Nonlinear Analysis: Modelling and Control, Vol. 28, No. 2, 250–263
https://doi.org/10.15388/namc.2023.28.31434

Vilnius University Press

Performance of the supervised generative classifiers of spatio-temporal areal data using various spatial autocorrelation indexes

Marta Karaliutė[a,b], **Kęstutis Dučinskas**[a,b]

[a]Institute of Data Science and Digital Technologies,
Vilnius University, Vilnius, Lithuania
marta.karaliute@ku.lt

[b]Faculty of Marine Technologies and Natural Sciences,
Klaipėda University, Klaipėda, Lithuania

Received: September 2, 2022 / **Revised:** January 19, 2023 / **Published online:** February 22, 2023

Abstract. This article is concerned with a generative approach to supervised classification of spatio-temporal data collected at fixed areal units and modeled by Gaussian Markov random field. We focused on the classifiers based on Bayes discriminant functions formed by the log-ratio of the class conditional likelihoods. As a novel modeling contribution, we propose to use decision threshold values induced by three popular spatial autocorrelation indexes, i.e., Moran's I, Geary's C and Getis–Ord G. The goal of this study is to extend the recent investigations in the context of geostatistical and hidden Markov Gaussian models to one in the context of areal Gaussian Markov models. The classifiers performance measures are chosen to be the average accuracy rate, which shows the percentage of correctly classified test data, balanced accuracy rate specified by the average of sensitivity and specificity and the geometric mean of sensitivity and specificity. The proposed methodology is illustrated using annual death rate data collected by the Institute of Hygiene of the Republic of Lithuania from the 60 municipalities in the period from 2001 to 2019. Classification model selection procedure is illustrated on three data sets with class labels specified by the threshold to mortality index due to acute cardiovascular event, malignant neoplasms and diseases of the circulatory system. Presented critical comparison among proposed approach classifiers with various spatial autocorrelation indexes (decision threshold values) and classifier based hidden Markov model can aid in the selection of proper classification techniques for the spatio-temporal areal data.

Keywords: separable covariance function, Bayes discriminant function, spatial weights, confusion matrix, decision threshold values.

1 Introduction

Spatial classification is an important task in spatio-temporal data mining. A comprehensive literature survey on state-of-the-art advances in spatio-temporal data mining is proposed by Hamdi et al. [15]. Compared with the general classification problem, spatial classification needs to consider the location information of the data and the interaction among feature and label variables at every time moment. Bayesian classifiers are one of the most common spatial classification tools. There are currently a variety of ways to achieve the classification goal, but one of the most effective one is to use the discriminant function.

In this study, we focus on the generative approach to supervised classification based discriminant functions formed by log-ratio of spatio-temporal Gaussian conditional distributions for features to enrich and generalize the traditional discriminant function theory.

Often before analyzing spatio-temporal datasets, spatio-temporal discretization (or aggregation) is applied. The discretization is useful to summarize information and help in extracting features within a spatio-temporal range rather than measuring a single point [14].

The generative approach to supervised classification is studied by numerous authors (see, e.g., [21, 26]). In the framework of this approach, the labeling of observations is based on the class-conditional feature distributions and information about various adjacency relationships. Spatial contextual classification problems arising in geospatial domain has been studied in the vast literature (see, e.g., [1, 27]). It is usually assumed that once the class label is known, feature observations are independent. This approach comprises the generative classifiers based on naive Bayes or hidden Markov models (HMMs) (see, e.g., [21]). Dučinskas [8] proposed and explored supervised generative classification rules for spatial Gaussian data by avoiding assumption of conditional independence. Comprehensive overview of methods for statistical classification and discrimination of Gaussian spatial data is provided by Berret and Calder [2]. The novel approach to classification of Gaussian Markov random fields (GMRF) observation on the lattice is developed by Dreižienė and Dučinskas [7] and Dučinskas and Dreižienė [9]. Some authors have investigated the performance of the Bayes classification rules when training samples consist of temporally dependent observations (see, e.g., [19]).

In the environmental agricultural and other research, data are often collected across space and through time. The spatio-temporal data are usually recorded at regular time intervals (time lags) and at irregular stations (areas) in compact area (see, e.g., [6]). Modeling and prediction of a such type of data has been studied by the numerous authors (see, e.g., [3, 5]).

However, generative statistical classification of spatio-temporal data has been rarely considered previously (see, e.g., [17, 33]). It should be noted that deep learning methods based on discriminative approach has been intensively studied now [34, 35], but these methods are off scope of this paper.

In the present paper, we focus on the classification of spatio-temporal areal data modeled by GMRF with separable spatio-temporal covariance specified by spatial margins and discrete temporal margins (see, e.g., [5]). Areal data consist of observations collected at subregions that form a partition of a region of interest. This data usually represent summaries of a quantity interest over subregions over a period of time.

Our approach comprises the cases when the *class label can vary across areal units and over time*. This assumption essentially widens the application area of the proposed classification method, especially for the cases with the imbalanced data. Separability of covariances was assumed for the sake of reduction of complexity due to interdependencies between features. The general objective of this study is to extend the previous investigations of spatio-temporal point referenced geostatistical data to spatio-temporal areal data. As a novel modeling contribution, we propose to use decision threshold values induced by three popular spatial autocorrelation indexes, i.e., Moran's I, Geary's C and Getis–Ord G.

Performance criterion of the classifier based on plug-in Bayes discriminant function realized on previously described spatio-temporal data models is chosen to be the accuracy rate (ACC), which shows the percentage of correctly classified test data, balanced accuracy rate (BAC) and geometric accuracy rate (GAC) evaluated from the confusion matrices. For numerical illustrations, the HCAR model for spatial covariance (see, e.g., [29]) and AR(p) model for temporal covariance are considered. This is the extension of the generative classification method for point referenced spatio-temporal data developed by Karaliutė and Dučinskas [16,17]. Critical analysis of the proposed classifiers is carried out for decision threshold values induced by several class label distributions. Critical comparison among proposed approach classifiers with various spatial autocorrelation indexes (decision threshold values) and classifier based HMM is performed.

In order to carry out the study, this paper is organized as follows. First, Section 2 introduces the models for spatial-temporal Gaussian data. Next, Section 3 introduces classifiers based on Bayes discriminant functions. The numerical analysis of annual death rate data collected by the Institute of Hygiene of the Republic of Lithuania from the 60 municipalities in the period from 2001 to 2019 is presented in the experimental study section. Finally, conclusions and comments are made in the last section.

2 Spatio-temporal data models and conditional distributions

The main objective of this paper is to classify scalar feature observations of GMRF $\{Z(s;t)\colon\ s \in D \subset \mathbb{R}^2,\ t \in D_T = [0,\infty]\}$, where s and t define spatial and temporal indexes, respectively. Let $\{Y(s;t)\colon\ s \in D \subset \mathbb{R}^2,\ t \in D_T\}$ be a random field that represents class label and takes only the value 0 or 1 with prior probabilities $\pi_0(s,t)$ and $\pi_1(s,t)$, respectively (see, e.g., [30]).

In this study, we assume that for $l = 0, 1$, the model of observation $Z(s;t)$ *conditional on* $Y(s;t) = l$ is

$$Z(s;t) = \mu_l(s;t) + \varepsilon(s;t),$$

where $\mu_l(s;t)$ is deterministic spatio-temporal mean. The error term is assumed to be generated by the univariate zero-mean GMRF $\{\varepsilon(s;t)\colon\ s \in D \subset \mathbb{R}^2,\ t \in D_T\}$, with covariance function defined by model $\operatorname{cov}(\varepsilon(s;t), \varepsilon(u;r)) = C(s,u;t,r)$ for all $s,u \in D$ and $t,r \in T$.

In the present paper, we restrict our attention to the separable spatio-temporal covariance model $C(s,u;t,r) = C_S(s,u)C_T(t,r)$, where $C_S(s,u)$ denotes pure spatial covariance between observations in areas s and u, and $C_T(t,r)$, denotes pure temporal covariance between observations at time points t and r.

Under this assumption, the spatio-temporal covariance structure factors into a purely spatial and a purely temporal component, which allows for computationally efficient estimation and inference. Therefore, separable covariance models have been popular even in situations in which they are not physically justifiable. Many statistical tests for separability have been proposed recently and are based on parametric models (see, e.g., [4, 12]) or spectral methods [25].

Let $S = \{s_i \in D, i = 1, \dots, n\}$ be a set of n nonoverlapping areal units (areas). Assume that S is endowed with a neighbourhood system $N = \{N_i, i = 1, \dots, n\}$, where N_i denotes the collection of areas that are neighbours of area s_i. Usually, the neighborhood N_i could be defined to be those areas with which area s_i shares a common border. Data is recorded for each areal unit at consecutive time periods at $t \in D_T = 1, 2, \dots, m, m+1$.

At every time moment t, denote by n_{lt} the number of areas with $Y(s,t) = l$. Thus $n = n_{0t} + n_{1t}$ for every $t \in D_T$. Hence a set of class labels at any time moment can differ in composition.

As it follows, the transpose of any matrix A is denoted by the symbol A'.

Suppose the training sample consisting of m temporal feature observations in S_n is specified by $n \times m$ matrix $Z = (Z_1, \dots, Z_m)$, where $Z_t = (Z(s_1,t), \dots, Z(s_n,t))'$.

This structure of data presentation is motivated by a model that assumes multivariate (in space) time series. Denote by $z_t = (z_t^1, \dots, z_t^n)$ and $y_t = (y_t^1, \dots, y_t^n)$ the realized value of Z_t and $Y_t = (Y(s_1,t), \dots, Y(s_n,t))'$, respectively.

Set $\bar{z}_t = \sum_{i=1}^n z_t^i/n$ and $\tilde{z}_t = z_t - \bar{z}_t 1_n$, where 1_n denotes an n-dimensional column vector whose elements are all equal 1.

Specify n-dimensional quadratic matrix $W = (w_{ij})$, $i,j = 1, \dots, n$, with (i,j) element

$$w_{ij} = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{if } s_j \in N_i, \\ 0 & \text{otherwise.} \end{cases}$$

To be specific, n-dimensional random vector Z_t follows the Gaussian Markov graphical model specified by the family of joint distributions defined on the undirected graph $G = (S; E)$, with S as set of nodes and edges E corresponding the endowed neighborhood system N with spatial weights matrix W.

Define by $H = D - W$ the Laplace matrix, where D is the diagonal matrix with diagonal elements $d_{ii} = \sum_{s_j \in N_i} w_{ij}$, $i = 1, \dots, n$.

Global Moran's I, Geary's C and Getis–Ord G for n areas at the time moment t are

$$I(t) = \frac{n}{S_0} \frac{\tilde{z}_t' W \tilde{z}_t}{\tilde{z}_t' \tilde{z}_t}, \qquad C(t) = \frac{(n-1)}{2S_0} \frac{z_t' L z_t}{\tilde{z}_t' \tilde{z}_t}, \qquad G(t) = \frac{z_t' W z_t}{z_t'(J - I) z_t},$$

where

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}, \qquad J = 1_n 1_n'.$$

It should be noted that the expected value of Moran's I is $-1/(n-1)$. Values of I that exceed $-1/(n-1)$ indicate positive spatial autocorrelation in which similar values, either high values or low values, are spatially clustered. Values of I below $-1/(n-1)$ indicate negative spatial autocorrelation in which neighboring values are dissimilar.

The theoretical expected value for Geary's C is 1. A value of Geary's C less than 1 indicates positive spatial autocorrelation, while a value larger than 1 points to negative spatial autocorrelation.

Geary's C is inversely related to Moran's I, but it is not identical. Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation. Geary's C is also known as Geary's contiguity ratio or simply Geary's ratio.

The Getis–Ord G-statistic (see, e.g., [11,13]) distinguishes between hot spots and cold spots. It identifies spatial concentrations, i.e., G is relatively large if high values cluster together and G is relatively low if low values cluster together.

In what follows, with an insignificant loss of generality, we focus on the linear independent of time mean $\mu_l(s;t) = \beta_l' x(s)$, where $x(s) = (x_1(s), \dots, x_q(s))'$ is the vector of an explanatory variables, and β_l is a q-dimensional vector of parameters, $l = 0, 1$.

Denote by X the $n \times 2qm$ matrix $X = (X_{(1)}, X_{(2)}, \dots, X_{(m)})$, where

$$X_{(t)} = \begin{pmatrix} x_1'(1 - y_t^1) & x_1' y_t^1 \\ x_2'(1 - y_t^2) & x_2' y_t^2) \\ \vdots & \vdots \\ x_n'(1 - y_t^n) & x_n' y_t^n \end{pmatrix},$$

and $x_i = x(s_i)$, $i = 1, \dots, n$.

Then the matrix model for Z conditional on $\{Y_t = y_t,\ t = 1, \dots, m\}$ is

$$Z = XB + E, \tag{1}$$

where $B = I_m \otimes \beta$ with $2q \times 1$-dimensional parameter vector $\beta = (\beta_0', \beta_1')'$, and $n \times m$ matrix of Gaussian errors $E = (\varepsilon(s_i;t),\ i = 1, \dots, n,\ t = 1, \dots, m)$, and I_m is $m \times m$ identity matrix.

The covariance separability assumption implies that $\mathrm{vec}(E) \sim N_{nm}(0,\ C_T \otimes C_s)$ with $C_T = (c_T^{tr} = C_T(t,r),\ t,r = 1, \dots, m)$ denoting the $m \times m$ matrix of pure temporal covariances and $C_S = (c_S^{ij} = C_S(s_i, s_j),\ i,j = 1, \dots, n)$ denoting the $n \times n$ matrix of pure spatial covariances.

Then under spatio-temporal data model specified in Eq. (1) it follows that

$$Z \sim N_{n \times m}(XB,\ C_T \otimes C_S).$$

We concern with the problem of classification of the observations $Z(s_i,\ m+1)$, $i = 1, \dots, n$, into one of two classes with given joint training sample Z. Or in other words, based on the training sample information, we want to predict label of an observation of Z in every location at the time moment $t = m + 1$.

Set $c_T^{m+1,r} = C_T(m+1,\ r)$, $r = 1, \dots, m$, $c_T^{m+1} = (c_T^{m+1,1}, \dots, c_T^{m+1,m})'$ and e_i' – the ith row of identity matrix I_n.

Then we can conclude that the conditional distribution of $Z(s_i, m+1)$ given $Z = z$ and $Y(s_i, m+1) = l$ is Gaussian, i.e.,

$$\big((Z(s_i, m+1) \mid Z = z;\ Y(s_i, m+1) = l\big) \sim N\big(\mu_{li(z)}^{m+1}, \Sigma_{m+1,i(z)}\big), \quad l = 0, 1, \tag{2}$$

where

$$\mu_{li(z)}^{m+1} = \beta_l' x_i + \big(\big(c_T^{m+1}\big)' C_T^{-1} \otimes e_i'\big) \operatorname{vec}(E),$$
$$\Sigma_{m+1,i(z)} = \operatorname{var}\big(Z(s_i, m+1)\big) - c_s^{ii}\big(c_T^{m+1}\big)' C_T^{-1} c_T^{m+1} = c_S^{ii} \rho_{m+1}$$

with

$$\rho_{m+1} = c_T^{m+1,m+1} - \big(c_T^{m+1}\big)' C_T^{-1} c_T^{m+1}.$$

In this study, we assume that the conditional distribution of label $Y(s_i, m+1)$, $i = 1, \dots, n$, *given the joint training sample does not depend on features* Z *values in training areas, i.e.,* $\mathbf{P}(Y(s_i, m+1) = l \mid Z = z, Y = y) = \mathbf{P}(Y(s_i, m+1) = l \mid Y = y) = \pi_l(s_i, m+1)$, $l = 0, 1$. *These conditional probabilities specify the label distribution and induce decision threshold values.*

3 Classification of regressive spatio-temporal models

In the present paper, we focus on the spatio-temporal Gaussian areal data with Markov-type model for pure spatial covariance [5] when for geostatistical point referenced data, the Matérn or powered-exponential class (see, e.g., [5, Sect. 4.1.1], [31, p. 31]) are the most popular among statisticians.

Specific attention is given to the Gaussian spatio-temporal model with pure spatial covariances belonging to the HCAR models and with pure temporal covariance of stationary AR(p) model.

This spatial covariance matrix for n areas $C_s = \sigma_s^2 R$, where $R = (r_{ij})$ denotes the spatial correlation matrix with $R = (I_n + \eta H)^{-1}$. Here $\eta \geqslant 0$ is a spatial dependence parameter, and $\sigma_s > 0$ is a scale parameter. Then spatial precision matrix is defined by $\Omega_S = C_s^{-1} = (I_n + \eta H)/\sigma_s^2$.

For AR(p) model of temporal covariance, we have the following conditional moment functions: $\mu_{li(z)}^{m+1} = \mu_{li}^{m+1} + ((0, \dots, 0, \alpha_p, \dots, \alpha_2, \alpha_1)' \otimes e_i') \operatorname{vec}(E)$ and $\Sigma_{m+1,i(z)} = c_S^{ii} \sigma_T^2$, where $c_S^{ii} = \sigma_s^2 (1 + \eta h_i)^{-1}$.

Under the assumption that the classes are completely specified, the well-known Bayes discriminant function (BDF) minimizing the total probability of misclassification is formed by the log-ratio of conditional likelihood of distributions (see [19]) specified in Eq. (2), that is,

$$W_Z\big(Z(s_i, m+1)\big) = L_Z\big(Z(s_i, m+1)\big) - \gamma_i(m+1), \tag{3}$$

where $\gamma_i(m+1) = \ln(\pi_0(s_i, m+1)/\pi_1(s_i, m+1))$ and

$$L_Z\big(Z(s_i, m+1)\big) = \left(Z(s_i, m+1) - \frac{\mu_{1i(z)}^{m+1} + \mu_{0i(z)}^{m+1}}{2}\right) \times \Sigma_{m+1,\, i(z)}^{-1} \big(\mu_{1i(z)}^{m+1} - \mu_{0i(z)}^{m+1}\big).$$

So BDF classifies the observation $Z(s_i,\, m+1)$ in the following way: class label takes value 1 if $L_Z(Z(s_i,\, m+1) \geqslant \gamma_i(m+1)$, and 0 otherwise.

So $L_Z(Z(s_i,\, m+1)$ *is a linear term, and* $\gamma_i(m+1)$ *plays role of a decision threshold.*

The probability of misclassification for $W_Z(Z(s_i,\, m+1))$ is optimal under the criterion of the minimum of misclassification probability.

However, in practical applications all statistical parameters of populations are rarely known.

Then the estimators of unknown parameters found from the given training sample are plug-in BDF specified in Eq. (3).

Replacing parameters with their estimators for BDF, we denote the plug-in Bayes discriminant function (*PBDF*):

$$\widehat{W}_Z\bigl(Z(s_i,\, m+1)\bigr) = \left(Z(s_i,\, m+1) - \frac{\hat{\mu}_{1i(z)}^{m+1} + \hat{\mu}_{0i(z)}^{m+1}}{2}\right) \times \Sigma_{m+1,i(z)}^{-1}\bigl(\hat{\mu}_{1i(z)}^{m+1} - \hat{\mu}_{0i(z)}^{m+1}\bigr) - \gamma_i(m+1).$$

So PBDF has the same threshold as BDF, but differs in the linear terms.

In the present paper, we apply the averaged ML estimator of β and σ_s^2 that for any fixed $\eta \geqslant 0$, is, respectively,

$$\widehat{\beta} = \sum_{t=1}^{m} \frac{\widehat{\beta}_{(t)}}{m}, \qquad \widehat{\sigma}_s^2 = \sum_{t=1}^{m} \frac{\widehat{\sigma}_{s(t)}^2}{m},$$

where

$$\widehat{\beta}_{(t)} = \bigl(X'_{(t)} R^{-1} X_{(t)}\bigr)^{-1} X'_{(t)} R^{-1} Z_t$$

and

$$\widehat{\sigma}_{s(t)}^2 = \frac{1}{n-2q}\bigl(Z_t - X_{(t)}\widehat{\beta}\bigr)' R^{-1}\bigl(Z_t - X_{(t)}\widehat{\beta}\bigr).$$

Three models label distribution for observation in s_i at $t = m+1$ are proposed. They differ on the level of the incorporated spatio-temporal information.

It is obvious that each model for label distribution specifies specific decision threshold value $\gamma_i(m+1)$ for considered classifier.

Label distribution based on Moran's I, Geary's C and Getis–Ord G is denoted by

$$\pi_{M1t}(s_i,\, m+1) = \frac{1}{1 + \mathrm{e}^{-I(m)y_i^*(m)}}, \qquad \pi_{C1t}(s_i,\, m+1) = \frac{1}{1 + \mathrm{e}^{-C(m)y_i^*(m)}}$$

and

$$\pi_{G1t}(s_i,\, m+1) = \frac{1}{1 + \mathrm{e}^{-G(m)y_i^*(m)}},$$

respectively, where $y_i^*(t) = 2y_i(t) - 1$.

Recall that HMM approach of classification is based on assumption of conditional independence for feature observations and first-order Markov property for labels. In this article, we restricted our attention on Gaussian observation with regression mean model

Table 1. Confusion matrix for two-class problem.

	$\widehat{Y}(s_i, m+1)$	
$Y(s_i, m+1)$	0	1
0	a	b
1	c	d

and constant variance for each areal unit [20, 22]. The spatially weighted estimators of regression coefficients, variances and transition probabilities are inserted in the PBDF.

Performance criteria of the generative classifier based on PBDF is evaluated by confusion matrix formed for test data and fixing the results of correctly and incorrectly recognized test observations of each class.

This procedure is realized via partitioning the observed data into training and testing sets. Then classifier being designed on the training data, and its accuracy being validated on the test data. In this paper, our focus is on using m temporal observations for training, and the observations at time moment $t = m + 1$ are used for testing.

Label prediction in areal unit s_i at time moment $t = m + 1$ is $\widehat{Y}(s_i, m+1) = H(\widehat{W}_Z(Z(s_i, m+1))$, where $H(\cdot)$ is the Heaviside step function.

Let

$$\sum_{i=1}^{n} I\bigl(Y(s_i, m+1) = 0\bigr) I\bigl(\widehat{Y}(s_i, m+1) = 0\bigr) = a,$$
$$\sum_{i=1}^{n} I\bigl(Y(s_i, m+1) = 0\bigr) I\bigl(\widehat{Y}(s_i, m+1) = 1\bigr) = b,$$
$$\sum_{i=1}^{n} I\bigl(Y(s_i, m+1) = 1\bigr) I\bigl(\widehat{Y}(s_i, m+1) = 0\bigr) = c,$$
$$\sum_{i=1}^{n} I\bigl(Y(s_i, m+1) = 1\bigr) I\bigl(\widehat{Y}(s_i, m+1) = 1\bigr) = d,$$

where $I(A)$ denotes the indicator of event A.

The confusion matrices that will be applied for the assessment of the proposed classifier performances are shown in Table 1.

Traditionally, the most commonly used empirical measure of classifier performance is called *accuracy rate* that shows the percentage of correctly classified test data given by the formula

$$\mathrm{ACC} = \frac{a + d}{a + b + c + d}. \tag{4}$$

However, in practice, situations with significant disbalance between the majority and minority class examples frequently occur (see, e.g., [18, 23, 24]). Then the evaluation of the classifiers' performance must be carried out using specific metrics to take into account the class distribution. In this article, we also used two other performance evaluation measures based on confusion matrix.

The first approximation of AUC is usually called balanced accuracy (see, e.g., [18,28]) and is specified by the formula

$$\mathrm{BAC} = \frac{1}{2}\left(\frac{a}{a+b} + \frac{d}{c+d}\right). \tag{5}$$

The second G-mean is important to measure the avoidance of the overfitting to the minority class and the degree to which the majority class is marginalized (see, e.g., [32]). It is specified by the formula

$$\mathrm{GAC} = \sqrt{\frac{a}{a+b} \cdot \frac{d}{c+d}}. \tag{6}$$

Below is a receiver operator characteristic (ROC) chart [25] that allows to visualize the trade-off between *sensitivity* (*true positive rate*), i.e., $\mathrm{TPR} = a/(a+b)$ and 1-*specificity* (*false positive rate*), i.e., $\mathrm{FPR} = d/(c+d)$ for any confusion matrix corresponding to selected decision threshold value.

4 Experimental study

The numerical analysis of annual death rate data collected by the Institute of Hygiene of the Republic of Lithuania from the 60 municipalities in the period from 2001 to 2019 is carried out.

For numerical illustrations of obtained results, we considered the Gaussian spatio-temporal model with pure spatial exponential covariances and with pure temporal covariance of stationary AR(1).

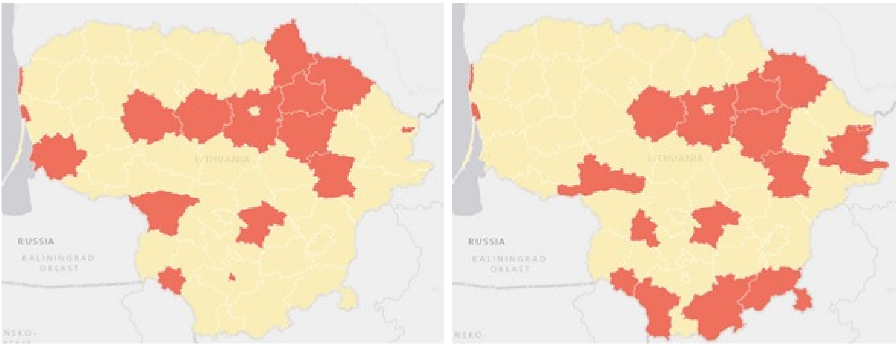
Crude death rate for each municipality measured in units per one hundred thousand population is considered as variable Z. Three class label variables Y are specified by the threshold to mortality index due to acute cardiovascular event, malignant neoplasms and diseases of the circulatory system. Denote these label variables by ACE, MN and CSD, respectively. For the cases with index values less than conventional threshold, label variable takes value 0, and takes value 1 otherwise. Here we consider the case with constant mean, i.e., $\mu(s;t) = \beta_l x(s)$ with $x(s) = 1$.

Then for $t = 1, \dots, 18$,

$$X_{(t)} = \begin{pmatrix} 1-y_1^t & y_1^t \\ 1-y_2^t & y_2^t \\ \vdots & \vdots \\ 1-y_n^t & y_n^t \end{pmatrix}.$$

Numerical illustrations are performed on 60 areas in two-dimensional areas that are depicted in Fig. 1.

Data in the period from 2001 to 2018 ($t = 1, \dots, 18$) are used for training, and remaining data (period 2019) are used for testing. Hence $m = 18$ and $n = 60$, i.e., we consider $18 \cdot 60$ observations for training and 60 for testing.



(a) (b)

Figure 1. Classified Lithuanian municipalities in 2001 (a) and 2002 (b). Yellow color areas indicate municipalities with low level of mortality due to ACE (with label value 0), and red areas indicate municipalities with high level of mortality due to ACE (with label value 1).

Table 2. Annual imbalance ratio for various mortality reasons (class label variables).

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
ACE	1.73	2.33	2.16	2	1.61	4.45	2.16	3.29	2.33	1.73
MN	2.33	3	2.53	3.62	1.5	1.31	1.5	1.4	4	2.16
CSD	4	4	2.75	4.45	5	3.62	3	3	5	4.45

	2011	2012	2013	2014	2015	2016	2017	2018	2019
ACE	2.33	1.4	1.22	1.14	1.86	0.76	0.71	0.5	0.3
MN	5	3	2.33	4.45	4.45	5	3.62	5.67	7.57
CSD	4	19	6.5	5	5.67	5	4	4.45	3

For $t = 1, \dots, m, m + 1$, denote by $\mathrm{IR}_t = n_1^t / n_0^t$ the imbalance ratio. The graph of IR_t is depicted in Table 2.

As we can see in Table 2, in majority of time periods, CSD has the highest IR, the next in the row is MN case, and the ACE has the lowest IR.

The values performance measures of the proposed classifier specified in Eq. (4)–(6) *for three label distribution models and various class label variables and classifier based on spatial* HMM (*see* [10]) *are presented in Table* 3.

As it might be seen from Table 3, the methods based on the incorporation of the spatial index Geary's C has an advantage against two others in all performance measures for case with label variable ACE (i.e., the case with lowest IR), incorporation of Moran's I ensures the highest performance for the label variable MN (i.e., the case with lowest IR), and at last, the incorporation of Getis–Ord G ensures the highest performance for the label variable CSD (i.e., the case with highest IR).

ROC plot of presented in Fig. 2 visualizes the performances of proposed classifier for decision threshold values induced by different spatial autocorrelation indexes and classifier based on HMM. Depicted points represent four classifiers and a random classifier depicted by dotted line. It is easy to check that the area under the curve in the ROC plot is equal to the performance measure BAC.

Table 3. Performance measures of classifiers based on PBDF (numbers in bold indicate highest values of performance measures).

		Moran's I	Geary's C	Getis–Ord G	HMM
ACE	ACC	0.7333	**0.7667**	0.7500	0.6500
	BAC	0.7019	**0.7236**	0.6630	0.6723
	GAC	0.6994	**0.7191**	0.6427	0.6711
MN	ACC	0.8000	0.8333	**0.8500**	0.7667
	BAC	0.5768	0.7197	**0.7291**	0.5800
	GAC	0.4980	0.7042	**0.7119**	0.4870
CSD	ACC	**0.8000**	0.7667	0.7333	0.7833
	BAC	**0.6889**	0.6667	0.6222	0.6778
	GAC	**0.6521**	0.6360	0.5812	0.6441

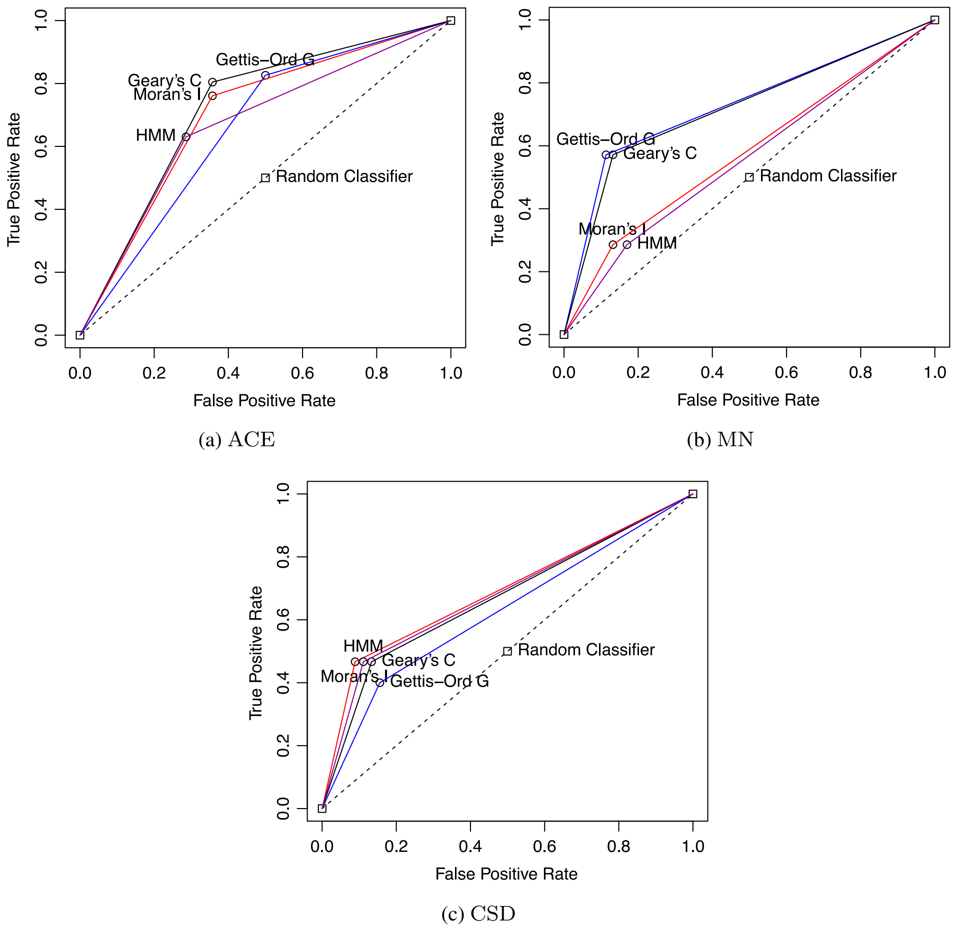


Figure 2. ROC plots for the problems with class label variables ACE, MN, CSD and HMM.

5 Conclusions

In this paper, we developed the novel generative approach to classification of areal Gaussian data based on Bayes discriminant functions and separable spatio-temporal covariances comprising the situations when the class label can vary across areal units and over time. A real data study has been conducted, and critical comparison of the performance of the classifiers with decision threshold values induced by different spatial autocorrelation indexes and classifiers based on HMM are performed.

The proposed methodology has several attractive features that make it compare favorably against other approaches to generative supervised spatial classification. First, the method is applicable to both regular and irregular areal units (or lattice models of spatial data) and avoids strict conditional independence assumption as in HMM. Second, the approach provides an easily interpretable measure of competing classifiers performance through various spatial autocorrelation indexes. Third, spatial data model can be easily specified in terms of undirected graphs.

The obtained results with real data showed that the value of IR significantly influenced the performance of proposed classifiers:

- The methods based on the incorporation of the spatial index Geary's C has an advantage against two others in all performance measures for case with label variable ACE (i.e., the case with the lowest IR);
- The incorporation of Moran's I ensures the highest performance for the label variable MN (i.e., the case with an intermediate IR);
- At last, the incorporation of Getis–Ord G ensures the highest performance for the label variable CSD (i.e., the case with highest IR).

Finally, in majority cases of considered real data examples, the proposed approach to generative classification shows better performance than classification based on HMM.

There are several reasons for further research. First, there is further scope for exploring techniques supervising classification of spatio-temporal Gaussian data with nonseparable covariance models. Second, future research may, also, include implementation of the proposed classification technique in the context of spatio-temporal non-Gaussian models. Finally, further research involving more covariates could help gain more insights into the relative strength of the rival classifiers.

References

1. P.M. Atkinson, P. Lewis, Geostatistical classification for remote sensing: an introduction, *Comput. Geosci.*, **26**(4):361–371, 2000, `https://doi.org/10.1016/S0098-3004(99)00117-X`.

2. C. Berrett, C.A. Calder, Bayesian spatial binary classification, *Spatial Stat.*, **16**:72–102, 2016, `https://doi.org/10.1016/j.spasta.2016.01.004`.

3. M. Blangiardo, M. Cameletti, *Spatial and Spatio-Temporal Bayesian Models with R-Inla*, John Wiley & Sons, 2015.

4. P.E. Brown, K.F. Karesen, G.O. Roberts, S. Tonellato, Blur-generated non-separable space-time models, *J. R. Stat. Soc., Ser. B*, **62**(4):847–860, 2000, https://doi.org/10.1111/1467-9868.00269.
5. N. Cressie, C.K. Wikle, *Statistics for Spatio-Temporal Data*, Wiley, Hoboken, NJ, 2011.
6. S.S. Demel, J. Du, Spatio-temporal models for some data sets in continuous space and discrete time, *Stat. Sin.*, **25**:81–98, 2015, https://doi.org/10.5705/ss.2013.223w.
7. L. Dreižienė, K. Dučinskas, Comparison of spatial linear mixed models for ecological data based on the correct classification rates, *Spatial. Stat.*, **35**, 2020, https://doi.org/10.1016/j.spasta.2019.100395.
8. K. Dučinskas, Approximation of the expected error rate in classification of the Gaussian random field observations, *Stat. Probab. Lett.*, **79**:138–144, 2009, https://doi.org/10.1016/j.spl.2008.07.042.
9. K. Dučinskas, L. Dreižienė, Risks of classification of the Gaussian Markov random field observations, *J. Classif.*, **35**:422–436, 2018, https://doi.org/10.1007/s00357-018-9269-7.
10. K. Dučinskas, M. Karaliutė, L. Šaltytė-Vaisiauskė, Spatially weighted Bayesian classification of spatio-temporal areal data based on Gaussian-hidden Markov models, *Mathematics*, **11**(2), 2023, https://doi.org/10.3390/math11020347.
11. M.M. Fischer, J. Wang, *Spatial Data Analysis: Models, Methods, and Techniques*, Springer, New York, 2011, https://doi.org/10.1007/978-3-642-21720-3.
12. M. Fuentes, Testing for separability of spatial-temporal covariance functions, *J. Stat. Plann. Inference*, **136**:447–466, 2006, https://doi.org/10.1016/j.jspi.2004.07.004.
13. A. Getis, J.K. Ord, The analysis of spatial association by use of distance statistics, *Geogr. Anal.*, **24**(3):189–206, 1992.
14. F. Giannotti, D. Pedreschi, *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*, Springer, 2008, https://doi.org/10.1007/978-3-540-75177-9.
15. A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S.K. Rumi, F.D. Salim, Spatiotemporal data mining: A survey on challenges and open problems, *Artif. Intell. Rev.*, **55**:1441–1488, 2022, https://doi.org/10.1007/s10462-021-09994-y.
16. M. Karaliutė, K. Dučinskas, Classification of Gaussian spatio-temporal data with stationary separable covariances, *Nonlinear Anal. Model. Control*, **26**(2):363–374, 2021, https://doi.org/10.15388/namc.2021.26.22359.
17. M. Karaliutė, K. Dučinskas, Supervised linear classification of Gaussian spatio-temporal data, *Liet. Mat. Rink., Proc. Lith. Math. Soc., Ser. A*, **62**:9–154, 2021, https://doi.org/10.15388/LMR.2021.25214.
18. V. López, A. Fernández, S. García, V. Palade, F. Herrera, An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics, *Inf. Sci.*, **250**:113–141, 2013, https://doi.org/10.1016/j.ins.2013.07.007.
19. G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York, 2004.

20. S.K. Kadhem nad P. Hewson, I. Kaimi, Using hidden Markov models to model spatial dependence in a network, *Aust. N. Z. J. Stat.*, **60**(4):423–446, 2018, https://doi.org/10.1111/anzs.12250.
21. A.Y. Ng, M.I. Jordan, On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes, in S. Becker, S. Thrun, K. Obermayer (Eds.), *Advances in Neural Information Processing System 15, December 9–14, 2002, Vancouver, Canada*, MIT Press, 2022, pp. 841–848.
22. L. Nguyen, Continuous observation hidden Markov model, *Revista Kasmera*, **44**(6):65–149, 2016.
23. A. Orriols-Puig, E. Bernadó-Mansilla, Evolutionary rule-based systems for imbalanced datasets, *Soft Comput.*, **13**(3):213–225, 2009, https://doi.org/10.1016/j.jmva.2006.01.009.
24. J. Ortigosa-Hernández, I. Inza, J.A. Lozano, Measuring the class-imbalance extent of multi-class problems, *Pattern Recognit. Lett.*, **98**:32–38, 2017, https://doi.org/10.1016/j.patrec.2017.08.002.
25. L. Scaccia, R.J. Martin, Testing axial symmetry and separability of lattice processes, *J. Stat. Plann. Inference*, **131**:19–39, 2005.
26. M. Sensoy, L. Kaplan, F. Cerutti, M. Saleki, Uncertainty-aware deep classifiers using generative models, in *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 04*, AAAI Press, Palo Alto, CA, 2020, pp. 5620–5627, https://doi.org/10.1609/aaai.v34i04.6015.
27. S. Shekhar, P.R. Schrater, R.R. Vatsavai, W. Wu, S. Chawla, Spatial contextual classification and prediction models for mining geospatial data, *IEEE Trans. Multimedia*, **4**(2):174–188, 2002, https://doi.org/10.1109/TMM.2002.1017732.
28. R. Soleymani, E. Granger, G. Fumera, *F-measure curves: A tool to visualize classifier performance under imbalance*, Pattern Recognit., 2020.
29. J.J. Song, V. De Oliveira, Bayesian model selection in lattice models, *Stat. Methodol.*, **9**:228–238, 2012, https://doi.org/10.1016/j.statmet.2011.01.003.
30. G. Stabingis, K. Dučinskas, L. Stabingienė, Comparison of spatial classification rules with different conditional distributions of class label, *Nonlinear Anal. Model. Control*, **19**(1):109–117, 2014, https://doi.org/10.15388/NA.2014.1.7.
31. M.L. Stein, *Interpolation of Spatial Data*, Springer, New York, 1999.
32. C. Su, Y. Hasio, An evaluation of the robustness of MTS for imbalanced data, *IEEE Trans. Knowl. Data Eng.*, **19**(10):1321–1332, 2007.
33. J. Šaltytė-Benth, K. Dučinskas, Linear discriminant analysis of multivariate spatial-temporal regressions, *Scand. J. Stat.*, **32**:281–294, 2005.
34. M. Zhao, C.H. Chang, W. Xie, Z. Xie, J. Hu, Cloud shape classification system based on multi-channel CNN and improved FDM, *IEEE Access*, **8**:44111–44124, 2020, https://doi.org/10.1109/ACCESS.2020.2978090.
35. Q. Zheng, M. Yang, J. Yang, Q. Zhang, X. Zhang, Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification, *Neural Comput. Appl.*, **33**:7723–7745, 2021, https://doi.org/10.1007/s00521-020-05514-1.