Spatial Statistics 28 (2018) 216–225

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Statistical classification of multivariate conditionally autoregressive Gaussian random field observations

L. Dreiziene [a,b,*], K. Ducinskas [a,b], L. Saltyte-Vaisiauske [a]

[a] *Klaipeda University, Klaipeda, Lithuania*
[b] *Vilnius University, Vilnius, Lithuania*

ARTICLE INFO

Article history:
Received 31 October 2017
Accepted 22 March 2018
Available online 29 March 2018

Keywords:
Gaussian Markov random field
Bayes discriminant function
Mahalanobis distance
Actual and expected error rate

ABSTRACT

The problem of supervised classifying of a multivariate Gaussian Markov random field (GMRF) observation into one of two populations specified by different regression mean models is considered. We focus on a multivariate conditionally autoregressive model, a subclass of GMRF with parametrical structure proposed by Pettitt et al. (2002) and generalized by Jin et al. (2005). For complete parametric certainty and with fixed training sample locations, the formula of the Bayes error rate is derived. Plug-in Bayes discriminant function obtained by replacing the regression and scale parameters with their ML estimators in the Bayes discriminant function is investigated. The novel formulas for the actual error rate and the approximation of the expected error rate (AER) associated with plug-in Bayes discriminant function are derived. These results are multivariate generalizations of the univariate ones. A numerical analysis of the derived formulas is implemented for the bivariate conditionally autoregressive model sampling on a regular 2-dimensional unit spacing lattice with the neighborhood structure based on the Euclidean distance. A comparison of the proposed classifier performance for different parametric structures of populations is done by studying the accuracy of the derived AER. Finally, the influence of the neighborhood order on the accuracy of the derived AER is examined.

* Corresponding author at: Klaipeda University, Klaipeda, Lithuania.
E-mail address: l.dreiziene@gmail.com (L. Dreiziene).

https://doi.org/10.1016/j.spasta.2018.03.006

1. Introduction

The problems of discriminant analysis of spatially correlated Gaussian data were considered previously by several authors (e.g. Switzer, 1980; Mardia, 1984; Klein and Press, 1992; McLachlan, 2004). The error rates in classification of Gaussian random field observation via plug-in Bayes discriminant functions (PBDF) are explored by Ducinskas (2009, 2011). Spatial–temporal case has been considered by Saltyte-Benth and Ducinskas (2005). Numerical comparison of the performances for different spatial classification rules is performed by Berret and Calder (2016). However, in the works mentioned above, the main attention is paid to the so-called the geostatistical models (GS) with directly specified Matern type or other parametric covariance function models. It is well known that the models which include covariance matrices require a large number of computer operations. In contrast to GS case GMRF is based on the direct specification of sparse precision matrix. Thus spatial interpolation and classification problems require far fewer calculations.

In the present paper we are concerned with classification problems for a multivariate GMRF observed over a finite lattice. Mardia (1988) introduced a multivariate GMRF, and more recently Jin et al. (2005) and Sain and Cressie (2007) explored these models. We focus on a subclass of a multivariate GMRF so-called a multivariate conditionally autoregressive (MCAR) model with parametrical structure proposed by Pettitt et al. (2002). They extended the univariate spatial model to the multivariate one, maintaining computational simplicity but modeling the essential aspects of dependence between the multivariate components and spatial dependence between sites. De Oliveira and Ferreira (2011) showed that this parametric structure is well-suited to the case of small samples and ensures good frequentist properties of ML estimators of parameters. Their results strengthen our motivation for using these models.

The classifier associated with PBDF obtained by substituting all model parameters by their ML estimators into Bayes' rule is examined. The main purpose of this study is to derive the analytic expressions for Bayes error rate, the actual error rate and the approximation of the expected error rate for the classifier based on PBDF. These functions are used as performance measures of the aforementioned classifier. Note that the obtained results are the multivariate generalizations of the previously explored univariate cases (Ducinskas and Dreiziene, 2016, 2018).

The bivariate GMRF sampled on a regular 2-dimensional unit spacing lattice endowed with the neighborhood structure based on the Euclidean distance between sites is used for a simulation experiment. The sampling properties of the ML estimators of parameters and the accuracy of the derived AER for various values of spatial dependence parameters and Mahalanobis distance are studied. The influence of the neighborhood size on the accuracy of the proposed AER for a particular case of the parametric structure is examined.

This paper is organized as follows: the problem description and definitions of classification error rates are displayed in Section 2; derivation of the AER is presented in Section 3; simulation experiments are done in Section 4; and finally, the conclusions are in Section 5.

2. The main concepts

The main objective of this paper is to classify a single observation of multivariate GMRF $\{Z(s) : s \in D \subset R^2\}$ when the training sample is given. The model of observation $Z(s)$ in population Π_l is

$$Z(s) = B_l' x(s) + \varepsilon(s), \tag{1}$$

where $x(s)$ is a $q \times 1$ vector of non-random regressors and B_l is a $q \times p$ matrix of parameters, $l = 1, 2$. The error term is generated by p-variate zero-mean multivariate stationary GMRF $\{\varepsilon(s) : s \in D\}$ and is specified with respect to the undirected graph (nodes with the neighborhood system) that will be described later.

Suppose that $\{s_i \in D; i = 0, 1, \dots, n\}$ is the set of spatial locations (nodes) where $Z(s)$ observations are taken. Indexing the spatial locations by integers i.e. $s_i = i$, $i = 0, 1, \dots, n$ denote the set of training locations by $S_n = S^{(1)} \cup S^{(2)}$, where $S^{(l)}$ are the subsets of S_n that contains n_l observations of $Z(s)$ from Π_l, $l = 1, 2$, $n = n_1 + n_2$. The location of the observation to be classified is indexed by $\{0\}$.

Assume that lattice $S_n^0 = S_n \cup \{0\}$ is endowed with a neighborhood system, $N^0 = \{N_k^0 : k = 0, 1, \ldots, n\}$ and lattice S_n is endowed with a neighborhood system, $\{N_k : k = 1, \ldots, n\}$ where N_k denotes the collection of sites that are neighbors of site s_k. Then define spatial weight $w_{kl} > 0$ ($w_{kl} = w_{lk}$) as a measure of similarity between sites k and l and let $w_0' = (w_{01}, \ldots, w_{0n})$, $w_i' = (w_{i1}, \ldots, w_{ii-1}, w_{ii+1}, \ldots, w_{in})$, $i = 1, \ldots, n$. Further we construct matrices $H_0 = (h_{kl}^0 : k, l = 0, 1, \ldots, n)$ and $H = (h_{kl} : k, l = 1, \ldots, n)$ with dimensions $(n+1) \times (n+1)$ and $(n \times n)$, respectively. The elements of these matrices are defined as follows

$$h_{kl}^0 = \begin{cases} h_k^0 & \textit{if } k = l \\ -w_{kl} & \textit{if } k \in N_l^0 \\ 0 & \textit{otherwise} \end{cases}, \qquad h_{kl} = \begin{cases} h_k & \textit{if } k = l \\ -w_{kl} & \textit{if } k \in N_l \\ 0 & \textit{otherwise} \end{cases},$$

$$\text{where } h_k^0 = \sum_{l \in N_k^0} w_{kl},\ k, l = 0, 1, \ldots, n \ \text{ and } h_k = \sum_{l \in N_k} w_{kl},\ k, l = 1, \ldots, n.$$

The main objective of this paper is to classify a single observation of multivariate GMRF $\{Z(s) : s \in D \subset R^2\}$ specified on lattice S_n^0.

For simplicity, we put $Z(s_i) = Z_i$, $\varepsilon(s_i) = \varepsilon_i$, $x(s_i) = x_i$, $i = 0, \ldots, n$ and then let $\varepsilon_{-0} = \left(\varepsilon_1', \ldots, \varepsilon_n'\right)'$, $Z_{-0} = \left(Z_1', \ldots, Z_n'\right)'$ and for $i = 1, \ldots, n$ $\varepsilon_{-i} = \left(\varepsilon_1', \ldots, \varepsilon_{i-1}', \varepsilon_{i+1}', \ldots \varepsilon_n'\right)'$ and $Z_{-i} = \left(Z_1', \ldots, Z_{i-1}', Z_{i+1}', \ldots, Z_n'\right)'$.

Suppose that Λ denotes the $p \times p$ correlation type matrix with ones on the diagonal and off-diagonal entries $\{-\lambda_{ij}\}$ but plays a role of a precision matrix. Hence putting the matrix element equal to zero, gives conditional independence between the components of Z. The full conditionals for $i = 0, \ldots, n$ are specified as

$$\varepsilon_i | \varepsilon_{-i} \sim N(\mu_l^i, \Sigma^i), \tag{2}$$

$$\mu_l^i = \left(\alpha_i' \otimes I_p\right) \varepsilon_{-i}, l = 1, 2,\ \Sigma^i = \rho_i \Lambda^{-1}. \tag{3}$$

Here $\alpha_i' = \alpha w_i'/(1 + \alpha h_i)$ and $\rho_i = \sigma^2/(1 + \alpha h_i)$. $\alpha \geq 0$ is a spatial dependence parameter and $\sigma > 0$ is a scale parameter. Then the covariance matrix of vector $Z = (Z_0', Z_1', \ldots, Z_n')$ is $var(Z) = \sigma^2 \left(I_{n+1} + \alpha H^0\right)^{-1} \otimes \Lambda^{-1}$.

Denote a training sample in vector form by $T = Z_{-0}$ and in matrix form by $T^* = (Z_1, \ldots, Z_n)'$. Design matrix X for training sample T is specified by $X_1 \oplus X_2$, where X_l is the $n_l \times q$ matrix of regressors for training sample observations from Π_l, $l = 1, 2$. Then under some regularity conditions (Mardia, 1988), the joint distribution for training sample in vector form is $T \sim N_{np}(vec\left(B'X'\right), \sigma^2 R(\alpha) \otimes \Lambda^{-1})$ and in matrix form follows $n \times p$ matrix Gaussian distribution $T^* \sim N_{p \times n}(XB, \sigma^2 R(\alpha) \otimes \Lambda^{-1})$, where $B' = (B_1', B_2')$ and $R(\alpha) = (I_n + \alpha H)^{-1}$ denotes the spatial correlation matrix for T.

Consider the problem of classification of observation vector Z at location s_0 denoted by $Z_0 = Z(s_0)$ into one of two populations specified above with given training sample T. It follows from (1)–(3) that for given training sample realization $T = t(T^* = t^*)$, the conditional distribution of observation Z_0 in population Π_l is p-variate Gaussian

$$(Z_0 | T = t; \Pi_l) \sim N_p(\mu_{lt}^0, \Sigma^0), \tag{4}$$

where for $l = 1, 2$

$$\begin{aligned} \mu_{lt}^0 &= E\left(Z_0 | T = t; \Pi_l\right) = B_l' x_0 + \left(\alpha_0' \otimes I_p\right)\left(t - vec\left(XB\right)\right) = \\ &\quad B_l' x_0 + \left(\alpha_0' \otimes I_p\right)\left(t^* - XB\right)' \alpha_0, \end{aligned} \tag{5}$$

$$\Sigma^0 = \rho_0 \Lambda^{-1}. \tag{6}$$

In the following let P_{0l} denote the conditional distribution specified in (4)–(6), for $l = 1, 2$.

The squared Mahalanobis distance between the populations based on the conditional distributions defined in (4)–(6) for the observation taken at location $s = s_0$ is $\Delta_0^2 = \left(\mu_1^0 - \mu_2^0\right)' \Lambda \left(\mu_1^0 - \mu_2^0\right) / \rho_0$, where $\mu_l^0 = B_l' x_0, l = 1, 2$. This distance is usually used for determination the level of statistical separation between populations.

Then Bayes discriminant function (BDF) (Ducinskas, 2009), minimizing the probability of misclassification is formed by log-ratio of conditional likelihood of distribution specified in (4)–(6), that is

$$W(Z_0) = (1 + \alpha h_0)\left(Z_0 - \frac{1}{2}\left(\mu_{1t}^0 + \mu_{2t}^0\right)\right)' \Lambda \left(\mu_{1t}^0 - \mu_{2t}^0\right)/\sigma^2 + \gamma, \tag{7}$$

where $\gamma = \ln(\pi_1/\pi_2)$.

Let $V = (I_q, I_q)$ and $G = (I_q, -I_q)$, where I_q denotes the identity matrix of order q. Using (5), (6) and replacing t^* by T^* in (7) we get

$$W(Z_0) = (1 + \alpha h_0)\left(Z_0 - \left(T^* - XB\right)'\alpha_0 - B'V'x_0/2\right)' \Lambda B'G'x_0/\sigma^2 + \gamma. \tag{8}$$

This discriminant function is optimal under the minimum misclassification probability criterion. The probability of misclassification for $W(Z_0)$ will be called Bayes error rate or optimal error rate and will be denoted by P_0.

Let $\Phi(x)$ and $\varphi(\cdot)$ be the standard Gaussian distribution and density functions, respectively.

Lemma 1. *Bayes error rate for $W(Z_0)$ specified in (8) is $P_0 = \sum_{l=1}^{2} \pi_l \Phi(Q_l)$, where $Q_l = -\Delta_0/2 + (-1)^l \gamma/\Delta_0$.*

The proof of Lemma 1 is presented in the Appendix.

3. The error rates for PBDF

In practical applications not all statistical parameters of populations are known. Then the estimators of the unknown parameters can be found from the training sample. PBDF is formed by plugging estimators of parameters into BDF. In this paper we assume that the true values of parameters B and σ^2 are unknown and the estimators $\hat{B}$ and $\hat{\sigma}^2$ based on T are used. Then the set of parameters that are to be estimated and the set of their estimators are denoted by $\Psi = \{B, \sigma^2\}$ and $\{\hat{\Psi} = \hat{B}, \hat{\sigma}^2\}$, respectively. After replacing Ψ by $\hat{\Psi}$ in (8), we get the PBDF

$$\hat{W}(Z_0) = W\left(Z_0; \hat{\Psi}\right) = (1 + \alpha h_0)\left(Z_0 - \left(T^* - X\hat{B}\right)'\alpha_0 - \hat{B}'V'x_0/2\right)' \Lambda \hat{B}'G'x_0/\hat{\sigma}^2 + \gamma. \tag{9}$$

Then the actual error rate for BPDF $\hat{W}(Z_0)$ is defined as

$$P\left(\hat{\Psi}\right) = \sum_{l=1}^{2} \pi_l P_{0l}\left((-1)^l \hat{W}(Z_0) > 0\right). \tag{10}$$

Lemma 2. *The actual error rate for PBDF specified in (9) is*

$$P\left(\hat{\Psi}\right) = \sum_{l=1}^{2} \pi_l \Phi(\hat{Q}_l), \tag{11}$$

where

$$\hat{Q}_l = (-1)^l \frac{(1 + \alpha h_0)\left(x_0'\left(B_l - V\hat{B}/2\right) + \alpha_0' X\left(\Delta\hat{B}\right)\right) \Lambda \hat{B}'G'x_0 + \gamma\hat{\sigma}^2}{\hat{\sigma}\sqrt{x_0' G\hat{B}\Lambda\hat{B}'G'x_0(1 + \alpha h_0)}} \tag{12}$$

with $\Delta\hat{B} = \hat{B} - B$.

Proof. Proof of lemma is established by essentially exploiting the proof of Lemma 1, formulas (9), (10) and properties of multivariate Gaussian distribution.

Definition. The expectation of the actual error rate with respect to the distribution of T, designated as $E_T\{P(\hat{\Psi})\}$, is called the expected error rate (EER).

So the EER for considered problem of Z_0 classification by BPDF is specified by

$$E_T\left\{P\left(\hat{\Psi}\right)\right\} = E_T\left\{\sum_{l=1}^{2} \pi_l \Phi\left(\hat{Q}_l\right)\right\}. \tag{13}$$

In this paper we use the ML estimator of the regression coefficients $2q \times p$ matrix $\hat{B} = \left(X'R(\alpha)^{-1}X\right)^{-1} X'R(\alpha)^{-1}T^*$ and the bias adjusted ML estimator of feature covariance matrix $\hat{\sigma}^2 = \left(T - vec\left(\left(X\hat{B}\right)'\right)\right)' R(\alpha)^{-1} \otimes \Lambda^{-1}(T - vec((X\hat{B})'))/(np - 2q)$.

Using the properties of the matrix-variate normal distribution it is easy to show that

$$\hat{B} \sim N_{2q\times p}(B, \sigma^2 R_B \otimes \Lambda^{-1}) \tag{14}$$

and

$$\hat{\sigma}^2 \sim \sigma^2 \chi^2(np - 2q)/(np - 2q), \tag{15}$$

where $R_B = \left(X'R(\alpha)^{-1}X\right)^{-1}$.

Theorem. *Suppose that observation Z_0 to be classified by BPDF specified in (9). Then the approximation of EER is*

$$\begin{aligned}\mathrm{AER} = &\sum_{l=1}^{2} \pi_l \Phi(-\Delta_0/2 + (-1)^l \gamma/\Delta_0) + \pi_1 \varphi(-\Delta_0/2 - \gamma/\Delta_0)\times \\ &\{F_0' R_B F_0 \Delta_0/k_0 + (p-1)\, x_0' G R_B G' x_0/(k_0\Delta_0) + 2\gamma^2/\Delta_0(np - 2q)\}/2,\end{aligned}$$

where $F_0 = X'\alpha_0 - (V'/2 + \gamma G'/\Delta_0^2)x_0$ and $k_0 = 1/(1 + \alpha h_0)$.

The proof of the Theorem is presented in the Appendix.

4. Simulation study

In order to investigate the accuracy of the proposed classifier and to demonstrate the relevance of the asymptotic expansion of EER a numerical example is considered. The influence of statistical parameters of populations on the proposed actual error rate and the approximation of the expected error rate in the finite training sample case are examined. The influence of parameter θ that defines the dependence between the components of variable Z is analyzed as well.

The observations are assumed to arise from bivariate Gaussian random field on 2-dimensional square lattice with unit spacing and component precision matrix $\Lambda = \begin{pmatrix} 1 & -\theta \\ -\theta & 1 \end{pmatrix}$, $|\theta| < 1$. Mean vectors of populations are assumed constant, i.e. $\mu_1^0 = (0, 0)'$, $\mu_2^0 = (0, d)'$, $d > 0$ with $\sigma^2 = 1$.

Bivariate GMRFs are sampled on the 11×11 regular unit spacing lattice S_{120}^0 with the focal location in the center of the lattice. The endowed neighborhood system N^0 is based on the 2nd order of the nearest neighbors (2-NN). Lattice S_{120}^0 with the 2-NN neighborhood for the focal location is depicted in Fig. 1. Spatial weights w_{ij} typically reflect the spatial influence of observation from site i on observation from site j. Here we use power distance weights of the form $w_{ij} = d_{ij}^{-m}$, where d_{ij} refers to the Euclidean distance between sites i and j, and m is any positive integer.

Performance of the proposed PBDF is evaluated by empirical estimator of the expected error rate $\overline{ER} = \sum_{l=1}^{M} P(\hat{\Psi}_l)/M$. Here M is a number of training sample T realizations that are simulated according to the model specified in Eq. (1). $\hat{\Psi}_l$ denotes the appropriate ML estimates of parameter Ψ for each simulated realization of $T = t_l$, $l = 1, \ldots, M$. The simulations ($M = 100$) of bivariate GMRF were performed using the algorithm based on Cholesky factorization (Rue and Held, 2005).

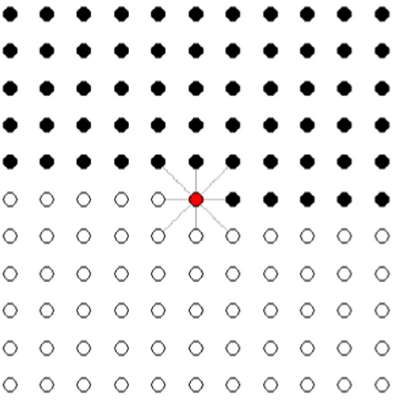

Fig. 1. Sample training locations. The points indicated by ● belong to Π_1, and the points indicated by ○ belong to Π_2. Sign ● denotes s_0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
$\overline{ER}$ and AER values for different values of α and θ.

α	$\overline{ER}$			AER			$\kappa = \left\|\text{AER} - \overline{ER}\right\| / \overline{ER}$		
	$\theta = 0.1$	$\theta = 0.5$	$\theta = 0.9$	$\theta = 0.1$	$\theta = 0.5$	$\theta = 0.9$	$\theta = 0.1$	$\theta = 0.5$	$\theta = 0.9$
0.1	0.19503	0.26466	0.40194	0.19524	0.26348	0.38914	0.00105	0.00449	0.03185
0.2	0.15395	0.22767	0.38524	0.15218	0.22642	0.36887	0.01146	0.00549	0.04249
0.3	0.12317	0.19834	0.37269	0.12100	0.19599	0.35162	0.01762	0.01183	0.05652
0.4	0.09926	0.17369	0.35448	0.09756	0.17036	0.33643	0.01716	0.01912	0.05090
0.5	0.08114	0.15182	0.34128	0.07936	0.14874	0.32277	0.02192	0.02032	0.05423
0.6	0.06624	0.13518	0.33166	0.06495	0.13052	0.31030	0.01944	0.03444	0.06441
0.7	0.05455	0.11906	0.32012	0.05340	0.11514	0.29880	0.02117	0.03292	0.06660
0.8	0.04530	0.10595	0.30298	0.04406	0.10203	0.28811	0.02721	0.03694	0.04909
0.9	0.03767	0.09558	0.29838	0.03647	0.09075	0.27812	0.03178	0.05061	0.06791
1	0.03117	0.08598	0.29108	0.03027	0.08093	0.26873	0.02894	0.05876	0.07679

The parameters to be varied in the simulation experiment are the spatial dependence parameter α and the component dependence parameter θ and the Euclidean distance between the mean vectors of populations, d. Accuracy of the AER is evaluated by relative error, $\kappa = \left|\text{AER} - \overline{ER}\right| / \overline{ER}$.

Table 1 shows the values of the AER, $\overline{ER}$ and κ calculated with respect to α and θ. The results show that all AER and $\overline{ER}$ values are decreasing while the values of α are increasing and the values of θ are decreasing. Fig. 2 confirms this conclusion.

The results in Table 2 are designated to analyze the influence of Euclidean distance between the mean vectors of populations, d with different spatial dependence parameters ($\alpha = 0.1, 0.5, 0.9$). According to the results in Table 2, it is easy to notice that all AER and $\overline{ER}$ values are decreasing while the values of d are increasing and the values of α are increasing. The accuracy of the AER evaluated by absolute error η is calculated for different α and d values and presented in Table 2 as well. Fig. 3 shows that absolute error of AER decreases monotonically as d increases.

Finally we examined the neighborhood size influence on the values of the AER for the fixed values $d = 1, \theta = 0.5$ *and* $\alpha = 0.5$. Fig. 4 shows that the AER monotonically decreases when the neighborhood size determined by NN order k increases.

5. Discussion

In this paper we discuss the problem of statistical classification of multivariate conditionally autoregressive Gaussian random field observations. We provide formulas for actual error rate and

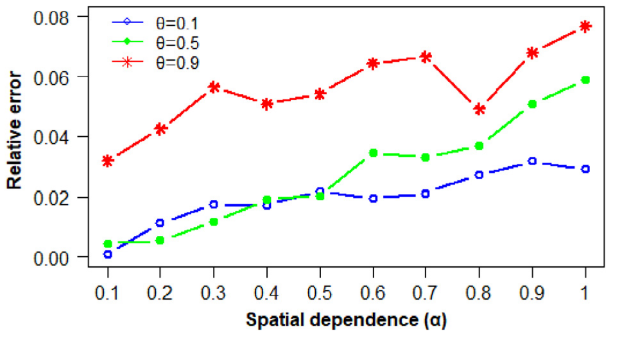

Fig. 2. Relative error of the AER for various values of α and three values of θ.

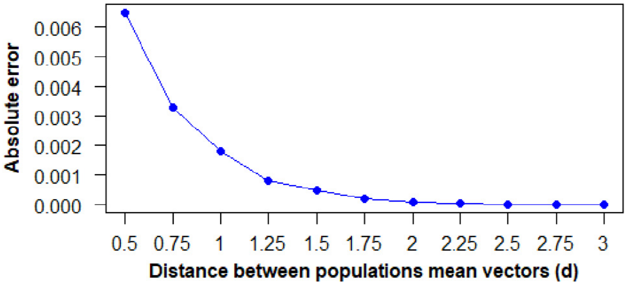

Fig. 3. The absolute difference between $\overline{ER}$ and AER for $\theta = 0.5$, $\alpha = 0.9$.

Table 2
$\overline{ER}$ and AER values for different values of α and d.

d	$\overline{ER}$			AER			$\eta = \lvert\overline{ER} - \text{AER}\rvert$		
	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
0.5	0.383661	0.312682	0.261479	0.376464	0.303430	0.254985	0.007197	0.009252	0.006493
0.75	0.320317	0.222843	0.162252	0.318272	0.218395	0.158963	0.002046	0.004448	0.003290
1	0.261834	0.149275	0.092507	0.263476	0.147503	0.090698	0.001642	0.001772	0.001808
1.25	0.211906	0.096580	0.048204	0.212638	0.094598	0.047390	0.000732	0.001982	0.000813
1.5	0.167300	0.058599	0.022973	0.167139	0.057542	0.022490	0.000161	0.001058	0.000483
1.75	0.130103	0.033561	0.009890	0.128636	0.033003	0.009668	0.001467	0.000557	0.000222
2	0.098932	0.018278	0.003844	0.097325	0.017820	0.003757	0.001608	0.000457	0.000086
2.25	0.073772	0.009234	0.001357	0.072229	0.009048	0.001318	0.001543	0.000185	0.000039
2.5	0.053568	0.004434	0.000427	0.052449	0.004317	0.000417	0.001119	0.000118	0.000010
2.75	0.038067	0.001989	0.000122	0.037235	0.001933	0.000119	0.000831	0.000056	0.000003
3	0.026518	0.000839	0.000031	0.025835	0.000812	0.000030	0.000682	0.000026	0.000001

the approximation of the expected error rate. Through the simulation study with a moderate size of training samples we found out that:

- The accuracy of the AER is monotonically decreasing with respect to the increase in α for fixed component dependence parameters θ.
- The accuracy of the AER is monotonically decreasing with respect to the increase in θ for fixed spatial dependence parameters α.

That means the greater spatial dependence and the smaller component dependence give better accuracy of the proposed AER.

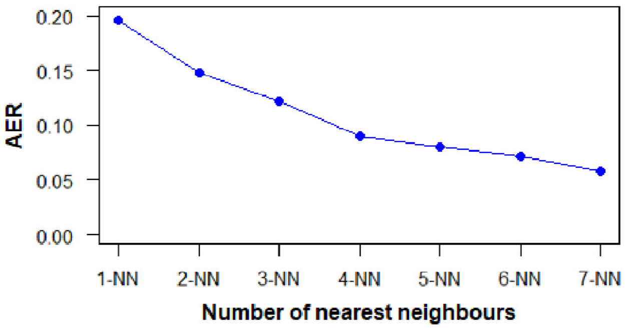

Fig. 4. The influence of the order of the neighborhood on the AER. k-NN indicates kth order of the nearest neighbors. $d = 1, \theta = 0.5, \alpha = 0.5$.

- The accuracy of the AER is better when the distance between populations mean vectors increases. That means that for slightly overlapping populations the performance of the proposed classifier is higher.
- For a particular value of spatial dependence parameter the AER values monotonically decrease with a specified neighborhood size (the order of the NN structure). That means the more neighbors are included the better performance of the proposed classifier is observed.

Therefore, the results of the simulation study give us strong arguments to expect that the proposed AER is of a good accuracy, especially for small values of spatial and component dependence parameters, and it could be effectively used for performance evaluation of classifiers based on the PBDF applied to the multivariate GMRF observations. This suggests that the proposed AER can be applied as an objective function in designing spatial training samples.

It should also be noted that the developed approach for spatial classification could be easily generalized to more than two populations (see Ducinskas et al., 2015). Extension to the conditionally specified auto-binomial and auto-gamma models (see Moller, 1999) are left for future work. Practical applications of proposed classifier will be concentrated to the mapping of presence and absence of invasive species (zebra mussel *Dreissena polymorpha*) in the Curonian lagoon, a shallow water body in the South-East Baltic. The aim of these investigations is to increase the accuracy of estimation of the invasive alien species impact on biotic indices and environmental quality assessments (see Zaiko and Daunys, 2015).

Appendix

Proof of Lemma 1. Recall, that the Bayes error rate for $W(Z_0)$ is defined as $P_0 = \sum_{l=1}^{2} \pi_l P_{0l}((-1)^l W(Z_0) > 0)$. It is obvious that in population Π_l the conditional distribution of $W(Z_0)$, given $T = t$ is Gaussian, i.e., $W(Z_0|T = t, \Pi_l) \sim N((-1)^{l+1}\Delta_0^2/2 + \gamma, \Delta_0^2)$, $l = 1, 2$. Then using the properties of normal distribution we complete the proof of Lemma 1.

Proof of Theorem. Let $\Delta\hat{B} = \hat{B} - B$, $\Delta\hat{\sigma}^2 = \hat{\sigma}^2 - \sigma^2$. Using (14)–(15) it is easy to show (e.g. Magnus and Neudecker, 2002) that

$$E_T\left(\Delta\hat{B}\right) = 0, \quad E_T\left(\Delta\hat{\sigma}^2\right) = 0, \quad E_T\left(\Delta\hat{\sigma}^2\Delta\hat{B}\right) = 0 \tag{16}$$

and

$$E_T\left(vec\left(\Delta\hat{B}'\right)\left(vec\left(\Delta\hat{B}'\right)\right)'\right) = R_B \otimes \Sigma, \tag{17}$$

$$E_T\left(\Delta\hat{\sigma}^2\right) = \sigma^4/(np - 2q). \tag{18}$$

For the proof of Theorem we use the following notations. Let $\hat{b}_{\alpha\beta}$ designate the elements of $\hat{B}$. Denote by $P_B^{(1)}(\alpha,\beta) = \partial P(\hat{\Psi})/\partial \hat{b}_{\alpha\beta}$, $P_B^{(2)}(\alpha\beta,\gamma\delta) = \partial^2 P(\hat{\Psi})/\partial \hat{b}_{\alpha\beta}\partial \hat{b}_{\gamma\delta}$, $P_\sigma^{(1)} = \partial P(\hat{\Psi})/\partial \hat{\sigma}_{ij}$, $P_\sigma^{(2)} = \partial_\sigma^2 P(\hat{\Psi})/\partial^2 \hat{\sigma}^2$ and $P_{B,\sigma}^{(2)}(\alpha\beta) = \partial^2 P(\hat{\Psi})/\partial \hat{b}_{\alpha\beta}\partial \hat{\sigma}^2$ the partial derivatives of $P(\hat{\Psi})$ with respect to the corresponding parameters evaluated at $\hat{B} = B$, $\hat{\sigma}^2 = \sigma^2$. Analogous notations will be used for partial derivatives of $\hat{Q}_l$, $l = 1, 2$.

Make a Taylor expansion of $P(\hat{\Psi})$ at points $\hat{B} = B$ and $\hat{\sigma}^2 = \sigma^2$ up to the second order partial derivatives and use the Lagrange remainder term. Taking the expectation with respect to the distribution of T and using (16) we get

$$E_T\left\{P\left(\hat{\Psi}\right)\right\} = P_0 + \left(\sum_{\alpha,\gamma=1}^{2q}\sum_{\beta,\delta=1}^{p} P_B^{(2)}(\alpha\beta,\gamma\delta)\, E_T\left\{\Delta\hat{b}_{\alpha\beta}\Delta\hat{b}_{\gamma\delta}\right\} + P_\sigma^{(2)} E_T\left\{\Delta\hat{\sigma}^2\right\}\right)/2 + R_3, \quad (19)$$

where R_3 is the expectation of remainder term. Note that

$$\pi_1\varphi(Q_1) = \pi_2\varphi(Q_2). \quad (20)$$

By using the chain rule and Eq. (20) we have

$$P_B^{(2)}(\alpha\beta,\gamma\delta) = \pi_1\varphi(Q_1)\sum_{l=1}^{2}(-1)^l\left(Q_l Q_{lB}^{(1)}(\alpha\beta)\, Q_{lB}^{(1)}(\gamma\delta) - Q_{lB}^{(2)}(\alpha\beta,\gamma\delta)\right), \quad (21)$$

$$P_\sigma^{(2)} = \pi_1\varphi(Q_1)\sum_{l=1}^{2}(-1)^l\left(Q_l\left(Q_{l\sigma}^{(1)}\right)^2 - Q_{l\sigma}^{(2)}\right). \quad (22)$$

Taking the appropriate partial derivatives by elements of matrices, we have

$$Q_{lB}^{(1)}(\alpha\beta) = (-1)^l\left(X'\alpha_0 - \left(V'/2 + \gamma G'/\Delta_0^2\right)x_0\right)J^{\alpha\beta}\Lambda B' G x_0/(\Delta_0 k_0 \sigma^2), \quad (23)$$

$$\sum_{l=1}^{2} Q_{lB}^{(2)}(\alpha\beta,\gamma\delta) = \left(\sigma^2\right)^{-1}\left(x_0' G J^{\alpha\gamma} G' x_0 - x_0' G J^{\alpha\beta}\Lambda B' G' x_0 x_0' G J^{\gamma\delta}\Sigma^{-1} B' G' x_0/\Delta_0^2 k\right)/(\Delta_0 k_0), \quad (24)$$

$$Q_{l\sigma}^{(1)} = (-1)^l \gamma x_0' G B \Sigma^{-1}\Lambda B' G x_0'/(\Delta_0^3 k_0 \sigma^2), \quad (25)$$

$$\sum_{l=1}^{2} Q_{l\sigma}^{(2)} = \left(x_0' G B \Sigma^{-1}\Lambda B' G' x_0 - x_0' G B \Lambda B' G' x_0 x_0' G \Lambda B' G' x_0/\Delta_0^2 k_0 \sigma^4\right)/(\Delta_0 k_0), \quad (26)$$

where J^{ij} is matrix of zeros except element (i, j) that is equal 1.

Remember, that the Lagrange remainder is the third-order polynomial with respect to the components of $\Delta\hat{B}$ and $\Delta\hat{\sigma}^2$. Coefficients of this polynomial are the third-order partial derivatives of $P(\hat{\Psi})$ with respect to components of $\hat{B}$ and $\hat{\sigma}^2$ estimated in the neighborhood of their true values. It is obvious from (14) and (15), that all third order moments of normally distributed components of $\Delta\hat{B}$ are equal to 0 and all third order moments of $\Delta\hat{\sigma}^2$ components are of order $O(1/n^2)$. Third order partial derivatives of $\Phi(\hat{Q}_l)$ with respect to elements of $\hat{B}$ and $\hat{\sigma}^2$ are bounded by the uniformly bounded functions in the neighborhood of point $\hat{B} = B$, $\hat{\sigma}^2 = \sigma^2$. So we can scrap R_3 in (19).

Finally, putting (20)–(26) into (19) and using (17)-(18) we complete the proof of the theorem.

References

Berret, C., Calder, C.A., 2016. Bayesian spatial binary classification. Spatial Stat. 16, 72–102.

De Oliveira, V., Ferreira, M.A.R., 2011. Maximum likelihood and restricted maximum likelihood estimation for a class of Gaussian Markov random fields. Metrika 74 (2), 167–183.

Ducinskas, K., 2009. Approximation of the expected error rate in classification of the Gaussian random field observations. Statist. Probab. Lett. 79, 138–144.
Ducinskas, K., 2011. Error rates in classification of multivariate Gaussian random field observation. Lith. Math. J. 51 (4), 477–485.
Ducinskas, K., Dreiziene, L., 2016. Expected error rates in classification of Gaussian CAR observations. In: Computer Data Analysis and Modeling: Theoretical and Applied Stochastics: Proc. of the 11th Intern. Conf., vol. 1, pp. 127–130.
Ducinskas, K., Dreiziene, L., 2018. Risks of classification of the gaussian Markov random field observations. J. Classification 35, 422–436.
Ducinskas, K., Dreiziene, L., Zikariene, E., 2015. Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. Statist. Probab. Lett. 98, 107–114.
Jin, X., Carlin, B.P., Banerjee, S., 2005. Generalized hierarchical multivariate CAR models for areal data. Biometrics 61, 950–961.
Klein, R., Press, S.J., 1992. Adaptive Bayesian classification of spatial data. J. Amer. Statist. Assoc. 87, 844–851.
Magnus, J.R., Neudecker, H., 2002. Matrix Differential Calculus and Applications in Statistics and Econometrics. Wiley, New York.
Mardia, K.V., 1984. Spatial discrimination and classification maps. Comm. Statist. Theory Methods 13, 2181–2197.
Mardia, K.V., 1988. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. J. Multivariate Anal. 24, 265–284.
McLachlan, G.J., 2004. Discriminant Analysis and Statistical Pattern Recognition. Wiley, New York.
Moller, J., 1999. Perfect simulation of conditionally specified models. J. R. Stat. Soc. Ser. B Stat. Methodol. 61, 251–264.
Pettitt, A.N., Weir, I.S., Hart, A.G., 2002. A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. Stat. Comput. 12, 353–367.
Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall, Boca Raton.
Sain, S.R., Cressie, N., 2007. A spatial model for multivariate lattice data. J. Econom. 140, 226–259.
Saltyte-Benth, J., Ducinskas, K., 2005. Linear discriminant analysis of multivariate spatial–temporal regressions. Scand. J. Stat. 32, 281–294.
Switzer, P., 1980. Extension of discriminant analysis for statistical classification of remotely sensed satellite imagery. J. Int. Assoc. Math. Geol. 12, 86–95.
Zaiko, A., Daunys, D., 2015. Invasive ecosystem engineers and biotic indices: Giving a wrong impression of water quality improvement? Ecol. Indic. 52, 292–299.