

LINEAR DISCRIMINANT ANALYSIS OF MULTIVARIATE SPATIAL-TEMPORAL REGRESSIONS

JŪRATĖ ŠALTYTĖ-BENTH AND KĘSTUTIS DUČINSKAS

ABSTRACT. We consider classification of a realization of the multivariate spatial-temporal Gaussian random field into one of two populations with different regression mean models and factorized covariance matrices. Unknown means and common feature vector covariance matrix are estimated from training samples with observations correlated in space and time, assuming spatial-temporal correlations to be known. We present the first-order asymptotic expansion of the expected error rate associated with linear plug-in discriminant function. Our results are applied to (ecological) data collected from the Lithuanian economical zone in the Baltic sea.

1. INTRODUCTION

Ordinary discriminant analysis (DA) assumes that observations of feature vectors at different locations are independent and, most often, identically distributed. However, in practical situations it is usually not so. Data that are close together in time and/or space, are likely to be correlated. Thus, the classification problem, with both temporal and spatial dependencies, is very important. For example, in weather prediction the weather may be divided into three classes: fair, rain and possible rain. Then the problem is to classify tomorrow's weather into one of these three classes on the basis of data from satellites, where weather masses are observed. This is an example of DA which is often used in, for example, pattern recognition.

When classes are completely specified and feature vectors are independent, an optimal classification rule in the sense of minimum misclassification probability is the Bayesian classification rule (BCR) (Anderson, 1958). In practice, however, the complete description of classes is usually not possible and for the estimation of probabilistic characteristics of each class training samples are required. When estimators of unknown parameters are used, the expressions for the expected error rate are very cumbersome even for the simplest procedures of DA. This makes it difficult to build some qualitative conclusions. Therefore, asymptotic expansions of the expected error rate are especially important.

Date: 21 July 2003.

Key words and phrases. Asymptotic expansion, Bayes classification rule, error rate, spatial-temporal.

We are grateful to the Center of Marine Research (CMR) in Klaipėda (Lithuania) for kindly providing us with data. Rasa Jokūbauskaitė, head of Ecotoxicological Department at CMR, and Ovidijus Stulpinas, head of Hydrological Department at CMR, are thanked for consultations. Jūratė Šaltytė-Benth is supported by the Norwegian Research Council under grant NFR: 155120/432.

Many authors have investigated the performance of the plug-in version of the BCR when parameters are estimated from training samples with independent observations or time series observations (see, e.g., Okamoto, 1963; Lawoko and McLachlan, 1985; Dučinskas, 1997). A good review of works in this field can be found in McLachlan (1992). The pioneering work on the classification of spatial data is Switzer's (1980) paper. An extension of Switzer's work is presented in Mardia (1984). However, neither of these authors analyse the error rate of classification. The asymptotic expansion of the expected error rate in spatial context was investigated in, e.g., Šaltytė (2001), Dučinskas and Šaltytė (2002).

In this paper, we consider the performance of the plug-in linear discriminant function (LDF) when the parameters are estimated from training samples being realizations of spatial-temporal Gaussian random field. We use the maximum likelihood (ML) estimators of unknown means and common feature vector covariance matrix assuming spatial-temporal correlation function to be known. We find an asymptotic expansion of the expected error rate associated with linear plug-in LDF up to the order $O(N^{-2})$, where $N = N_1 + N_2$ and N_l , $l = 1, 2$, is the size of training sample. Also we apply the classification procedure and obtained asymptotic expansion of the expected error rate to (ecological) data collected in the Lithuanian economical zone in the Baltic sea.

2. MODEL AND DEFINITIONS

Spatial-temporal data can be considered as a realization of a random field

$$\{Z(\mathbf{s}; t) : \mathbf{s} \in D, t \in [0, \infty)\},$$

where \mathbf{s} and t define spatial and temporal coordinates, respectively.

Suppose that the model of $Z(\mathbf{s}; t)$ in population Ω_l is

$$Z(\mathbf{s}; t) = B_l^T x(\mathbf{s}; t) + \varepsilon_l(\mathbf{s}; t),$$

where $x(\mathbf{s}; t) = (x_1(\mathbf{s}; t), \dots, x_q(\mathbf{s}; t))^T$ is a $q \times 1$ vector of nonrandom regressors and B_l , $l = 1, 2$, is the unknown parameter matrix of order $q \times p$. Assume that

$$\{\varepsilon_l(\mathbf{s}; t) : \mathbf{s} \in D \subset R^2, t \in [0, \infty)\}$$

is a p -variate zero-mean intrinsically stationary spatial-temporal Gaussian random field with stationary (in space and time) spatial-temporal covariance function defined by a parametric model

$$\text{cov}\{\varepsilon_l(\mathbf{s}; t), \varepsilon_l(\mathbf{s} + \mathbf{h}_s; t + h_t)\} = \sigma(\mathbf{h}_s, h_t; \theta_l) = \sigma_l(\mathbf{h})$$

for all $\mathbf{s}, \mathbf{s} + \mathbf{h}_s \in D$, $t, t + h_t > 0$, where $\mathbf{h} = (\mathbf{h}_s, h_t)^T$, $\theta_l = (\theta_l^s, \theta_l^t)^T \in \Theta$ is a $k \times 1$ parameter vector, Θ being an open subset of R^k , and θ_l^s , θ_l^t subvectors of parameters of spatial and temporal covariance function components, respectively, $l = 1, 2$. Here we restrict our attention to the homoscedastic models, i.e. to models with $\sigma_l(\mathbf{0}) = \Sigma$ for each $\theta_l \in \Theta$, $l = 1, 2$. Then, in Ω_l the mean function at location \mathbf{s} and time t is

$$\mu_l(\mathbf{s}; t) = B_l^T x(\mathbf{s}; t)$$

and the spatial-temporal covariance function is

$$\sigma_l(\mathbf{h}) = c_l(\mathbf{h}) \Sigma,$$

where $c_l(\mathbf{h}) = c(\mathbf{h}; \theta_l)$ is the spatial-temporal correlation function, $l = 1, 2$. It is assumed that the function $c_l(\mathbf{h})$ is positive definite (Cressie and Huang, 1999).

Assume that, for all $\mathbf{s}, \mathbf{s} + \mathbf{h}_s \in D$, $t, t + h_t > 0$, $\mathbf{h}_s \neq 0$, $h_t \neq 0$,

$$(2.1) \quad \text{cov} \{ \varepsilon_1(\mathbf{s}; t), \varepsilon_2(\mathbf{s} + \mathbf{h}_s; t + h_t) \} = 0.$$

Consider the problem of classification of p -variate observation vector $Z^0 = Z(\mathbf{s}_0; t_0)$ with $\mathbf{s}_0 \in D_0 \subset D$, $t_0 > 0$, into one of two populations specified above. Denote by π_1^0 and π_2^0 the prior probabilities of populations Ω_1 and Ω_2 , respectively ($\pi_1^0 + \pi_2^0 = 1$). Here the superscript 0 is used to emphasize that prior probabilities can depend on the position (in space and time) of the observation to be classified. Under the assumption that the populations are completely specified and for known prior probabilities, the BCR $d_B(\cdot)$ minimizing the probability of misclassification (PMC) is

$$(2.2) \quad d_B(z^0) = \arg \max_{\{l=1,2\}} \pi_l^0 p_l(z^0),$$

where z^0 is a realization of Z^0 and

$$p_l(z^0) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} (z^0 - \mu_l^0)^T \Sigma^{-1} (z^0 - \mu_l^0) \right)$$

is a probability density function (p.d.f.) of Z^0 in Ω_l , $l = 1, 2$. Here $\mu_l^0 = \mu_l(\mathbf{s}_0; t_0) = B_l^T x^0$, with $x^0 = x(\mathbf{s}_0; t_0)$, is a p -variate mean vector and Σ is a covariance matrix of Z^0 .

Denote by P_B the PMC of BCR, which is often called Bayes error rate.

In practical applications the parameters of the p.d.f. are usually not known. Then the estimators of unknown parameters can be found from training samples T_1 and T_2 taken separately from Ω_1 and Ω_2 , respectively. When estimators of unknown parameters are used, the plug-in version of BCR is obtained.

Suppose that the spatial-temporal random field is observed at $N = N_1 + N_2$ spatial-temporal coordinates in region $D_1 \subset D$, i.e. we observe the training sample $T = (T_1^T, T_2^T)^T$, where T_l is the $N_l \times p$ matrix of N_l observations of p -variate $Z(\mathbf{s}, t)$ from Ω_l , $l = 1, 2$. Then T is the $N \times p$ matrix.

Assume that D_1 is beyond the zone of influence of D_0 . Then Z^0 is independent on T .

Let \widehat{B}_1 , \widehat{B}_2 and $\widehat{\Sigma}$ be the estimators of B_1 , B_2 and Σ , respectively, based on T , and let $\widehat{\mu}_l(\mathbf{s}; t) = \widehat{B}_l^T x(\mathbf{s}; t)$, $l = 1, 2$. Similarly, $\widehat{\mu}_l^0$ is the estimator of μ_l^0 , $l = 1, 2$. Put $\Psi = (\mu_1^0, \mu_2^0, \Sigma)$ and $\widehat{\Psi} = (\widehat{\mu}_1^0, \widehat{\mu}_2^0, \widehat{\Sigma})$.

The plug-in rule $d_B(z^0; \widehat{\Psi})$ is obtained by replacing the parameters in (2.2) with their estimators. Then the corresponding sample LDF is defined as

$$W(z^0; \widehat{\Psi}) = \left(z^0 - \frac{1}{2} (\widehat{\mu}_1^0 + \widehat{\mu}_2^0) \right)^T \widehat{\Sigma}^{-1} (\widehat{\mu}_1^0 - \widehat{\mu}_2^0) + \gamma,$$

where $\gamma = \ln \frac{\pi_1^0}{\pi_2^0}$.

Definition 2.1. The *actual error rate* for $d_B(z^0; \widehat{\Psi})$ is defined as

$$P(\widehat{\Psi}) = \sum_{l=1}^2 \pi_l^0 \int \left(1 - \delta(l, d_B(z^0; \widehat{\Psi})) \right) p_l(z^0; \Psi) dz^0,$$

where $\delta(\cdot, \cdot)$ is the Kronecker delta.

In the considered case the actual error rate for $d_B(z^0; \widehat{\Psi})$ can be rewritten as

$$P(\widehat{\Psi}) = \sum_{l=1}^2 \pi_l^0 \Phi \left((-1)^l \frac{(\mu_l^0 - \frac{1}{2}(\widehat{\mu}_1^0 + \widehat{\mu}_2^0))^T \widehat{\Sigma}^{-1} (\widehat{\mu}_1^0 - \widehat{\mu}_2^0) + \gamma}{\sqrt{(\widehat{\mu}_1^0 - \widehat{\mu}_2^0)^T \widehat{\Sigma}^{-1} \Sigma \widehat{\Sigma}^{-1} (\widehat{\mu}_1^0 - \widehat{\mu}_2^0)}} \right),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Definition 2.2. The expectation of the actual error rate with respect to the distribution of T , designated as $E_T \left\{ P(\widehat{\Psi}) \right\}$, is called the *expected error rate* (EER) for the $d_B(z^0; \widehat{\Psi})$.

In the following section we find the ML estimators for means and covariance matrix and derive the asymptotic expansion for the EER.

3. MAIN RESULTS

Denote by C_l the spatial-temporal correlation matrix of order $N_l \times N_l$ with elements $c_l(\mathbf{h})$ and suppose that C_l is known, $l = 1, 2$. Let X_l be an $N_l \times q$ regressor matrix of the training sample T_l , $l = 1, 2$. Then for $l = 1, 2$ the model of T_l is

$$T_l = X_l B_l + E_l,$$

where E_l is the $N_l \times p$ matrix of random errors that are distributed according to a matrix multivariate normal distribution $N_{N_l \times p}(0, C_l \otimes \Sigma)$ (see Kai-Tai and Yao-Ting, 1990, chapter 3). So we can write $T_l \sim N_{N_l \times p}(X_l B_l, C_l \otimes \Sigma)$.

Lemma 3.1. For $l = 1, 2$, ML estimators of B_l and Σ based on T are

$$(3.1) \quad \widehat{B}_l^{ML} = (X_l^T C_l^{-1} X_l)^{-1} X_l^T C_l^{-1} T_l$$

and

$$(3.2) \quad \widehat{\Sigma}_{ML} = \frac{1}{N} \sum_{l=1}^2 \left(T_l - X_l \widehat{B}_l^{ML} \right)^T C_l^{-1} \left(T_l - X_l \widehat{B}_l^{ML} \right).$$

Proof. The log-likelihood of T_l is

$$L(B_l, \Sigma) = \text{const} - \frac{1}{2} (p \ln |C_l| + N_l \ln |\Sigma|) - \frac{1}{2} \text{tr} \left(C_l^{-1} (T_l - X_l B_l) \Sigma^{-1} (T_l - X_l B_l)^T \right).$$

After setting the partial derivatives of $L(B_l, \Sigma)$ with respect to B_l equal to zero we obtain normal equations

$$X_l^T C_l^{-1} X_l B_l \Sigma^{-1} = X_l^T C_l^{-1} T_l \Sigma^{-1},$$

$l = 1, 2$, which lead us to (3.1).

Now solving the matrix equation $\sum_{l=1}^2 \frac{\partial \ln L_l}{\partial \Sigma} = 0$ we obtain (3.2). Taking matrix derivatives we used results from chapter 9 in Magnus and Neudecker (2002). \square

Corollary 3.2. *The ML estimator $\widehat{\Sigma}_{ML}$ is a biased estimator and*

$$(3.3) \quad E \left\{ \widehat{\Sigma}_{ML} \right\} = \frac{N - 2q}{N} \Sigma.$$

Proof. The estimator $\widehat{\Sigma}_{ML}$ given in (3.2) can be rewritten in the following way

$$\widehat{\Sigma}_{ML} = \frac{1}{N} \sum_{l=1}^2 T_l^T (I - P_l)^T C_l^{-1} (I - P_l) T_l,$$

where

$$P_l = X_l (X_l^T C_l^{-1} X_l)^{-1} X_l^T C_l^{-1}$$

is an idempotent matrix, $l = 1, 2$. Using properties of idempotent matrices and Gaussian random vectors (see, e.g., Magnus and Neudecker, 2002), we obtain (3.3). Thus, further we will use the bias adjusted estimator

$$\widehat{\Sigma} = \frac{N}{N - 2q} \widehat{\Sigma}_{ML}.$$

\square

Put $\Delta \widehat{\mu}_l^0 = \widehat{\mu}_l^0 - \mu_l^0 = \left(\widehat{B}_l^{ML} - B_l \right)^T x^0$, $l = 1, 2$, and $\Delta \widehat{\Sigma} = \widehat{\Sigma} - \Sigma$. Let $\varphi(\cdot)$ be the standard normal p.d.f. Denote by $P_l^{(1)} = \partial P \left(\widehat{\Psi} \right) / \partial \widehat{\mu}_l^0$, $P_{l,k}^{(2)} = \partial^2 P \left(\widehat{\Psi} \right) / \partial \widehat{\mu}_l^0 \partial \left(\widehat{\mu}_k^0 \right)^T$, $P_{\widehat{\sigma}_{ij}}^{(1)} = \partial P \left(\widehat{\Psi} \right) / \partial \widehat{\sigma}_{ij}$, $P_{\widehat{\sigma}_{ij}, \widehat{\sigma}_{mn}}^{(2)} = \partial^2 P \left(\widehat{\Psi} \right) / \partial \widehat{\sigma}_{ij} \partial \widehat{\sigma}_{mn}$ and $P_{l,m, \widehat{\sigma}_{ij}}^{(2)} = \partial^2 P / \left(\widehat{\Psi} \right) \partial \widehat{\mu}_{lm}^0 \partial \widehat{\sigma}_{ij}$ the partial derivatives up to second order of $P \left(\widehat{\Psi} \right)$ with respect to the corresponding parameters evaluated at $\widehat{\mu}_1^0 = \mu_1^0$, $\widehat{\mu}_2^0 = \mu_2^0$ and $\widehat{\Sigma} = \Sigma$, where $\widehat{\mu}_{lm}^0$ denotes the m component of the vector $\widehat{\mu}_l^0$ and $\widehat{\sigma}_{ij}$ is ij element of the matrix $\widehat{\Sigma}$, $l, k = 1, 2$, $i, j, m, n = 1, \dots, p$.

Let $\lambda_{N_l}(C_l)$ be the largest eigenvalue of C_l , $l = 1, 2$. We make the following assumptions:

(A1) $(x^0)^T (X_l^T X_l) x^0 = O\left(\frac{1}{N_l}\right)$, as $N_l \rightarrow \infty$.

(A2) $\lambda_{N_l}(C_l) < \nu_l$, $\nu_l < \infty$, as $N_l \rightarrow \infty$, $l = 1, 2$.

(A3) $\frac{N_1}{N_2} \rightarrow v$, as $N_1, N_2 \rightarrow \infty$, $0 < v < \infty$, $l = 1, 2$.

Theorem 3.3. *Suppose (A1)-(A3) hold for training samples T_1, T_2 . Then the asymptotic expansion of the expected error rate for the $d_B(z^0; \hat{\Psi})$ is*

$$E_T \left\{ P(\hat{\Psi}) \right\} = P_B + \sum_{l=1}^2 a_l + \frac{b}{N-2q} + O(N^{-2}),$$

where

$$P_B = \sum_{l=1}^2 \pi_l^0 \Phi\left(-\frac{\Delta}{2} + (-1)^{l+1} \frac{\gamma}{\Delta}\right),$$

$$a_l = \frac{1}{2\Delta} \pi_1^0 \varphi\left(-\frac{\Delta}{2} - \frac{\gamma}{\Delta}\right) \left(\frac{\Delta}{2} + (-1)^{l+1} \frac{\gamma}{\Delta}\right)^2 (x^0)^T (X_l^T C_l^{-1} X_l)^{-1} x^0,$$

$$b = \pi_1^0 \varphi\left(-\frac{\Delta}{2} - \frac{\gamma}{\Delta}\right) \left(\frac{\gamma^2}{\Delta} + (p-1)\Delta\right)$$

and

$$\Delta = \sqrt{(\mu_1^0 - \mu_2^0)^T \Sigma^{-1} (\mu_1^0 - \mu_2^0)}$$

is the Mahalanobis distance, $l = 1, 2$.

Proof. Since $P(\hat{\Psi})$ is invariant under linear transformations of data we use the more convenient canonical form of

$$\mu_l^0 = (-1)^{l+1} \frac{\Delta}{2} \mathbf{1}_0,$$

$$(3.4) \quad \Sigma = I,$$

where $\mathbf{1}_0$ is a p -variate vector of zeroes except first coordinate which is equal to 1, $l = 1, 2$.

Expand $P(\hat{\Psi})$ in a Taylor series about the point $\mu_1^0 = \frac{\Delta}{2} \mathbf{1}_0$, $\mu_2^0 = -\frac{\Delta}{2} \mathbf{1}_0$, $\Sigma = I$. Taking the expectation with respect to the distribution of T and dropping the third order terms we have

$$E_T \left\{ P(\hat{\Psi}) \right\} = P_B + \sum_{l=1}^2 \left(P_l^{(1)} \right)^T E_T \left\{ \Delta \hat{\mu}_l^0 \right\} + \sum_{i,j=1}^p \left(P_{\hat{\sigma}_{ij}}^{(1)} \right)^T E_T \left\{ \Delta \hat{\sigma}_{ij} \right\} +$$

$$+ \frac{1}{2} \sum_{l,k=1}^2 \text{tr} \left(P_{l,k}^{(2)} E_T \left\{ \Delta \hat{\mu}_l^0 (\Delta \hat{\mu}_k^0)^T \right\} \right) +$$

$$(3.5) \quad \begin{aligned} & + \frac{1}{2} \sum_{i,j,m,n=1}^p \left(P_{\hat{\sigma}_{ij}, \hat{\sigma}_{mn}}^{(2)} \right)^T E_T \{ \Delta \hat{\sigma}_{ij} \Delta \hat{\sigma}_{mn} \} + \\ & + \frac{1}{2} \sum_{l,k=1}^2 \sum_{i,j,m=1}^p \left(P_{l_m, \hat{\sigma}_{ij}}^{(2)} \right)^T E_T \{ \Delta \hat{\sigma}_{ij} \Delta \hat{\mu}_{lm}^0 \}. \end{aligned}$$

Since $P(\hat{\Psi})$ is minimized at (3.4), then, for $l = 1, 2$,

$$(3.6) \quad P_l^{(1)} = \mathbf{0}_p,$$

where $\mathbf{0}_p$ is p -dimensional vector of zeroes, and

$$(3.7) \quad P_{\hat{\sigma}_{ij}}^{(1)} = 0.$$

Using Lemma 3.1 we get that, for $l = 1, 2$,

$$(3.8) \quad E_T \left\{ \Delta \hat{\mu}_l^0 (\Delta \hat{\mu}_l^0)^T \right\} = (x^0)^T (X_l^T C_l^{-1} X_l)^{-1} x^0 \cdot \Sigma.$$

From (2.1) it follows, that

$$(3.9) \quad E_T \left\{ \Delta \hat{\mu}_1^0 (\Delta \hat{\mu}_2^0)^T \right\} = \mathbf{0}_{p \times p},$$

where $\mathbf{0}_{p \times p}$ is $p \times p$ matrix of zeroes. Also

$$(3.10) \quad E(\Delta \hat{\sigma}_{ij} \Delta \hat{\sigma}_{mn}) = \frac{1}{N - 2q} (\sigma_{im} \sigma_{jn} + \sigma_{in} \sigma_{jm})$$

and

$$(3.11) \quad E_T \left\{ \Delta \hat{\sigma}_{ij} \Delta \hat{\mu}_{lm}^0 \right\} = 0,$$

because of the properties of Gaussian variables (Magnus and Neudecker, 2002).

Note that

$$(3.12) \quad P_{l,l}^{(2)} = \frac{\pi_1^0}{\Delta} \varphi \left(-\frac{\Delta}{2} - \frac{\gamma}{\Delta} \right) \left(I + \left(\left(\frac{\Delta}{2} + (-1)^{l+1} \frac{\gamma}{\Delta} \right)^2 - 1 \right) 1_0 1_0^T \right),$$

for $l = 1, 2$.

Using assumptions (A1) and (A2) and the inequality

$$(x^0)^T (X_l^T C_l^{-1} X_l)^{-1} x^0 < \lambda_{N_l}(C_l) \cdot (x^0)^T (X_l^T X_l)^{-1} x^0,$$

we can conclude that $a_l = O\left(\frac{1}{N_l}\right)$, as $N_l \rightarrow \infty$, $l = 1, 2$.

Under assumptions (A1)-(A3) the residual term of (3.5) is of order $O(N^{-2})$. Then putting (3.6)-(3.12) into (3.5) we complete the proof of the theorem. \square

As the contribution from higher order terms in the presented asymptotic expansion is negligible, we use (for the evaluation of the performance of LDF) the asymptotic expected error regret

$$AEE R = \sum_{l=1}^2 a_l + \frac{b}{N - 2q},$$

where a_l , $l = 1, 2$, and b are defined earlier. Minimum of $AEE R$ could also be used as a criterion for optimal training sample design.

Remark. In the case of (spatial) Gaussian random field with multivariate constant mean the asymptotic expansion of EER is (Dučinskas and Šaltytė, 2002):

$$E_T \left\{ P \left(\widehat{\Psi} \right) \right\} = P_B + \sum_{l=1}^2 \frac{\alpha_l}{c_l^{\ddot{}}} + \frac{\beta}{N - 2} + O(N^{-2}),$$

where

$$\begin{aligned} \alpha_l &= \pi_1^0 \varphi \left(-\frac{\gamma}{\Delta} - \frac{\Delta}{2} \right) \left(p - 1 + \left(-\frac{\Delta}{2} + (-1)^l \frac{\gamma}{\Delta} \right) \right) / (2\Delta), \\ \beta &= \pi_1 \varphi \left(-\frac{\gamma}{\Delta} - \frac{\Delta}{2} \right) \left(\frac{\gamma^2}{\Delta} + (p - 1) \frac{\Delta}{2} \right) \end{aligned}$$

and

$$c_l^{\ddot{}} = 1_{N_l}^T C_l^{-1} 1_{N_l},$$

here 1_{N_l} is N_l -dimensional vector of ones, $l = 1, 2$.

In the last section we apply the classification procedure used in this paper and calculate $AEE R$ values to a real data set.

4. APPLICATION

Many different factors affect the quality of water, and herewith the conditions for species living there. Here we analyse quarterly data collected in the Baltic sea in Lithuanian economical zone in the period 1994-2001. The factors we are interesting in are amount of oxygen (ml/l), salinity (PSU) and activity of the hydrogen ions (pH) in the water. According to hydrological laws for coastal and open sea waters, it is reasonable to distinguish the following groups of stations where measurements are performed: coastal zone, dumping site and Klaipėda transect (map of stations can be found in, for example, Report, 2002). Stations in the Klaipėda transect are between coastal zone and dumping site and are spread in different distances towards the open sea. It may be interesting to see properties of which zone are more common to them. Thus, we say that there are two classes of stations (coastal zone and dumping site) at which the three mentioned above factors are measured.

Measurements are made at different depths. As in this paper the influence of depth on factors is not investigated, we deal only with the observations made at 1 meter depth.

We have 32 time periods when measurements in each station were performed. In the coastal zone (Ω_1) there are 10 stations, and in the dumping site (Ω_2) only 3. In the Klaipėda transect we have 4 stations; these are stations to be classified. Thus $N_1 = 320$, $N_2 = 96$

and $p = 3$. We assume that the data are realizations of the spatial-temporal Gaussian random field.

In this paper we deal with a special case of regression mean model, namely the first-order trend surface model (Haining, 1990). There was no temporal trend detected, so we say that the regressor matrix for each time period is the same.

In our model we use the assumption of known spatial-temporal correlation matrices. However, in practice this is usually not the case. Therefore, we perform further calculations with correlation matrices obtained by using parametric spatial-temporal correlation functions fitted to empirical ones. Fitting of correlation function in spatial-temporal context is more difficult than in the spatial case. Thus, we consider spatial-temporal correlation models which are separable in a multiplicative sense (Cressie and Huang, 1999; De Cesare *et.al.*, 2001), i.e.

$$c_l(\mathbf{h}) = c_l(\mathbf{h}_s) c_l(h_t),$$

where $c_l(\mathbf{h}_s) = c(\mathbf{h}_s; \theta_l^s)$ and $c_l(h_t) = c(h_t; \theta_l^t)$, $l = 1, 2$, are spatial and temporal correlation functions, respectively.

According to our model, the multivariate structure of a set of features is independent on the spatial correlation, which is the same for all features. This is a so called intrinsic multivariate correlation, because the correlation between features does not depend upon spatial scale (Wackernagel, 1995). This fact we use when estimating the spatial correlation function for our data. We use the following procedure. For every pair of locations (stations) in geographical space a correlation value between two time series can be calculated. This we do for all three features separately. After plotting all obtained empirical correlations vs. distance we easily fit parametrical model by using the *nlinfit* function in Matlab.

Here we consider the following spatial correlation functions (see, e.g., Journel and Huijbregts, 1978):

Spherical

$$c_l^S(|\mathbf{h}_s|) = \begin{cases} 1, & |\mathbf{h}_s| = 0, \\ \frac{\theta_{l1}^s}{\theta_{l0}^s + \theta_{l1}^s} \left(1 - \frac{3}{2} \frac{|\mathbf{h}_s|}{\theta_{l2}^s} + \frac{1}{2} \left(\frac{|\mathbf{h}_s|}{\theta_{l2}^s} \right)^3 \right), & 0 < |\mathbf{h}_s| < \theta_{l2}^s, \\ 0, & |\mathbf{h}_s| \geq \theta_{l2}^s, \end{cases}$$

where θ_{l0}^s is the nugget effect, $\theta_{l0}^s + \theta_{l1}^s$ - sill and θ_{l2}^s - range of the spherical correlation function.

Exponential

$$c_l^E(\mathbf{h}_s) = \begin{cases} 1, & \mathbf{h}_s = 0, \\ \frac{\theta_{l1}^s}{\theta_{l0}^s + \theta_{l1}^s} \exp\left(-\frac{1}{\theta_{l2}^s} \sqrt{\mathbf{h}_{s1}^2 + \alpha_l \mathbf{h}_{s2}^2}\right), & \mathbf{h}_s > 0, \end{cases}$$

Gaussian

$$c_l^G(\mathbf{h}_s) = \begin{cases} 1, & \mathbf{h}_s = 0, \\ \frac{\theta_{l1}^s}{\theta_{l0}^s + \theta_{l1}^s} \exp\left(-\frac{1}{\theta_{l2}^s} (\mathbf{h}_{s1}^2 + \alpha_l \mathbf{h}_{s2}^2)\right), & \mathbf{h}_s > 0, \end{cases}$$

TABLE 1. Parameters of spatial correlation functions in Ω_1 and Ω_2

Parameters	S	$E1$	E	$G1$	G
$\frac{\theta_{11}^s}{\theta_{10}^s + \theta_{11}^s}$	0.6617	0.6731	0.7209	0.6202	0.6521
θ_{12}^s	34.3976	0.0546	0.2129	0.0057	0.1154
α_1	—	1	0.0816	1	0.0520
MSE_1	0.0720	0.0718	0.0711	0.0730	0.0719
$\frac{\theta_{11}^s}{\theta_{10}^s + \theta_{11}^s}$	0.9999	0.9659	0.9658	0.9596	0.9597
θ_{22}^s	0.4285	0.0319	0.0105	0.0336	0.0036
α_2	—	1	10.0502	1	10.3926
MSE_2	0.5539	$2.41e - 04$	$2.40e - 04$	$2.45e - 04$	$2.44e - 04$

where $\mathbf{h}_s = (\mathbf{h}_{s1}, \mathbf{h}_{s2})^T$, θ_{l0}^s and $\theta_{l0}^s + \theta_{l1}^s$ are the nugget effect and sill, respectively, θ_{l2}^s in the last two models is parameter describing the rate of change in spatial dependence, and α_l is the parameter of anisotropy, $l = 1, 2$.

Parameters of fitted spatial correlation functions in Ω_1 and Ω_2 are presented in Table 1. We use the following abbreviations: S for spherical, $E1$ for isotropical exponential, E for anisotropical exponential, $G1$ for isotropical Gaussian and G for anisotropical Gaussian correlation function. As criteria for choosing the function which fits best the empirical data the mean square error is used. In Table 1 it is denoted by MSE_1 for Ω_1 and MSE_2 for Ω_2 .

As we see from the Table 1, the anisotropical exponential spatial correlation function for both classes fits data best. Thus in the further analysis we use this function.

In order to fit a time correlation function we calculate autocorrelation functions up to lag 12 for time series at each location. Then we average obtained autocorrelations over all stations in certain class. We repeat this procedure for each feature separately. Plotting all three obtained empirical autocorrelation functions and again using *nlinfit* function we find the best fit. As the behaviour of the empirical autocorrelation function was very similar to the exponential one, we will consider only the exponential temporal correlation function:

$$c_l^E(h_t) = \begin{cases} 1, & h_t = 0, \\ \exp\left(-\frac{1}{\theta_l^t} h_t\right), & h_t > 0, \end{cases}$$

where θ_l^t is parameter describing how fast the temporal correlation function decreases.

Fitted values of parameters for the temporal exponential correlation function are: $\theta_1^t = 0.2796$ and $\theta_2^t = 0.0654$.

Suppose we are interesting in classifying the observations made in spring 2001 at 4 stations in the Klaipėda transect. We denote these 4 stations by letters A, B, C and D . Mahalanobis distances denoted by Δ_K , $K = A, B, C, D$, between each of these stations and classes Ω_1 and Ω_2 are given in Table 2.

TABLE 2. Mahalanobis distances

	Ω_1	Ω_2
Δ_A	3.5636	5.8857
Δ_B	3.7109	5.7976
Δ_C	3.8008	5.6678
Δ_D	3.8717	5.5021

TABLE 3. Classification results and values of $AEEER$

Station	Δ	π_1^0	Class	P_B	$AEEER$	$AEEER_{ind}$
A	1.2530	0.5	2	0.2654	0.1046	0.0931
B	4.3919	0.5	1	0.0140	0.0013	0.0009
C	11.4030	0.5	1	0.0011	0.0001	0.0000
D	20.5466	0.5	1	0.0000	0.0000	0.0000
A	1.2530	0.75	2	0.4657	0.0576	0.0490
B	4.3919	0.75	1	0.0212	0.0014	0.0010
C	11.4030	0.75	1	0.0018	0.0001	0.0000
D	20.5466	0.75	1	0.0000	0.0000	0.0000
A	1.2530	0.38	2	0.3085	0.4360	0.3912
B	4.3919	0.39	1	0.0153	0.0012	0.0008
C	11.4030	0.40	1	0.0001	0.0001	0.0000
D	20.5466	0.41	1	0.0000	0.0000	0.0000

In order to classify these stations, we use the plug-in Bayes classification rule. As we don't know true values of means of both classes and feature vector covariance matrix, we will use their ML estimators from (3.1) and (3.2). The classification results and the values of $AEEER$ are given in Table 3. Here we consider different values of prior probabilities:

1) equal prior probabilities, i.e. $\pi_1^0 = 0.5$; 2) the value of $\pi_1^0 = 0.75$ was chosen as a proportion of number of points in Ω_1 ; and 3) the value π_1^0 proportional to the Mahalanobis distance from the point to be classified to the Ω_1 .

As we see from Table 3 and as one can expect, the Bayes error rate P_B and $AEEER$ decreases when the distance between classes increases. The smallest values (almost zero) of P_B and $AEEER$ are obtained in the case when a station D is classified. Such small values correspond to the big Mahalanobis distance between classes. In other words, big distance between classes means that it is very unlikely to make a mistake when station is classified.

In all considered cases of π_1^0 station A is assigned to the dumping site and the 3 other stations to the coastal zone. In order to obtain the smaller $AEEER$ value for the considered data set one should not use the prior probabilities proportional to the number of points in the corresponding class.

As it was expected, the $AEEER$ values in the case of independent (in time and space) observations (denoted by $AEEER_{ind}$) are the smallest ones. Thus, it is very important take

into consideration the temporal and spatial dependency factor when practical problems are solved.

The same calculations were performed for the observations made at the same 4 stations in the autumn 2001. Even though the numerical values were slightly different, the general tendencies remained the same.

REFERENCES

- [1] Anderson T.W. (1958). *An introduction to multivariate statistical analysis*. Wiley, New York.
- [2] Cressie N., Huang H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc. Theory Methods*, **94**(448), 1330-1340.
- [3] De Cesare L., Myers D.E., Posa D. (2001). Estimating and modeling space-time correlation structures. *Statist. Probab. Lett.*, **51**, 9-14.
- [4] Dučinskas K. (1997). An asymptotic analysis of the regret risk in discriminant analysis under various training schemes. *Lithuanian Math. J.*, **37**(4), 337-351.
- [5] Dučinskas K., Šaltytė J. (2002). The effect of spatial autocorrelation on the error rates of the linear discriminant functions. *Lithuanian Math. J.*, **42**(2), 169-178.
- [6] Haining R.P. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.
- [7] Journel A.G., Huijbregts Ch.J. (1978). *Mining geostatistics*. Academic Press.
- [8] Kai-Tai F., Yao-Ting Z. (1990). *Generalized multivariate analysis*. Springer.
- [9] Lawoko C.R.O., McLachlan G.J. (1985). Discrimination with autocorrelated observations. *Pattern Recognition*, **18**(2), 145-149.
- [10] Magnus J.R., Neudecker H. (2002). *Matrix differential calculus with applications in statistics and econometrics*. Wiley, New York.
- [11] Mardia K.V. (1984). Spatial discrimination and classification maps. *Comm. Statist. Theory Methods*, **13**(18), 2181-2197.
- [12] McLachlan G.J. (1992). *Discriminant analysis and statistical pattern recognition*. Wiley, New York.
- [13] Okamoto. (1963). M. An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.* **34**, 1286-1301.
- [14] Report (2002). Results of oceanographic, hydrochemical, ecotoxicological and hydrobiological investigations of Baltic sea. *Ministry of the Environment of the Republic of Lithuania, Center of Marine Research*.
- [15] Šaltytė J. (2001). The asymptotic expansion of the expected risk for the LDA of spatially correlated Gaussian observations. *PhD thesis*, Klaipėda university.
- [16] Switzer P. (1980). Extension of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Math. Geol.*, **12**(4), 367-376.
- [17] Wackernagel H. (1995). *Multivariate geostatistics. An introduction with applications*. Springer.

(Jūratė Šaltytė-Benth)

CENTRE OF MATHEMATICS FOR APPLICATIONS

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF OSLO

P.O. BOX 1053, BLINDERN

N-0316 OSLO, NORWAY

AND

DEPARTMENT OF SYSTEM RESEARCH

KLAIPĖDA UNIVERSITY

H. MANTO 84

LT-5808 KLAIPĖDA, LITHUANIA

E-mail address: jurate@math.uio.no

(Kęstutis Dučinskas)

DEPARTMENT OF SYSTEM RESEARCH

KLAIPĖDA UNIVERSITY

H. MANTO 84

LT-5808 KLAIPĖDA, LITHUANIA

E-mail address: duce@gmf.ku.lt