PO_4–P in September. These months were chosen because their concentrations had characteristic distributions in that the concentrations have large spatial variations over the Ariake Sea. In these figures, the black areas are the land and the white areas in the sea are where the concentrations could not be estimated because of a lack of sufficient data.

The results are compared with those obtained from single variable space-time ordinary kriging (Liu et al. 2004; Koike et al. 2005), as shown in the lower column. General trends with respect to the locations of high and low concentrations are similar between the multivariate and single variable kriging models. However, it is conspicuous that the multivariate model can express local changes of the concentrations while overcoming the smoothing effect shown in the single variable models. Another advantage of the model is that it can fill in the blank areas that the single variable models can not estimate because of the lack of data regarding the target component. This can be confirmed well from the northern parts of the concentration maps of NO_2–N in September 1996 and September 1999, NH_4–H in June 1995, and PO_4–P in September 1999.

Common to the four components and for all years, the concentrations are relatively high in the northern sea and near the land, and in particular near large cities such as Kumamoto (Fig. 1). It is noteworthy that the concentrations of NO_3–N and NH_4–N increased in the northern sea after 1995 and 1994, respectively, as such trends signify that the concentrations of nutritive salts can be influenced by urban development. Another interesting characteristic is found in the maps for 1996 and 1997, in which the concentrations of NO_2–N and PO_4–P occasionally became the highest or second

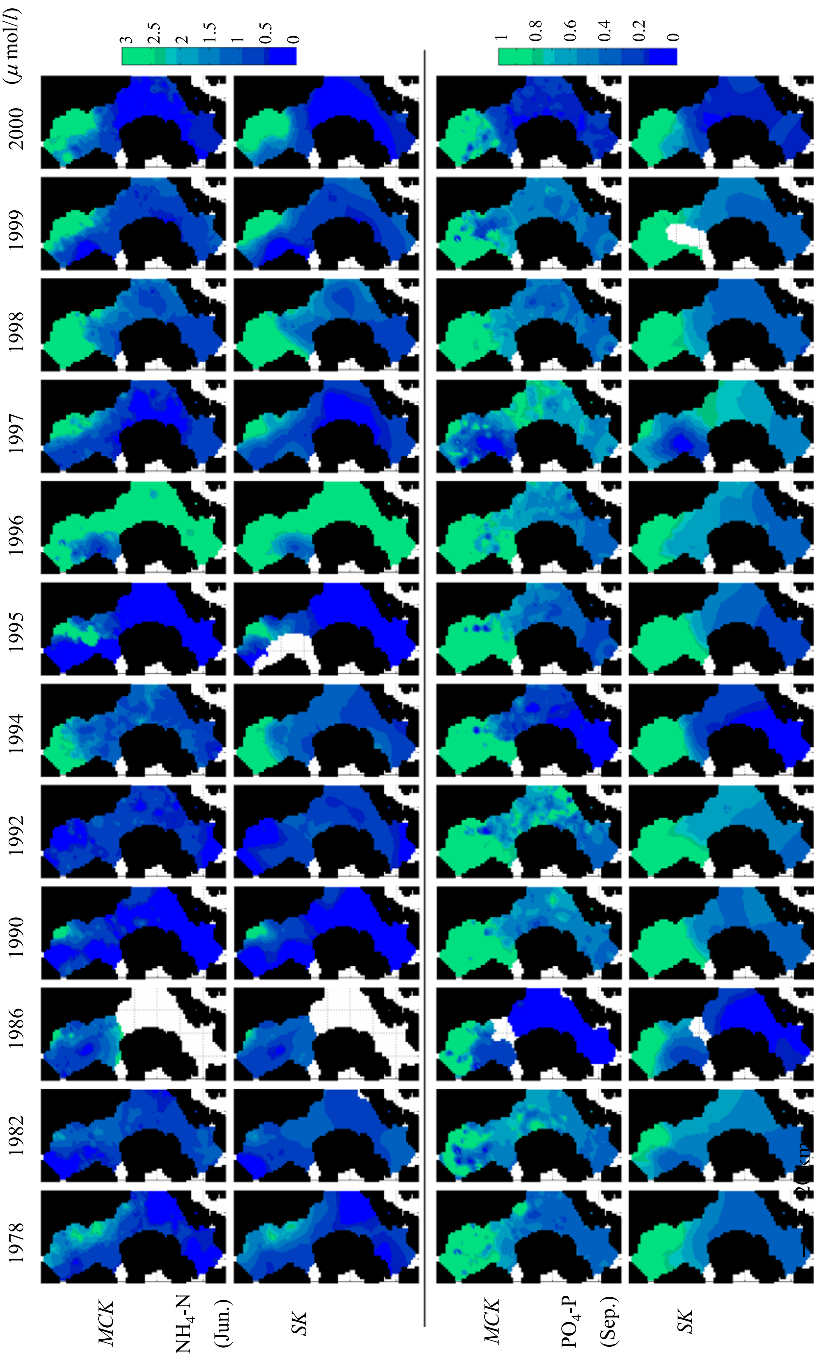

Fig. 6 Estimated distributions of the concentrations of four nutritive salts over the Ariake Sea. MCK and SK denote space-time models by multivariate standardized ordinary cokriging and single variable ordinary kriging

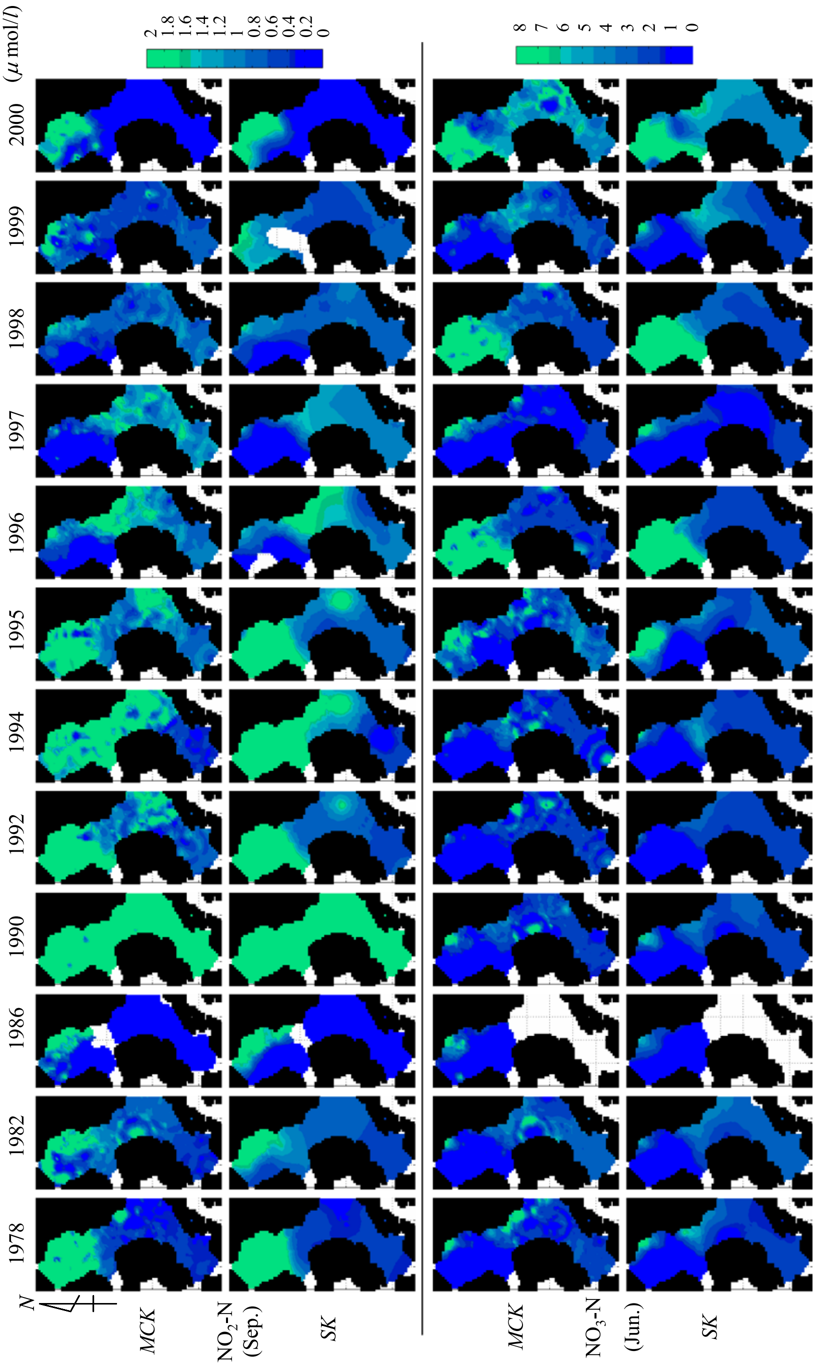

Fig. 6 (Continued)

highest in the middle sea near Kumamoto City. This trend was not observed for the other components.

Discussion

The validity of complicated distributions of nutritive salt concentrations by multivariate cokriging must be examined. To this end, we focused on the sea color as obtained through Landsat–5 Thematic Mapper (TM) images because color trends are correlated with water turbidity and quality. Three sources of visible band data (bands 1, 2, and 3) were used to make a true color composite image on 10 November 2000. The blue, green, and red colors are assigned to bands 1, 2, and 3, respectively. The turbidity is shown by the unclearness of the blue color. The image was compared with the estimated concentrations of NH_4–N in November 2000 by the cokriging and kriging models (Fig. 7). Because these visible blue, green, and red band data have relatively high correlations with the NH_4–N concentrations at the measurement stations, as shown in the correlation coefficients of 0.80, 0.79, and 0.76, respectively, the color composite image soft data was chosen for the check. This correlation may originate from the fact that large water turbidity is associated partly with dense sand and clay particles that are poured into the sea through rivers along with nutritive salts.

Two points of difference between the two models are noticeable in the northern sea. The cokriging model could clearly separate the area into two zones, the eastern side with the highest concentrations and the western side subordinate to the eastern side. In the kriging model, on the other hand, the concentrations in the area are

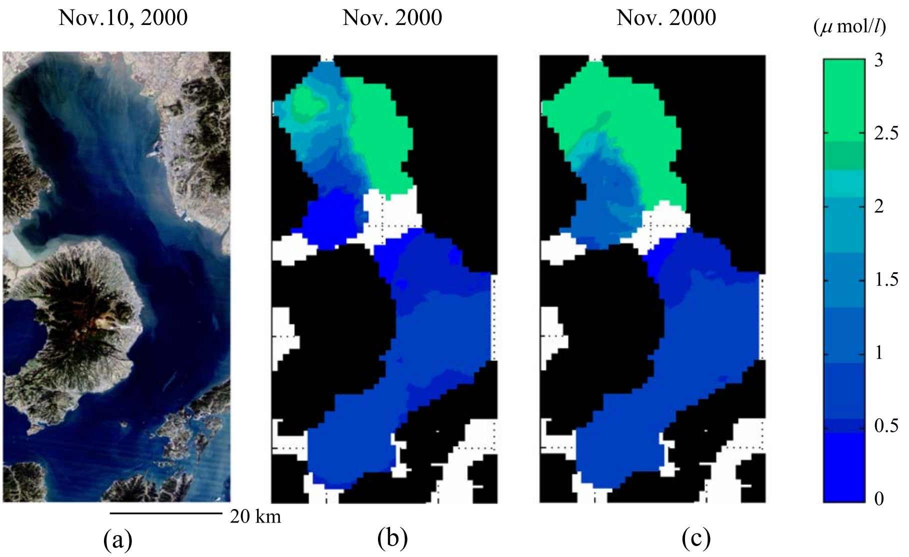

Fig. 7 Comparison of (**a**) true color composite image of band 1 (*blue*), 2 (*green*), and 3 (*red*) data of the Landsat–5 TM on 10 November 2000, with two estimation maps of the concentration of NH_4–N by (**b**) multivariate standardized ordinary cokriging and (**c**) single variable ordinary kriging in November 2000. The turbidity of the sea water, shown by the unclearness of the blue color, was found to have a correlation with the NH_4–N concentration at the stations shown in Fig. 1. (Colour figure online)

almost uniformly high. The other point of difference is that the concentrations obtained by the cokriging model are uniformly low except for the high concentration zones noted above, while the concentrations adjacent to the high zones obtained by the kriging model are higher than those of the cokriging model. This trend may result from the smoothing effect, in which the high concentration data influence the kriging estimates more strongly and widely. Those two characteristics in the cokriging model correspond with the turbidity trends that appeared in the color composite image. Consequently, it can be said that multivariate cokriging can produce estimations better than single variable ordinary kriging.

Conclusions

In this paper, variogram modeling by tensor product cubic smoothing surface and multivariate space-time standardized ordinary cokriging were proposed for improving the availability of multivariate geostatistics. The main results obtained are summarized as follows:

1. The tensor product cubic smoothing surface method is effective in modeling experimental space-time semivariograms because this method can balance between the closeness of the experimental data and the smoothness of the model. Further, when compared with the product-sum method, it can follow reasonably complicated space-time correlation structures.
2. Multivariate space-time standardized ordinary cokriging was formulated to consider irregular absences of data where patterns may change depending on the type of data. The advantages of this method over single variable ordinary kriging were demonstrated by a case study of a closed sea environment. The cokriging model expressed local changes of the concentrations of nutritive salts while overcoming the smoothing effect and filled in the blank areas where the ordinary kriging could not estimate because of the lack of data regarding the target component.
3. The validity of the cokriging model was confirmed because the correlation coefficients in the cross-validation were relatively high: NO_2–N = 0.79, NO_3–N = 0.80, NH_4–N = 0.70, and PO_4–P = 0.77. Furthermore, the sea color shown by the color composite image from the three visible band data of the Landsat–5 TM corresponded well with the NH_4–N concentration model, which may be related to the water turbidity. From the cokriging models, environmental changes, such as the recent increases observed in June in the northern sea for the NO_3–N and NH_4–N concentrations, were clarified.

Acknowledgements The authors express their sincere thanks to Prof. Kiyoshi Takikawa and Dr. Kenji Tanaka of Kumamoto University for their valuable suggestions and discussions regarding the data analysis. Sincere thanks are also extended to the Kumamoto Prefectural Fisheries Research Center, the Saga Prefectural Ariakekai Fisheries Research and Development Center, and the Fukuoka Fishers and Marine Technology Research Center for providing the environmental data of the Ariake Sea.

References

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V., 2000, Logratio analysis and compositional distance: Math. Geol., v. 32, no. 3, p. 271–275.

Christakos, G., and Hristopulos, D.T., 1998, Spatiotemporal environmental health modeling—a tractatus stochasticus: Kluwer Academic Press, Boston, 400 p.

Cressie, N., and Huang, H., 1999, Classes of nonseparable, spatial-temporal stationary covariance functions: Jour. Am. Stat. Assoc., v. 94, no. 448, p. 1330–1340.

De Cesare, L., Mayer, D., and Posa, D., 2001, Estimating and modeling space-time correlation structures: Stat. Probab. Lett., v. 52, no. 1, p. 21–28.

De Iaco, S., Myers, E.D., and Posa, D., 2002a, Nonseparable space-time covariance models: some parametric families: Math. Geol., v. 34, no. 1, p. 23–42.

De Iaco, S., Myers, E.D., and Posa, D., 2002b, Space-time variograms and a functional form for total air pollution measurements: Comput. Stat. & Data Analysis, v. 41, no. 2, p. 311–328.

De Iaco, S., Myers, E.D., and Posa, D., 2003, The linear coregionalization model and the product-sum space-time variogram: Math. Geol., v. 35, no. 1, p. 25–38.

Deutsch, C.V., and Journel, A.G., 1998, GSLIB: Geostatistical software library and user's guide: Oxford University Press, New York, 368 p.

Goovaerts, P., and Sonnet, P., 1993, Study of spatial and temporal variations of hydrogeochemical variables using factorial kriging analysis, *in* Soares, A., ed., Geostatistics Troia'92, v. 2: Kluwer Academic Publishers, Dordrecht, p. 745–756.

Goovaerts, P., 1993, Spatial orthogonality of the principle components computed from coregionalized variables: Math. Geol., v. 25, no. 3, p. 281–302.

Koike, K., Narikiyo, K., and Matsuda, S., 2002, Spatio-temporal distribution modeling of iron component concentrations in groundwater using geostatistics: Jpn. J. Appl. Stat., v. 31, no. 1, p. 41–58.

Koike, K., Liu, C.X., and Tamura, A., 2005, Application of spatio-temporal geostatistics to detection of secular changes of environments in a closed sea: Geoinformatics, v. 16, no. 1, p. 3–15.

Liu, C.X., Tamura, A., and Koike, K., 2004, Space-time geostatistical modeling for sea environment—with a case study with Ariake Sea, *in* Matsui, K., Shimada, H., and Gottfried, J., eds., Proceedings of 2nd International Workshop on Earth Science and Technology: Kyushu University, Fukuoka, Japan, p. 233–240.

Martín-Fernández, J.A., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 2003, Dealing with zeros and missing values in compositional data sets using nonparametric imputation: Math. Geol., v. 35, no. 3, p. 253–278.

Pawlowsky-Glahn, V., and Egozcue, J.J., 2002, BLU estimators and compositional data: Math. Geol., v. 34, no. 3, p. 259–274.

Rouhani, S., and Wackernagel, H., 1990, Multivariate geostatistical approach to space-time data analysis: Water Resources Res., v. 36, no. 4, p. 585–591.

Takikawa, K., Tanaka, K., Aoyama, C., Nishioka, R., and Hokamura, T., 2003, Long-term variation of water quality variables in Ariake sea over 25 years: Asian WATERQUAL2003 (IWA-Asia Pacific Regional Conf.), Oct. 19–23, 2003, Bangkok, Thailand, p. 1–8 (Pres. ID: 2QHD08).