Performance Evaluations of Gaussian Spatial Data Classifiers Based on Hybrid Actual Error Rate Estimators

Kestutis Ducinskas
Klaipeda University
Vilnius University

Lina Dreiziene
Lithuanian Maritime Academy
LCC International University

Abstract

Discrimination and classification of spatial data has been widely mentioned in the scientific literature, but lacks full mathematical treatment and easily available algorithms and software. This paper fills this gap by introducing the method of statistical classification based on Bayes discriminant function (BDF) and by providing original approach for the classifier performance evaluation. Supervised classification of spatial data with response variable modelled by Gaussian random fields (GRF) with continuous or discrete spatial index is studied. Populations are assumed to be with different regression parameters vectors. Classification rule based on BDF with inserted ML estimators of regression and scale parameters is studied. We focus on the derived actual error rate (AER) and the approximation of the expected error rate (AEER) for both types of models. These are used in the construction of hybrid actual error rate estimators that are spatial modifications of widely applicable D and O estimators applied in cases of independent observations.

Simulation experiments are used for comparison of proposed AER estimators by the minimum of unconditional mean squared error criterion for both types of GRF models.

Keywords: spatial index, Bayes discriminant function, probability of misclassification, actual error rate.

1. Introduction

Statistical classification and discriminant analysis of spatial data has been mentioned in the scientific literature, but lacks full mathematical treatment and easily available algorithms and software. This paper fills this gap by proposing the method of Gaussian spatial models evaluation and comparison based on classification error rate estimators and by providing novel formulas and algorithms, which allows to evaluate the influence of spatial information to the performance of proposed classifier. The actual and expected error rates for supervised classification of Gaussian random field (GRF) observations via plug-in Bayes discriminant function (PBDF) in partial parametric uncertainty case are studied by Ducinskas (2009) and in complete parametric uncertainty case by Ducinskas and Dreiziene (2011). Given training sample, the explored classifier is obtained by substituting model parameters with their estimators in the well-known Bayes rule. Spatial discrimination based on PBDF for feature observations

having elliptically contoured distributions is implemented by Batsidis and Zografos (2011). Numerical comparison of the performances for different spatial classification rules is performed by Berrett and Calder (2016). In the above mentioned papers, the main attention is paid to the so-called geostatistical models (GS) with continuous spatial index and directly specified covariance functions belonging to Matern type or other parametric models. However, many researchers intesivelly explored spatial Gaussian data observed over the lattice. Two classes of spatial linear models for lattice data are used in practice: conditional autoregressive model (CAR) and simultaneous regression model (SAR) following a neighbourhood structure on lattice. In spatial statistics literature CAR models are the most often used for the analysis of lattice/areal data since majority of authors declare that CAR models being subclass of Markov random fields are more general than SAR models (see, e.g., Sain and Cressie 2007). Recall that CAR models are a subclass of Markov random fields such that the spatial dependence is induced by conditional distributions of random errors at individual locations (sites). The practical use of Gaussian Markov random field (GMRF) (see, e.g., Rue and Held 2005) for modelling large scale spatial phenomena has significantly increased after recent advances on the efficient simulation. Without insignificant loss of generality, we restrict our attention to homogenious CAR (HCAR) lattice models (Song and De Oliveira 2012) with original parametric structure proposed by De Oliveira and Ferreira (2011). These are well-suited to the case of small samples, and ensures good frequentist properties of ML estimators of drift and scale parameters. Spatial classification based on PBDF for univariate HCAR models, imposed by the mentioned above structure, is recently explored by Ducinskas and Dreiziene (2018).

In the present paper we focus on linear classification problem of GRF observation for GS models as well as for HCAR models by using PBDF. The main theoretical objective of this study is to study the properties of two types actual error rate estimators based on previously derived analytic expressions for AER and AEER. The proposed hybrid estimators, say SD and SO, are formed by replacing theoretical parameters with their estimators in AER and AEER, respectively. This is a spatial modification of D and O estimators applied for classification problems of independent Gaussian observations (see, e.g., Lachenbruch and Mickey 1968; Snapinn and Knoke 1984; Egbo 2016). Comparison of the geostatical models for ecological data based on correct classification rates is performed by Ducinskas and Dreiziene (2019) and Dreiziene and Ducinskas (2020).

This paper is organized as follows: the problem description, definitions of classifiers based on BDF are displayed in the next section; specification of the actual error rates and their estimators are presented in Section 3. Section 4 illustrates the proposed methods by simulation experiments and, finally, the conclusions are in the last section.

2. Main concepts and definitions

In this paper we focus on classification of spatial data that can be considered as realization of random field $\{Z(s) : s \in S \subset R^2\}$. The universal kriging model belonging to the subclass of spatial linear models is explored. The model of observation $Z(s)$ for $s \in S$ is (Haining 1990; Cressie and Wikle 2011)

$$Z(s) = x'(s)\beta_l + \varepsilon(s),$$

where $x(s)$ is a $q \times 1$ vector of non-random regressors and β_l is a $q \times 1$ vector of unknown parameters, $l = 1, 2$, $\beta_1 \neq \beta_2$. The error term $\varepsilon(s)$ is zero-mean GRF $\{\varepsilon(s) : s \in S\}$ with known covariance function $\sigma(s, t) = cov(\varepsilon(s), \varepsilon(t))$, $s, t \in S$.

In this paper we have considered geostatistical model (GS) of random fields for which spatial index s is assumed to vary continuously throughout S.

Suppose that $\{s_i \in S, i = 0, 1, ..., n\}$ is the set of spatial sites where the observations of GRF are taken. Denote the set of training sites by $S_n = S^{(1)} \cup S^{(2)}$, where $S^{(1)} = \{s_1, s_2, ..., s_{n_1}\}$ and $S^{(2)} = \{s_{n_1+1}, ..., s_{n_1+n_2}\}$, $n = n_1 + n_2$. Suppose that $S^{(l)}$ are the subsets of S_n that

contains n_l observations of $Z(s)$ from Ω_l, $l = 1, 2$. The site of the observation to be classified is denoted by s_0 and will be called a focal location (see, e.g., Berrett and Calder 2016). Set $S_n^0 = S_n \cup \{s_0\}$.

Put $\beta' = (\beta_1', \beta_2')$, $\alpha_0 = \Sigma^{-1}c_0$, and denote by X_l the $n_l \times q$ matrix of regressors for observations from Ω_l, $l = 1, 2$. Then $n \times 2q$ design matrix of the training sample Z is specified by $X = X_1 \bigoplus X_2$, where symbol $\bigoplus$ denotes the direct sum of matrices and X_l is the $n_l \times q$ matrix of regressors for observations from Ω_l, $l = 1, 2$.

In what follows we use the following notations for $i, j = 0, ..., n$: $Z(s_i) = Z_i$, $\varepsilon(s_i) = \varepsilon_i$, $x(s_i) = x_i$, $\sigma_{ij} = cov(Z_i, Z_j)$. Setting $Z = (Z_1, ..., Z_n)'$ we have $c_0 = cov(Z_0, Z) = (\sigma_{01}, \sigma_{02}, ..., \sigma_{0n})'$ and $\Sigma = Var(Z) = (\sigma_{ij}, i, j = 1, ..., n)$. So the training sample Z has multivariate Gaussian distribution

$$Z \sim N_n(X\beta, \Sigma).$$

The problem considered in this paper is the following: for given training sample Z classify Z_0 into one of two described below populations. Let z denote the realization of Z. Then the conditional distribution of Z_0 given $Z = z$ in population Ω_l is Gaussian

$$Z_0 | Z = z, \Omega_l \sim N_1(\mu_{zl}^2, \sigma_{0z}^2), \tag{1}$$

$$\mu_{zl}^2 = x'\beta_l + \alpha_0'(z - X\beta), \tag{2}$$

$$\sigma_{0z}^2 = \sigma_{00} - c_0'\Sigma^{-1}c_0. \tag{3}$$

First consider GS model for spatial data. For this model we consider stationary random error case. Assume that covariance function is directly specified parametric function $\sigma(s, t) = \sigma^2 r(s - t)$, where $r(\bullet)$ is the spatial correlation function and $\sigma^2 = \sigma_{ii}$, $i = 0, ..., n$.

Let $r_0 = (r(s_0 - s_1), ..., r(s_0 - s_n))$ and $R = (r(s_i - s_j), i, j = 1, ..., n)$. Then $\alpha_0 = R^{-1}r_0$ and $\sigma_{0z}^2 = \sigma^2 \rho_0$, where $\rho_0 = (1 - r_0'R^{-1}r_0)$.

Next consider the lattice data − HCAR model. Assume that lattice $S_n^0 = S_n \cup \{s_0\}$ is endowed with a neighbourhood system $N^0 = \{N_i^0 : i = 0, 1, ..., n\}$, and lattice S_n is endowed with a neighbourhood system $N = \{N_i : i = 1, ..., n\}$, where N_i denotes the collection of sites that are neighbours of site s_i.

Define spatial weight $w_{ij} > 0$ as a measure of similarity between sites s_i and s_j, and put $w_{ij} = w_{ji}$, $h_i^0 = \sum_{j \in N_i^0} w_{ij}$, $i, j = 0, 1, ..., n$ and $h_i = \sum_{j \in N_i} w_{ij}$, $i, j = 1, ..., n$.

Suppose

$$h_{ij} = \begin{cases} h_i & if \ \ i = j \\ -w_{ij} & if \ \ i \in N_j \\ 0 & otherwise \end{cases}$$

and

$$h_{ij}^0 = \begin{cases} h_i^0 & if \ \ i = j \\ -w_{ij} & if \ \ i \in N_j^0. \\ 0 & otherwise \end{cases}$$

Then consider the $(n + 1) \times (n + 1)$ matrix $H^0 = (h_{ij}^0 : i, j = 0, 1, ..., n)$ and $n \times n$ matrix $H = (h_{ij} : i, j = 1, ..., n)$. Let $w_0' = (w_{01}, ..., w_{0n})$, $w_i' = (w_{i1}, ..., w_{ii-1}, ..., w_{in})$, $i = 1, ..., n$. So for HCAR model the covariance matrix is $\Sigma = \sigma^2 R$ with $R = (I_n + \alpha H)^{-1}$.

Define $Z^+ = (Z_0, Z')'$ and assume $var(Z^+) = \sigma^2(I_{n+1} + \alpha H^0)^{-1}$ (see, De Oliveira and Ferreira 2011). Here $\alpha \geq 0$ is a spatial dependence parameter and $\sigma > 0$ is a scale parameter. Then $\alpha_0' = \alpha w_0'/(1 + \alpha h_0)$, $\rho_0 = 1/(1 + \alpha h_0)$.

In the following for brevity we will use $\Psi = (\beta', \sigma^2)'$. Set the squared marginal Mahalanobis distance by $\Delta_0^2 = (x_0'(\beta_1 - \beta_2))^2/\sigma^2$. Then the squared conditional Mahalanobis distance between conditional distributions of Z_0 given $Z = z$ is specified by $\Delta_{0z}^2 = \Delta_0^2/\rho_0$. Recall that in the GS case $\rho_0 = (1 - r_0'R^{-1}r_0)$ and in the HCAR case $\rho_0 = 1/(1 + \alpha h_0)$.

Under the assumption of complete parametric certainty, the Bayes discriminant function (BDF) minimizing the probability of misclassification (PMC) is formed by log ratio of conditional likelihoods.

Denote by $\pi_1^0, \pi_2^0, (\pi_1^0 + \pi_2^0 = 1)$ the prior probabilities of the populations Ω_1 and Ω_2, respectively, for observation at focal location s_0, given training sample Z.

Then for $Z = z$, BDF is specified by (see, McLachlan 2004)

$$W_z(Z_0, \Psi) = \left(Z_0 - \frac{1}{2}(\mu_{1z}^0 + \mu_{2z}^0)\right)(\mu_{1z}^0 - \mu_{2z}^0)/\sigma_{0z}^2 + \gamma_0,$$

where $\gamma_0 = ln(\pi_1^0/\pi_2^0)$. So BDF allocates the observation in the following way: classifies observation Z_0 given Z to the population Ω_1 if $W_z(Z_0, \Psi) \geq 0$, and to the population Ω_2, otherwise.

Recall that by definition squared Mahalanobis distance between populations specified above at location s_0 is $\Delta_{0z}^2 = (\mu_{1z}^0 - \mu_{2z}^0)^2/\sigma_{0z}^2 = (x_0'(\beta_1 - \beta_2))^2/\sigma_{0z}^2$. The explicit expression of the PMC associated with BDF for GRF is derived in Ducinskas (2009). Hence for above specified classifier PMC has the following form:

$$P(\Psi) = \sum_{l=1}^{2} \left(\pi_l^0 \Phi(-\Delta_{0z}/2 + (-1)^l \gamma_0/\Delta_{0z})\right),$$

where $\Phi(\bullet)$ is the standard Gaussian distribution function.

3. Actual error rates and their estimators

Let $\hat{\mu}_{1z}$, $\hat{\mu}_{2z}$ and $\hat{\sigma}_{0z}^2$ be the estimators of μ_{1z}, μ_{2z} and σ_{0z}^2, respectively, obtained by replacing β and σ^2 in equations (2) and (3) with their estimators $\hat{\beta}$ and $\hat{\sigma}^2$ based on Z. Put $\hat{\Psi} = (\hat{\beta}', \hat{\sigma}^2)$. Then using (1)-(3) we get the estimator of conditional mean and conditional variance

$$\hat{\mu}_{lz}^0 = x_0'\hat{\beta}_l + \hat{\alpha}_0'(z - X\hat{\beta}), \quad l = 1, 2,$$

$$\hat{\sigma}_{0z}^2 = \sigma^2 \rho_0.$$

In the present paper we use ML estimator of regression parameters:

$$\hat{\beta} = (X'R^{-1}X)^{-1}X'R^{-1}Z \sim N_{2q}(\beta, \Sigma_\beta),$$

and bias adjusted ML estimator of scale parameter

$$\hat{\sigma}^2 = (Z - X\hat{\beta})'R^{-1}(Z - X\hat{\beta})/(n - 2q),$$

where $\Sigma_\beta = \sigma^2 R_\beta$ with $R_\beta = (X'R^{-1}X)^{-1}$. By replacing the parameters with their estimators in (5) we form PBDF

$$W_z(Z_0, \hat{\Psi}) = \left(Z_0 - \hat{\alpha}_0'(z - X\hat{\beta}) - \frac{1}{2}x_0'F^+\hat{\beta}\right)(x_0'F^-\hat{\beta})/\hat{\sigma}_{0z}^2 + \gamma_0$$

with $F^+ = (I_q, I_q)$ and $F^- = (I_q, -I_q)$, where I_q denotes the identity matrix of order q.
Let for $l = 1, 2$, $a_l = x_0'\beta_l - \alpha_0'X\beta$, $b = \alpha_0'X - x_0'F^+/2$.

Lemma 1. *The actual error rate (AER) for PBDF $W_{\tilde{z}}(Z_0; \hat{\Psi})$ is*

$$P(\hat{\Psi}) = \sum_{l=1}^{2} \left(\pi_l^0 \Phi(\hat{Q}_l) \right),$$

where $\hat{Q}_l = (-1)^l \left((a_l + b\hat{\beta}) sgn(x_0' F^- \hat{\beta}) / \sigma_{0z} + \gamma_0 \hat{\sigma}_{0z}^2 / (\sigma_{0z} |x_0' F^- \hat{\beta}|) \right).$

Proof of Lemma 1 is presented in Ducinskas (2009).

Note that $Q_l(\Psi) = -\Delta_{0z}/2 + (-1)^l \gamma_0 / \Delta_{0z}$.

The expectation of the actual error rate with respect to the distribution of Z is called the expected error rate (EER). The EER is useful in providing a guide to the performance of the plug-in classification rule before it is actually formed from the training sample. It can be considered as the performance measure to the PBDF similar as the mean squared prediction error (MSPE) is the performance measure to the plug-in kriging predictor (see, Diggle, Ribeiro, and Christensen 2003). These facts strengthen the motivation for deriving the AEER associated with PBDF.

AER is a function of discriminant function, but the distribution of PBDF based on unknown parameters is quite complicated and thus an analytical expressions for the error rates becomes difficult. Therefore AER should be estimated by different error rate estimators, which will be explained below. One type of AER estimators is based on approximation of AER (see, Ducinskas 2009).

In the present paper we use AEER that will be derived in Lemma 2.

Lemma 2. *The approximation of EER for PBDF $W_z(Z_0, \hat{\Psi})$ is*

$$AEER = P(\Psi) + \pi_1^0 \varphi(-\Delta_{0z}/2 - \gamma_0/\Delta_{0z}) \left(\Delta_{0z} K_\beta / 2 + \gamma_0^2 / (n - 2q)\Delta_{0z} \right),$$

where $K_\beta = \Lambda' R_\beta \Lambda$ and $\Lambda' = \alpha_0' X - x_0'(F^+/2 + \gamma_0 F^-/\Delta_{0z}^2)$.

Proof. The proof of Lemma 2 is based on the second order Taylor expansion of $P(\hat{\Psi})$ about the point $\hat{\Psi} = \Psi$, and is implemented by mimicking the proof of the Theorem from Ducinskas paper (2009). $\square$

Here we propose spatial modification of the widely used error rate estimators D and O and denote them by SD and SO. Note, that each of the methods of error rate estimation described in this section is given a symbol to identify it. The estimators are referred to by symbol as a superscript.

Definition 1. The actual error rate estimator, SD, is obtained by replacing Δ_{0z}^2 with the estimator

$$D_{0z}^2 = (\hat{\mu}_{1z} - \hat{\mu}_{2z})^2 / \hat{\sigma}_{0z}^2$$

in PMC. Therefore

$$\hat{P}^{SD} = \sum_{l=1}^{2} \left(\pi_l^0 \Phi(-D_{0z}/2 + (-1)^l \gamma_0 F^-/D_{0z}) \right). \tag{4}$$

Definition 2. The actual error rate estimator, SO, is obtained by replacing Δ_0^2 with $\hat{\Delta}_0^2$ in AER, i.e.

$$\hat{P}^{SO} = \hat{P}^{SD} + \pi_1^0 \varphi(D_{0z}/2 + \gamma_0/D_{0z})(D_{0z}\tilde{K}_\beta/2 + \gamma_0^2/D_{0z}(n - 2q)), \tag{5}$$

where $\tilde{K}_\beta = \tilde{\Lambda}' R_\beta \tilde{\Lambda}$ with $\tilde{\Lambda}' = \alpha_0' X - x_0'(F^+/2 + \gamma_0 F^-/D_{0z}^2)$.

Let $\hat{P}$ represent an arbitrary estimator of actual error rate. As it follows, we call them error rate estimators.

For effectiveness, criteria of these estimators is based on the unconditional mean squared error (UMSE) i.e.

$$UMSE(\hat{P}) = E(\hat{P} - P(\hat{\Psi}))^2. \tag{6}$$

4. Simulation experiments

In order to illustrate the results of previous section, a numerical example is considered. Comparison of the proposed AER estimators with respect to the minimum of UMSE for GS and HCAR models is demonstrated. We consider the empirical estimators of the error rates incurred by the rule based on the proposed PBDF for stationary GRF.

Assume that data are sampled on the 11×11 regular unit spacing lattice with the focal location in the centre of the lattice (see Figure 1). Equal-sized training samples with equal prior probabilities are considered, that is $n_1 = n_2 = 60$ and $\pi_1^0 = \pi_2^0 = 1/2$.
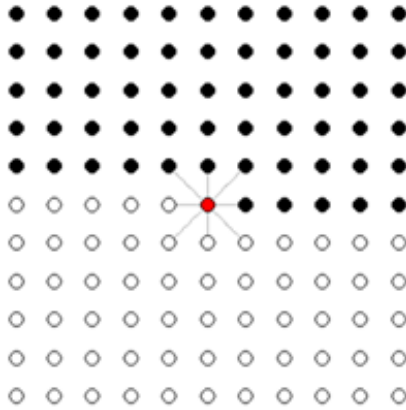

Figure 1: Sample training sites. The points indicated by $\bullet$ belong to $S^{(1)}$, and the points indicated by $\circ$ belong to $S^{(2)}$. The central point $\bullet$ denotes s_0 (focal location).

Spatial correlation for GS case is modelled by isotropic exponential covariance function given by $\sigma(h) = \sigma^2 exp(-h/\alpha)$. 100 replications ($M = 100$) were performed with true parameters $\sigma^2 = 1$, $\alpha = 2$ using geoR, package included into the R project for statistical computing.

Table 1: Actual error rate estimators for different Mahalanobis distance Δ_0 in GS case

Δ_0	$\hat{P}^{SD}$	$\hat{P}^{SO}$	$UMSE^{SD}$	$UMSE^{SO}$	$\Delta(UMSE)$
0.25	0.39846	0.398462	0.011071	0.011071	0.00000047
0.5	0.357336	0.357339	0.008786	0.008785	0.00000040
0.75	0.269909	0.269915	0.008140	0.008139	0.00000073
1	0.218281	0.218289	0.006300	0.006299	0.00000057
1.25	0.167597	0.167607	0.004703	0.004703	0.00000060
1.5	0.113626	0.113639	0.005681	0.005679	0.00000111
1.75	0.076875	0.076890	0.004482	0.004481	0.00000110
2	0.061583	0.061600	0.001693	0.001693	0.00000050

For HCAR case the simulations were performed according to the algorithm based on Cholesky factorisation proposed by Rue and Held (2005). Spatial weights w_{ij} typically reflect the spatial influence of observation from site s_i on observation from site s_j. Here we use power distance

weights of the form $w_{ij} = d_{ij}^{-2}$, where d_{ij} refers to the Euclidean distance between sites s_i and s_j.

Table 2: Actual error rate estimators for different Mahalanobis distance Δ_0 in HCAR case

Δ_0	$\hat{P}^{SD}$	$\hat{P}^{SO}$	$UMSE^{SD}$	$UMSE^{SO}$	$\Delta(UMSE)$
0.25	0.363667	0.363682	0.005282	0.005284	0.00000180
0.5	0.247142	0.247176	0.014470	0.014476	0.00000607
0.75	0.145962	0.146021	0.040068	0.040087	0.00001874
1	0.083951	0.084034	0.049309	0.049337	0.00002805
1.25	0.040081	0.040193	0.052304	0.052341	0.00003684
1.5	0.019177	0.019312	0.084212	0.084269	0.00005714
1.75	0.007670	0.007830	0.077168	0.077229	0.00006118
2	0.003044	0.003227	0.086850	0.086920	0.00007023

The values of the proposed actual error rate estimators (4), (5) and their UMSE (6) for both GRF types are presented in Table 1 and Table 2. The magnitude of UMSE for any given value of Mahalanobis distance Δ_0 show the advantage of SO estimator against SD estimator since $\Delta UMSE = UMSE^{SD} - UMSE^{SO} > 0$. The tables show that the effect of separation level between populations is evident, i.e., estimators and their UMSE decreases as Δ_0 increases.

5. Conclusions

For the spatial data modelled by two types of GRF the novel actual error rate estimators for classifiers based on PBDF are proposed and explored. Performances of the proposed estimators are evaluated by UMSE criterion via simulation experiments.

The simulation study with a moderate size of training samples shows that for both types of GRF models SO estimator has an advantage against SD estimator with respect to the minimum UMSE criterion.

Hence, adding the supplementary term to the $\hat{P}^{SD}$ slightly improves the effectiveness of estimator. That fact confirms the usefulness of deriving the expected error rate approximations for classifiers of spatial data modelled by GRF.

References

Batsidis A, Zografos K (2011). "Errors of Misclassification in Discrimination of Dimensional Coherent Elliptic Random Field Observations." *Statistica Neerlandica*, **65**(4), 446–461.

Berrett C, Calder CA (2016). "Bayesian Spatial Binary Classification." *Spatial Statistics*, **16**, 72–102.

Cressie N, Wikle CK (2011). *Statistics for Spatio-Temporal Data*. Wiley, Hoboken, NJ.

De Oliveira V, Ferreira MAR (2011). "Maximum Likelihood and Restricted Maximum Likelihood Estimation for a Class of Gaussian Markov Random Fields." *Metrika*, **74**(2), 167–183.

Diggle PJ, Ribeiro PJ, Christensen OF (2003). "In Moller J. (ed) Spatial Statistics and Computational Methods: Lecture Notes in Statistics." *R News*, **173**, 43–86.

Dreiziene L, Ducinskas K (2020). "Comparison of Spatial Linear Mixed Models Based on the Correct Classification Rates." *Spatial statistics*, **35**(online).

Ducinskas K (2009). "Approximation of the Expected Error Rate in Classification of the Gaussian Random Field Observations." *Statistics and Probability Letters*, **79**, 138–144.

Ducinskas K, Dreiziene L (2011). "Supervised Classification of the Scalar Gaussian Random Field Observations under a Deterministic Spatial Sampling Design." *Austrian Journal of Statistics*, **40**(1,2), 25–36.

Ducinskas K, Dreiziene L (2018). "Risks of Classification of the Gaussian Markov Random Field Observations." *Journal of Classification*, **35**, 422–436.

Ducinskas K, Dreiziene L (2019). "Spatial Model Selection Based on Hybrid Performance Measure of Linear Classifier." *Computer Data Analysis and Modeling: Stochastics and Data Science: Proc. of the Twelfth Intern. Conf., Minsk*, pp. 17–20.

Egbo I (2016). "Evaluation of Error Rate Estimators in Discriminant Analysis with Multivariate Binary Variables." *American Journal of Theoretical and Applied Statistics*, **5**(4), 173–179.

Haining R (1990). *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge University Press, Cambridge, UK.

Lachenbruch PA, Mickey MR (1968). "Estimation of Error Rates in Discriminant Analysis." *Technometrics*, **10**(1), 1–11.

McLachlan GJ (2004). *Discriminant Analysis and Statistical Pattern Recognition.* Wiley: New York.

Rue H, Held L (2005). *Gaussian Markov Random Fields: Theory and Applications.* Boca Raton: Chapman.

Sain SR, Cressie N (2007). "A Spatial Model for Multivariate Spatial Data." *Journal of Ekonometrics*, **140**, 229–259.

Snapinn SM, Knoke JD (1984). "Classification Error Rate Estimators Evaluated by Unconditional Mean Squared Error." *Technometrics*, **26**(2), 371–378.

Song JJ, De Oliveira V (2012). "Bayesian Model Selection in Spatial Lattice Models." *Statistical Methodology*, **9**, 228–238.

Affiliation:

Kestutis Ducinskas
Klaipeda University
H. Manto 84, Klaipeda, Lithuania
E-mail: k.ducinskas@gmail.com