https://doi.org/10.15388/vu.thesis.376
https://orcid.org/0000-0002-1865-8017

VILNIUS UNIVERSITY

Eglė Zikarienė

Spatial Contextual Classification Based on Conditional Distributions Belonging to Elliptical and Exponential Families

DOCTORAL DISSERTATION

Natural Sciences,
Informatics (N 009)

VILNIUS 2022

The dissertation was prepared between 2013 and 2021 at Vilnius University. The research was supported by the Research Council of Lithuania.

Academic supervisor – Prof. Dr. Kęstutis, Dučinskas (Vilnius University, Natural Sciences, Informatics, N 009).
Academic consultant – Prof. Dr. Julius, Žilinskas (Vilnius University, Natural Sciences, Informatics, N 009).

This doctoral dissertation will be defended in a public/closed meeting of the Dissertation Defence Panel:

Chairman – Prof. Habil. Dr. Gintautas Dzemyda (Vilnius University, Natural Sciences, Informatics, N 009).
Members:
Prof. Dr. Audronė Jakaitienė (Vilnius University, Natural Sciences, Informatics, N 009),
Prof. Dr. Vacius Jusas (Kaunas University of Technology, Technological Sciences, Informatics Engineering, T 007),
Prof. Dr. Elijah Liflyand (Bar-Ilan University, Israel, Natural Sciences, Mathematics, N 001),
Prof. Dr. Virginijus Marcinkevičius (Vilnius University, Natural Sciences, Informatics, N 009).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 1:00 p.m. on the 29th of September, 2022 in Room 203 of the Institute of Data Science and Digital Technologies of Vilnius University.
Address: Akademijos str. 4, LT-04812, Vilnius, Lithuania.

The text of this dissertation can be accessed at the libraries of Vilnius University, as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.376
https://orcid.org/0000-0002-1865-8017

VILNIAUS UNIVERSITETAS

Eglė Zikarienė

Erdvinis kontekstinis klasifikavimas naudojant sąlyginius eliptinių ir eksponentinių šeimų skirstinius

DAKTARO DISERTACIJA

Gamtos mokslai,
Informatika (N 009)

VILNIUS 2022

Disertacija rengta 2013 – 2021 metais Vilniaus universitete.
Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

Mokslinis vadovas – prof. dr. Kęstutis Dučinskas (Vilniaus universitetas, gamtos mokslai, informatika, N 009]).
Mokslinis konsultantas – prof. dr. Julius Žilinskas (Vilniaus universitetas, gamtos mokslai, informatika, N 009).

Gynimo taryba:
Pirmininkas – prof. habil. dr. Gintautas Dzemyda (Vilniaus universitetas, gamtos mokslai, informatika, N 009).
Nariai:
prof. dr. Audronė Jakaitienė (Vilniaus universitetas, gamtos mokslai, informatika, N 009),
prof. dr. Vacius Jusas (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija, T 007),
prof. dr. Elijah Liflyand (Bar-Ilan universitetas, Izraelis, gamtos mokslai, matematika, N 001),
prof. dr. Virginijus Marcinkevičius (Vilniaus universitetas, gamtos mokslai, informatika, N 009).

Disertacija ginama viešame Gynimo tarybos posėdyje 2022 m. rugsėjo mėn. 29 d. 13 val. Vilniaus universiteto Duomenų mokslo ir skaitmeninių technologijų instituto 203 auditorijoje.
Adresas: Akademijos g. 4, LT-04812 Vilnius, Lietuva.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: https://www.vu.lt/naujienos/ivykiu-kalendorius

ACKNOWLEDGEMENTS

First of all, I would like to sincerely thank my academic supervisor Prof. Dr. Kęstutis Dučinskas for comprehensive support and confidence, for shared knowledge and scientific guidance, for showing direction, and for all the help provided while preparing this dissertation.

Thank you to the academic consultant Prof. Dr. Julius Žilinskas for valuable suggestions and comments.

I am also thankful to dissertation reviewers Prof. Habil. Dr. Gintautas Dzemyda and Prof. Dr. Virginijus Marcinkevičius for accurate insights, critical remarks and beneficial suggestions for improving this work.

I would also like to thank my family members for their encouragement, patience and constant support.

ACRONYMS

AER	Actual Error Rate
BA	Bayes Analysis
BR	Bayes Rule
BDF	Bayes Discriminant Function
CDF	Cumulative Distribution Function
DIC	Deviance Information Criterion
EER	Expected Error Rate
FO	Focal Observation
GAM	Generalized Additive Model
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
GRF	Gaussian Random Field
LDF	Linear Discriminant Function
LS	Least Squares
ML	Maximum Likelihood
MPL	Maximum Pseudo-Likelihood
MRF	Markov Random Field
NN	Nearest Neighbourhood
PDF	Probability Density Function
PLDF	Plug-in LDF
PBDF	Plug-in BDF
SABE	Spatial Auto-Beta
SRF	Spatial Random Field
STL	Set of Training Locations
STLG	STL with the grouped label
STLM	STL with the mixed label
TLC	Training Labels Configuration
TRF	T-distributed Random Field
WAIC	Watanabe Akaike's Information Criterion
ZIB	Zero-Inflated Beta

LIST OF SYMBOLS

The following list describes several symbols that will be later used within the body of the document:

$c(\bullet)$	a covariance function.
D	a spatial domain, of interest, $D \subset \mathbb{R}^n$.
d^{lk}	a conditional Mahalanobis distance between Ω_l and Ω_k, $l \neq k$ at spatial location s_0 conditional on *Z*, for two-class case it is denoted as d.
d_{ij}	an Euclid distance between two spatial locations s_i, s_j.
$\overline{EER}^{\delta}$	an empirical estimate of the EER, δ denotes the estimation method.
$H(\bullet)$	a Heaviside step function.
$I(\bullet)$	an indicator function.
i,j	a spatial location index.
k,l	a class index, $l \neq k$.
L	a testing sample size, for a different class the subscript is added, L_l.
M	a number of simulations.
m	a number of classes.
N	a training sample size, for a different class a the subscript is added, N_l.
NN_i	sites belonging to the NN of s_i, upper superscript is added to denote the class *l*, NN_i^l.
$P(\bullet)$	a marginal probability distribution, with $(\bullet\|\bullet)$ denoting a conditional probability distribution, with $(\bullet,\bullet)$ denoting a joint probability distribution.
$P_{0z}^{\delta}(\bullet)$	a probability of misclassification of the observation $Z_0 \in \Omega_l$, for training sample realization *z*, δ denotes classification method, $\delta = B, L$, B denotes BDF based method, L denotes LDF based method. With actual

	parameters values Ψ, it is Bayes error rate $P_{0z}^{B}(\Psi)$; with estimated parameters values $\hat{\Psi}$, it is AER, $P_{0z}^{B}(\hat{\Psi})$.
$P_l(\bullet)$	a marginal probability mass function for class *l*; different situations for parameter definitions are equivalent for $P(\bullet)$.
P_{lz}	a conditional probability of incorrect classification, of the observation $Z_0 \in \Omega_l$ for training sample realization *z*.
$p_l(\bullet)$	a marginal density function for class *l*; different situations for parameter definitions are equivalent for $P(\bullet)$.
R	an $N \times N$ matrix of spatial correlation between components of observations *Z*.
$r(\bullet)$	a spatial correlation function.
r_0	a vector of spatial correlations between Z_0 and *Z*.
S_N	a STL, *N* denotes the number of the training sample, an upper superscript *l* denotes class, S_N^l
T	a training sample.
t	a realisation of the training sample.
W_{lk}^{δ}	a discriminant function which classifies observation into class l or k, $l,k=\overline{1,m}; l \neq k.$ δ denotes classification method, $\delta = B, L$; B denotes BDF based method W_{lk}^{B}; L denotes LDF based method W_{lk}^{L}; for two-class case it is denoted W^{δ}.
X	an $N \times mq$ design matrix of the training sample, it is specified by $\oplus_{l=1}^{m} X_l$, where symbol $\oplus$ denotes the direct sum of matrices and X_l is the $N_l \times q$ matrix of regressors for observations from Ω_l, $l=\overline{1,m}.$
Y	an $N \times 1$ class label vector, $Y(s_i)$ denotes class label value in spatial location s_i, in short Y_i.
y	a realisation of class label vector.

Z	an $N \times 1$ random feature observation vector, $Z(s_i)$ denotes observation value in spatial location s_i, in short Z_i; if $i = 0$, it is FO $Z_0 = Z(s_0)$.
z	a realisation of feature observation vector.
β_l	a vector of regression parameters, for the data model to describe the difference for class *l*; without index *l*, β denotes the vector of all regression parameters.
Δ^2_{lk}	a marginal squared Mahalanobis distance, between Ω_l and Ω_k, $l \neq k$ at the spatial location s_0; for two-class case, it is denoted Δ^2; without an upper superscript denoted marginal Mahalanobis distance Δ_{lk} or Δ.
γ_{lk}	a log ratio of class label probabilities for Z_0, $\gamma_{lk} = \ln(\pi_0^l / \pi_0^k)$ for the multiclass case $l, k = \overline{1, m},\ l \neq k$; for two-class case it is denoted as γ, $\gamma = \ln(\pi_0^1 / \pi_0^2)$.
μ^l_{0z}	a conditional mean parameter for conditional distribution of the observation Z_0 for training sample realisation *z*, different for class *l*, without *z*; μ_0^l denotes marginal mean parameter for marginal distribution of the observation Z_0 different for class *l*.
Σ	an $N \times N$ covariance matrix for observation vector *Z*.
Ω_l	a population (class) $l = \overline{1, m}$.
Ψ	a set of all unknown model parameters.

The symbol ' denotes transposed matrix and the symbol $\wedge$ denotes estimated values of unknown parameters and expressions including parameter estimates that are used in this thesis.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Machine Learning is a field of Artificial Intelligence, an essential component of the growing field of data science that relies on the idea that systems can learn from data, identify patterns and make decisions. Through statistical approaches, algorithms are trained to make classifications or predictions, uncovering critical insights within data mining projects.

Classification is a task that requires the use of machine learning algorithms that learn to identify the category of new observations based on training data. Machine learning algorithms with completely labelled data are attributed to a supervised learning category. Two main approaches in supervised learning are generative and discriminative (Bishop and Lasserre, 2007). The supervised generative models try to capture the distribution of each class separately. In short, it models how a particular class would generate input data. When a new observation is given to these classifiers, it predicts which class would have most likely generated the given observation. Mathematically, generative models try to learn the joint probability distribution $P(Z,Y)$. Using the Bayes rule (BR), (more in subsection 1.2) and the values of $P(Z|Y)$ and $P(Y)$ for the particular class, $P(Y|Z)$, conditional probability distribution (also called posterior class distribution) of *Y* given *Z*, can be calculated. Here *Z* is called a feature vector, *Y* is called a class label vector, $P(Z|Y)$ is called the class-conditional probability distribution of *Z*, and $P(Y)$ is the probability distribution of *Y*. Some of the best-known methods are Gaussian discriminant analysis, Bayes network, Hidden Markov model (Duda R.O. et al., 2001). In general, discriminative models try to find a decision boundary between the different classes during the learning process, model the $P(Y|Z)$ directly, or learn a direct map from inputs *Z* to the *Y*. Some of the best-known methods are logistic regression, conditional random field, Support Vector Machine, Random Forest and others (Theodoritis and Kountroumbas, 2009). This thesis focuses on supervised generative models and their realization algorithms for classification problems.

The expression of $P(Z|Y)$ depends on the choice of the probability distribution that describes the observed feature vector *Z* with known class label vector *Y*. Distribution is chosen considering the properties of *Z* values: continuous or discrete, infinite or restricted interval of analysed values. For example, it can be elliptical or exponential families distribution, the families

of distributions that are described in subsection 1.3. The expression of $P(Z|Y)$ also depends on whether the *Z* is considered independent or has a statistical dependency that can be specified using covariance functions, variograms, or $P(Z|Y)$ that can be a function of *Z* depending on the spatial locations of feature *Z*. The expression of $P(Y)$ reflects the researcher's prior knowledge of class labels *Y*. It might, for instance, depend on unobservable factors or on the environment in which the feature Z is collected, for example, spatial locations of feature *Z*. Various expressions of these functions allow to expand the application of supervised generative models to solve different classification problems.

Statement of the Problem

Usually, the basic assumption of independence of observations is followed when solving classification problems using generative models (Jana and Kumar, 2016). Spatial information plays a fundamental role in analysing and understanding various fields of science tasks, for example, ecological, biological processes (natural sciences), biomedicine, engineering and social sciences. Compared with the general classification problem, spatial classification needs to consider the location information of the data and the interaction among variables. In this work, spatial contextual information notation is used to define spatial information incorporation into the classifier's structure. The spatial contextual term that is commonly used in image classification, indicating the relationship between a classified pixel and its neighbouring pixels is incorporated into analyses. A detailed comparison of the spatial contextual classification methods for remote sensing image classification is provided (Li et al., 2014). Contextual classification models that exploit spatial information by quantifying region spatial relationships can be used for image classification and object detection (Sun et al., 2016; Stabingis, 2019). In this work, spatial trend models, covariation functions, and auto-models are used to define spatial information related to the relationship between a classified spatial location and its neighbouring spatial locations. Auto-models is the term proposed by Besag (1974) that defined an expression for conditional probability models of random variables with spatial dependence (details in subsection 1.3).

Spatial contextual information incorporation into supervised generative classification algorithms construction can be used for modelling feature value distribution and/or class label distribution. Using such generative models for

classification problems in this thesis is generally called spatial contextual classification.

Moving from independent models to models with statistical spatial dependence another basic premise in classification problems is introduced that require data to follow Gaussian distribution (Wang et al., 2020, Dučinskas and Dreižienė, 2021a). This requirement is rarely implemented in real data analysis. For this reason, the research must do some data transformations. For example, Box-Cox power transformation is a transformation of non-Gaussian dependent variables into a Gaussian shape. Logit and arcsine transformations are appropriate for the data obtained from a count, and the data are expressed as decimal fractions and percentages.

The transformed data models are more difficult to interpret, and the transformations themselves do not necessarily translate features into Gaussian distribution. To solve classification problems for features with spatial contextual information with non-Gaussian distribution, new classification methods are required that would extend the application of generative classification algorithms.

Research Object

Supervised generative classification approach, Bayes discriminant function (BDF), elliptical and exponential families conditional distributions.

Research Aim and Objectives

The research aims to construct supervised generative classification algorithms based on BDF for features with spatial contextual information distributions belonging to exponential and elliptic families.

For this aim, the following objectives should be achieved:

1. Expand the use of supervised generative classifiers based on BDF for features model by Gaussian Random Field (GRF).
2. Construct a supervised generative classification algorithm based on BDF for features model by T-distributed Random Field (TRF).
3. Construct supervised generative classification algorithms based on BDF for features with spatial contextual information distributions belonging to the exponential family.

Research Methods

A literature review is performed to evaluate the methods used for modelling feature values with statistical spatial contextual information as well as to evaluate the generative models used to solve the supervised classification problems for models with spatial contextual information. The inclusion of spatial contextual information in the probability distribution of feature and class label probability is defined. The properties of the auto-models used to describe the exponential family models are used.

The applied BDF function method is based on the ratio of the logarithms of univariate conditional probability functions to the proposed supervised generative classification algorithms. Expressions of error rates are obtained for the accuracy assessment.

In empirical research, the proposed classification algorithms are investigated using simulated and real data for the estimation of the Actual Error Rate (AER). The unknown population parameters are estimated using the Bayes estimator, Maximum Likelihood (ML) and Maximum Pseudo-Likelihood (MPL) method. The critical comparison of proposed classifiers is performed for various class label functions. Based on the properties of the values of the real data attribute, auto-beta and zero-inflated auto-beta models are chosen to solve the classification problem.

Scientific Contributions and Practical Value of the Research

This thesis contributes to developing a supervised generative approach for the statistical classification of spatial data. The main contributions of this thesis can be outlined as follows:

1. Extension to classification problem of GRF observed for parameter estimation using Bayes Analysis (BA) method.
2. Classification problem solution for univariate TRF observation.
3. Classification problem solution of univariate random field observation from exponential family distribution using auto-models.
4. Classification problem solution of univariate random field observation with the Zero-Inflated Beta (ZIB) distribution.

Defensive Claims

The following claims are defended in this thesis:

1. Bayes Analysis (BA) method for parameter estimation gives the advantage over the Maximum Likelihood (ML) method for classification problem for features modeled by GRF.
2. The ML method for parameter estimation gives the advantage over Least Squares (LS) method for the classification of univariate TRF observation.
3. Spatial contextual classification based on BDFs has an advantage over the one based on linear discriminant function (LDF) for beta distribution.
4. Including the training sample elements' spatial contextual information-based neighbourhood system in the environment of classification point through the class label distribution improves the classification accuracy.

Approbation of the Results

Results obtained in this thesis were published in 5 papers: 2 papers in periodic scientific journals indexed by Web of Science, one article in a periodic scientific journal, and two papers in reviewed scientific conference proceedings. The results were presented at three international and two non-international scientific conferences. The following list presents the publications and presentations in conferences:

Papers in periodic scientific journals indexed in The Web of Science:

[A1] Dučinskas K., Dreižienė L., Zikarienė E. 2015. Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. Statistics and probability letters. p. 107-114. Available online 2014-12-16.

[A2] Dučinskas K., Zikarienė E. 2015. Actual error rates in classification of the T-distributed random field observation based on plug in linear discriminant function. Informatica. vol. 26, no. 4, p. 557-568. https://doi.org/10.15388/Informatica.2015.64.

Paper in a periodic scientific journal:

[A3] Zikarienė E., Dučinskas K. 2021. Application of spatial auto-beta models in statistical classification. Lietuvos matematikos rinkinys. Proceedings of The Lithuanian Mathematical Society, 62(A), p. 36-43. https://doi.org/10.15388/LMR.2021.25219.

Papers in peer-reviewed scientific conference proceedings:

[A4] Zikarienė E., Dučinskas K. 2019. Implementation of generalized additive models for spatial beta regression. Proceedings of the XII International Conference. Computer data analysis and modelling: stochastics and data science. p. 341-343.

[A5] Dučinskas K., Zikarienė E., Dreižienė L. 2014. Comparison of Performances of Plug-in Spatial Classification Rules Based on Bayesian and ML Estimators. Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods. p. 161-166.

Presentations in scientific conferences:

- Dučinskas K., Zikarienė E., Dreižienė L. Comparison of Performances of Plug-in Spatial Classification Rules Based on Bayesian and ML Estimators. 3rd International Conference on Pattern Recognition Applications and Methods 2014, Angers, France, March 6-8, 2014.
- Zikarienė E., Dučinskas K. Application of spatial beta regression for modelling of the algae concentration index. Spatial statistics 2019, Sitges, Spain, July 9-13, 2019.
- Zikarienė E., Dučinskas K. Implementation of generalized additive models for spatial beta regression. Computer data analysis and modelling: stochastics and data science. Minsk, Belarus, September 18-22, 2019.
- Zikarienė E., Dučinskas K. Application of spatial auto-beta models in statistical classification. Lithuanian Mathematical Society 62nd conference. June 16-17, 2021.
- Zikarienė E., Dučinskas K. Supervised classification based on spatial auto-beta models of environmental data. Computer days 2021, Klaipėda, Lithuania, September 23- 24, 2021.

Outline of the Doctoral Thesis

This doctoral thesis consists of the introduction, 3 chapters, conclusions, bibliography, and summary in lithuanian language. The introduction section provides an introduction to the research and an overview of the dissertation. The first chapter is designated for the analysis of related works on supervised generative classification methods and spatial data models. A Supervised generative classification method, based on BDF, and AER expressions are presented in the chapter. Chapter 2 presents the main results of the thesis concerned with the proposed supervised generative classification algorithm, based on BDF. Univariate conditional density (probability) expressions for spatial data models are used for constructing these functions. Chapter 3 introduces the numerical experiments and applications. General conclusions are presented after Section 3; 113 bibliographic references are included at the end of the thesis. The dissertation consists of 150 pages, 21 figures and 10 tables.

1. LITERATURE REVIEW OF SPATIAL CONTEXTUAL CLASSIFICATION, CONDITIONAL ELLIPTICAL, AND EXPONENTIAL FAMILIES DISTRIBUTIONS

In this section, a review is carried out of related works on spatial data modelling and spatial contextual classification based on BR. The theoretical base of the proposed supervised generative classification algorithms is presented based on [A1], [A2], [A3].

1.1. Spatial Data Analysis and Contextual Classification

The analysis of spatially observed data is an increasingly important statistical activity. In general, real data models in geographically approximate areas will display spatial dependence. A general analysis of the spatial process is presented by Banerjee and Gelfand (2003). These authors analysed the smoothness of the spatial process characterized by the process covariance structure in univariate and multivariate cases. Due to their mathematical tractability, Gaussian random field (GRF) models have received the majority of attention within the statistical modelling literature (Cressie, 1993; Diggle and Ribeiro 2007; Chiles and Delfiner, 2012; Baltmiškyte and Dučinskas, 2013; Gelfand and Schliep, 2016). However, according to numerous authors, not all data behave as a realisation of Gaussian distribution (see Roislien and Omre, 2006). While Gaussian processes are convenient to use and difficult to criticise, particularly as random effects specifications, they are limited because marginal distributions are always symmetric and unimodal, with fixed tail behaviour. However, Gaussian processes can be employed to create more flexible extensions while still retaining computational convenience. These extensions can be used in geostatistical modelling (Gelfand and Schliep, 2016).

A well-known insufficiency of the Gaussian distributions is their light tails. There has been intense research in using non-Gaussian distributions to tackle the deficiencies of the Gaussian distributions. The class of elliptically contoured distributions is a particularly appealing family of multivariate symmetric distributions with simple density functions and possesses properties that provide a valuable competitor to the multivariate Gaussian model. This family of multivariate distributions includes particular cases of the multivariate normal, multivariate *t*-distribution, Pearson type II and VII, multivariate symmetric Kotz type distribution, scale mixtures of normal, etc. The multivariate *t*-distribution is suitable for models describing the random phenomenon involving high probability in the tails. In this respect, *t*-

distribution is a beneficial model in economics, actuarial sciences, and many other disciplines (Sutradhar and Ali, 1986).

Another motivation for considering the multivariate *t*-distribution is its widely recognised capability to handle outliers more readily than multivariate Gaussian distribution. In observing spatial phenomena, it is sometimes found that the TDF model is advantageous whenever multiple, sparsely sampled realisations of the random field are available. The known estimation, discrimination, and simulation methods for independent observations from multivariate *t*-distributions are reviewed by Nadarajah and Kotz (2008). Thus, it is essential to include spatial dependencies in the prediction and classification problem. Kim and Mallick (2003) considered spatial prediction problems using the elliptical distribution.

Statistical analysis of non-Gaussian data is associated with exponential family and is based on Generalized Linear Models (GLMs). These models are used for data where the response variable is restricted, for example, binary or countable. GLM model is made up of a linear predictor (linear function of explanatory variables) and two functions: the link function that describes how the response variable mean depends on the linear predictor and the variance function that describes how the variance depends on the mean (Fitzmaurice et al., 2004). GLM has been proposed as an alternative model-based approach to the analysis of presence/absence or count data (Warton et al., 2015). One way to choose the most suitable model is to compare them. For example, comparison of Poisson and negative binomial models is carried out by Dučinskas et al. (2012) and Chan et al. (2021). If a dataset with a large number of zeros and skewed distribution is considered, a mixture of distributions is used for analysis, for example, the zero-inflated models class (Ospina and Ferrari, 2010). Zero-Inflated Beta (ZIB) model is proposed for metagenomics data represented as proportions analysis (Peng X. et al., 2016). Spatially correlated observations do not satisfy the independence assumption central to GLM theory (McCullagh and Nelder, 1989). However, Generalized Linear Mixed Models (GLMMs) can accommodate spatial random effects and provide a flexible means of analysing spatially correlated observations (Breslow and Clayton, 1993, Diggle et al., 1998). GLMMs have been intensively explored by numerous researchers. For example, Zhang (2002) applies the spatial GLMM for the binomial model when an unobservable spatial random process is Gaussian isotropically stationary. Very large, non-Gaussian datasets have been considered by Sengupta and Cressie (2013). The authors applied a GLMM consisting of a conditional exponential family model for the data and an underlying geostatistical process for some transformation of the mean of the data model. Dimension reduction is

achieved, which results in substantial computational speed-ups. Spatial regression models based on GLMM with the Bayesian approach used for parameter estimation are considered (Cepeda and Nunez, 2013, Lagos-Álvarez et al., 2017, Kalhory and Mahammadzadeh, 2018; Paradinas et al., 2018). When applying GLMM, an unobservable spatial Gaussian random process is modelled. Conditionally on an unobservable spatial random process, the response variable is modelled as an independent process. To obtain this independence assumption in this work, specific attention is paid to multi-parameter auto-models. Auto-models proposed by Besag (1974) are such a class of models that allow modelling spatial processes describing conditional probability distributions associated with each spatial location, that belongs to the exponential. And probability structures of the spatial process are dependent only upon contributions from cliques containing no more than two spatial locations (pairwise-only dependence). Based on the auto-models conception Kaiser and Cressie (2000) address the problem of constructing and identifying a valid joint probability density function from a set of specified conditional densities. The authors show how relations between the joint and the conditional densities using MRFs may be used in reverse order to construct a valid model from the specification of conditional densities alone.

Lee et al. (2001) extend Besag's (1974) results by relaxing the pairwise-only dependence assumption and giving a necessary form that one-parameter exponential family conditional densities must take under more general conditions of multiway dependence. Kaiser et al. (2002) introduce a class of spatial models with general multi-parameter exponential family conditional distributions. The work of these authors is extended by Hardouin and Yao (2008) with a general parameterization of the multi-parameter auto-models. The authors discuss auto-models with beta distribution and analysed spatial cooperation as well as spatial competition according to suitable choices of their parameter values.

Moving from regression models to classification tasks, it is worth noting that Switzer (1980) is one of the first to examine the classification of spatial data. This work is extended by Mardia (1984), comparing classification errors when observations are considered dependent and independent. Problems of discriminant analysis and Gaussian spatial data statistical classification has been intensively studied by numerous authors (e.g. Klein and Press, 1992, Shekhar et al., 2002, McLachlan, 2004, Nishii and Eguchi, 2006). However, none of these authors pays attention to the analytical expressions of classification errors. They are evaluated only empirically. The asymptotic expansions in terms of the inverses of the training sample sizes for cross-validation, apparent and plug-in error rates are proposed by Dučinskas (1995).

An extension of the result is expansions of expected regret risk of gamma distributions for the different parametric structures of independent observations (Dučinskas, 1998). Dučinskas (1997) examined classification risk, associated with uncorrelated observations and various training sampling schemes. Šaltytė (2001) and Šaltytė and Dučinskas (2002) derived an asymptotic expression of the expected classification error when classifying observations from a Gaussian field with a continuous spatial index into two classes with different regression mean models and general variance. These results are extended to a multivariate space-time regression model (Šaltytė-Benth and Dučinskas, 2005). A detailed empirical comparison of different classification procedures can be found in Atkinson and Lewis (2000), Berret and Calder (2016).

Conde et al. (2004) examined the classification of exponentially distinct populations with two or more classes with independent variables. Kleinman (2004) proposed a GLMM-based approach to spatial data clustering using a logistic regression approach. However, in practical applications in the study of disease outbreaks, spatial dependence has been eliminated due to computational complexity and limitations of the software used. Kauermann et al. (2010) proposed a new classification algorithm based on factor selection and logistic regression applied to GLMM. This algorithm also uses non-spatial GLMMs. Berret and Calder (2016) define spatial binary classifiers based on the probit versions of the GLM and GLMM and compare these two classifiers. Andrews et al. (2011) proposed a classification technique of independent (or uncorrelated) observations based on mixtures of multivariate *t*-distributions.

So far, the considered classification examples have assumed that no relationship exists among the various classes. In this work, this assumption is removed, and it is presumed that the different classes are closely related. The first extensions for the case where the spatial correlation between Gaussian observations, which are classified, and the observation training sample is considered non-zero, are made by Dučinskas (2009, 2009b) and Dučinskas and Dreižienė (2011, 2018) and Dreižienė et al. (2018). The ML method is used n these works to estimate the model's unknown parameters. Dreižienė (2019) proposed to assess the risk of classification when not all population parameters are known, including anisotropy parameters. Also, a connection between BDF for spatial lattice models and geostatistical models is discussed. The approach for comparison and selection of spatial linear mixed models based on a hybrid estimator of actual correct classification rates is considered. The advantage of the proposed approach against the indicator kriging method is shown (Dreižienė and Dučinskas, 2020).

An approach for classifying spatial Gaussian data based on Bayesian discriminant functions in terms of semivariograms has been developed (Dučinskas and Dreižienė, 2021a). Universal kriging case when several populations are specified by different regression parameters of GRF with second-order properties is expressed in terms of semivariograms and variances (Dučinskas and Dreižienė, 2021b).

Wakaki (1998) considered the problem of classification of observation into one of the two multivariate elliptical populations with different mean vectors and different covariance matrices. Batsidis and Zografos (2011) derived the asymptotic approximation of the distribution function for the probabilities of misclassification of elliptic random field observations. However, their approach led to the expression containing implicit function, and they did not explore approximations and estimators of the Expected Error Rate (EER). Also, Batsidis (2012) studied the behaviour of the LD by comparing the distribution function of the errors of misclassification under the truncated t and truncated normal models.

Thompson et al. (2020) developed a supervised generative approach to the classification of independent observations following matrix-variate *t*-distribution. Matrix-variate distributions that can conveniently model matrix-valued observations are considered in this work. The authors developed a methodology for parameter estimation in the matrix-variate *t*-distribution and extended it to discriminant analysis and classification using matrix-variate *t*-distribution mixtures. There is a valid assumption that observations are independent identically distributed realizations from *t*-distribution.

Contextual classification, a topic of pattern recognition, is an approach to classification based on contextual information. Contextual means this approach focuses on the relationship of nearby data points, also called a neighbourhood. Contextual classification is the term often used in image recognition. In remote sensing literature, spatial information extracted and incorporated into classification approaches has been generally called "spatial-contextual" image classification, indicating that the relationship between a "target" pixel and its neighbouring pixels is incorporated into the analysis. Li et al. (2014) divided these spatial-contextual analysis techniques into three methodological approaches, including texture extraction, image segmentation, object-based image analysis, and MRFs modelling. Contextual information incorporation from sample size was discussed by numerous authors (Song el. al., 1998, Mackowiak et al., 2021). Shekhar et al. (2002) compared the spatial autoregressive model and MRF model in terms of spatial data mining, where the MRF model for class labels defines spatial context when feature values are independent. Contextual classification method based on a

maximum posterior approach and MRF were investigated by Jackson and Landgrebe (2002), Khedama and Belhadj-Aissa (2004) and Nishii and Eguchi (2006). Stationary GRFs can model feature values in image classification, and MRF can model class labels (Stabingienė and Dučinskas, 2010). The authors analysed GRF observation classification when class label conditional distribution depends only on labels for locations from the neighbourhood. Performance of the BDF and performance of Plug-in BDF (PBDF) are compared with ones ignoring spatial correlation among feature observations (Stabingienė et al., 2010). In Dučinskas and Stabingienė (2011), class labels are modelled by RF based on 0–1 divergence, and the formula of the expected Bayes error rate is derived. The effect of training sample size and the influence of statistical parameters on the error rate are numerically evaluated, and the results are generalised by Stabingienė (2012). In the dissertation of Stabingienė (2012), the main purpose is to use BDF for GRF observation classification in image recognition. The influence of the number of nearest neighbours on the quality of image classification is studied. Stabingis et al. (2014) proposed the spatial classification rule with distance. BDF and distance-based posterior distribution for class labels are used. The study of this proposed method for analysing the influence of noise and the choice of neighbours of the classified spatial location is presented in the dissertation (Stabingis, 2019). MRFs incorporate spatial-contextual information into a classifier by modifying the discriminant function with the addition of spatial correlation terms (Fauvel et al., 2013). Wang et al. (2020) proposed a contextual classification method of data for smart machining based on the Gaussian mixture model. A new family of methods is formed by combining generative models and deep neural networks. They are called deep generative model networks (Guo et al., 2021). Also, it should be noted that the popular Generative Adversarial Network combines generative and discriminative models of machine learning.

After the literature analysis, it can be concluded that spatial contextual classification based on BDF is performed using Gaussian discriminant functions. From the analysis of related works on the topic of spatial data analysis, it can be concluded that other probability models, such as *t*-distribution or exponential family distributions, are used to model data that do not satisfy GRF properties. New classification methods for non-Gaussian spatial data would allow expanding the possibilities of classification based on BDF. The next part of this chapter presents the contextual classification rule, based on the BR for spatial data classification constructed in this work.

1.2. Spatial Contextual Classification Rule

Machine learning techniques aim to teach computers to complete specific tasks without being explicitly programmed by using the collected data. The learning is performed by using some data or observations such as examples, direct experience, or instruction. Machine learning is usually divided into four main types: supervised, unsupervised learning, semi-supervised, and reinforcement learning. Supervised learning includes having a dataset with the correct output that is used to train the system. On the other hand, unsupervised learning includes trying to find relations among the points in the dataset without having the correct results during training. It means the algorithm tries to cluster points that it believes to be highly correlated under one label based on their statistical properties only. Semi-supervised learning combines the previous two types by training the system using a dataset containing labelled and unlabelled data points. The goal is to improve the performance of the model by making use of both types of data points. Last, reinforcement learning in contrast uses trial-and-error to discover the set of actions that maximize some cumulative reward metric (Moubayed et al. 2018).
Supervised learning holds that class membership information is completely given. With more common methods, a machine learning algorithm is trained on a labelled dataset in which each record includes the outcome information. In contrast, unsupervised learning algorithms learn from a dataset without labels. Semi-supervised machine learning is a combination of these methods. It allows the algorithm to deduce patterns and identify relationships between your target variable and the rest of the dataset based on the information it already has. In semi-supervised learning, an algorithm learns from a dataset that includes labelled and unlabelled data, usually primarily unlabelled. On the other hand, unlabelled data may result from new data points which have yet to be classified.

Supervised learning is a branch of machine learning algorithms in which a function is inferred based on a labelled training sample. The training sample is formed of a group of training examples, each of which is a pair (Z,Y) where Z is an input vector and Y is the output value. The algorithm produces a function that can be used for mapping future unknown inputs. In this thesis, the focus is on supervised classification algorithms, training sample defined as $T=\{(Z_i,Y_i)\}_{i=1}^{N}$, where Z_i is called a feature value for observation i, Y_i is called class label for observation i, $Y\in\{1,...,m\}$ with m being a number of classes, i is spatial location index, $i=\overline{1,N}$, N is the number of observations.

If $m = 2$, the classification is called binary classification; if $m > 2$, it is called multiclass classification (Murphy, 2012).

Generative and discriminative models represent two different ways of solving classification tasks. The generative models look at the training sample elements and try to build a model of what one class feature values look like, then look at the other class feature values and also build a separate model of what feature values look like. Finally, to classify a new feature value, match the new feature value with build models and see which of the built models is more appropriate for the new feature. In general, the generative model learn the joint probability distribution $P(Z,Y)$ on feature values *Z* and class label *Y*. The joint probability distribution can be written as:

$$P(Z,Y) = P(Z|Y)P(Y),$$

where $P(Z|Y)$ is a class-conditional probability distribution function of feature values, $P(Y)$ is a class label distribution. If these distributions are not known, they can also be estimated from the training sample. Having these probability distributions, the posterior distribution on *Y* given *Z* can be computed. Such calculation of posterior probability is called BR (Theodoritis and Kountroumbas, 2009):

$$P(Y|Z) = \frac{P(Z|Y)P(Y)}{P(Z)}, \tag{1.1}$$

where $P(Z|Y)$ is a class-conditional probability distribution or likelihood function, $P(Y)$ is a distribution of class label, $P(Z)$ is a marginal likelihood. Given a training sample, discriminative models try to find a line or surface that is a decision boundary that separates the training sample into a different class. The models learn class probability distribution $P(Y|Z)$ directly from the space of an input (feature values) *Z* to the class labels *Y*. In the construction of Bayes classification rule, based on BR, $P(Z)$ is not taken into account because it is the same for all classes, and it does not affect the decision. In this work, a supervised generative classification algorithm is suggested as a tool that uses BDF based on BR.

In spatial analysis context, classification refers to the process of allocating spatial locations in an area of interest to one of several distinct classes or groups based on feature values data associated with each spatial location. One way of getting around the dependencies in the training samples is to select the data locations so that they are sufficiently far apart for the correlation between

their data vectors to be negligible. While this will improve the sampling efficiency, it may lead to difficulties in obtaining the ground truth information for these scattered locations. The main assumption that is held in this work is that there is spatial statistical dependence in the data. Spatial models can be specified for both the distribution of class label $P(Y)$ in the neighbourhood and the class-conditional probability distribution $P(Z|Y)$ for the feature values, given the class label.

1.2.1. Spatial Contextual Information in a Probability of Class Label and in a Data Model

Denote by s_i, a spatial location for observation *i*, and denote by $Z(s_i)$ the feature value in spatial location, in general $Z(s)$. Denote by $S_N = \{s_i \in D; i = 1,...,N\}$, the Set of Training Locations (STL) where the feature values of the training sample $Z' = [Z(s_1),...,Z(s_N)]$ are taken. It specifies the spatial sampling design or spatial framework for the training sample (Shekhar et al., 2002). Assume that S_N is partitioned into a union of *m* disjoint subsets of locations, i.e., $S_N = \bigcup_{l=1}^{m} S^{(l)}$, where $S^{(l)}$ is a subset of S_N that contains N_l locations of the feature values observations $Z(s)$ from population Ω_l, $l = 1,...,m$. It shall be assumed that the deterministic spatial sampling design and all analyses are carried out conditional on the given STL. Feature values vector Z is considered to contain spatial location information further in this work.

Data model. Consider a Random Field (RF) $\{Z(s): s \in D \subset \mathbb{R}^d\}$ is specified by multivariate distributions. The primary purpose is to solve the problem of classification of a univariate RF observation $Z_0 = Z(s_0)$, $s_0 \in D$, to one of *m* populations Ω_l, $l = 1,...,m$. Following Beret and Calder (2016), the spatial location s_0 of the observation to be classified is called the focal location. Based on that, the observation Z_0 mentioned above is called a Focal Observation (FO). To properly describe a model joint probability distribution $P(Z_0, Z)$ has to be used, but this expression is not a closed form for spatial models with the exponential family distribution (Cressie, 1993). Besag (1974)

suggested auto-models, that describe spatial models for exponential family distributions using conditional probability distribution expressions. It is known that conditional Gaussian distribution for Z_0, given Z, is multivariate Gaussian distribution. (Arnold et al., 2001, Anderson, 2003). The conditional distribution of the TRF is also a multivariate *t*-distribution, with different degrees of freedom (Ding, 2016). Because of this reason conditional distributions are selected for constructing RF models.

Let $P_l\left(Z_0 \middle| Z = z; \Psi\right)$ denote a class-conditional probability mass function for discrete distribution of Z_0 and $p_l\left(Z_0 \middle| Z = z; \Psi\right)$ a class-conditional PDF for a continuous distribution of Z_0 given $Z = z$ in the population Ω_l, $l = 1,...,m$ with a set of unknown model parameters Ψ, describing the probability distribution of the feature vector in each population. The main assumption made in these studies is that the data has a spatial dependency and the class-conditional probability density (or mass) function incorporates information from features of neighbouring locations. The latter is also called contextual information. Analysed feature values random field is specified by multivariate distribution belonging to elliptical or exponential distribution family.

Probability of class label. Probabilities of class labels implicate the prior information the researcher has about the classification class. Let π^l denote the probability realisation of the class label distribution $P(Y)$ for the population Ω_l. Class label distribution models a priori knowledge about restrictions imposed on the simultaneous labelling of connected neighboring locations. This source of information reflects statistical dependencies between the labels of neighbouring locations. MRF theory (Kinderman and Snell, 1980) provides a convenient and consistent way to model context-dependent information. Solberg et al. (1996) has exploited spatial context in classification using MRF to obtain higher accuracies over their counterparts (i.e., non-contextual classifiers). MRFs can be used to model the spatial context in $P(Y)$ related to the neighbourhood of focal location *s* and with the energy function (Sun et al., 2016). The advantage of the classification rule with distance-based posterior distribution for class label against one ignoring spatial proximity between locations is shown visually and confirmed numerically when the stationary GRF model is considered (Stabingis et al., 2014). Further description of the method is presented in the doctoral thesis

(Stabingis, 2019). The obtained results show that if nearest neighbour schemes are used for class label distribution estimation, classification accuracy is improved.

When the class is balanced, priors are equal. An approach to efficiently represent class labels is to form probabilities of the class label from data samples (Theodoritis and Kountroumbas, 2009). An alternative is to assume that probabilities depend on the location of FO and their neighbours.

In this work, the probabilities of the class labels are chosen to be estimated using the deterministic function that includes distances between FO and their neighbours. The proposed approach calculates the distance between a new instance and the nearest neighbour from each class and estimates the probabilities of the class labels using the distances. The probabilities are inversely proportional to the distance. The general formula for the probability for the population Ω_l that includes information from neighbours using inverse distance:

$$\pi_i^l = \sum_{j \in NN_i^1} \left(1/d_{ij}\right) \Big/ \sum_{j \in NN_i} \left(1/d_{ij}\right), \tag{1.2}$$

where d_{ij} is the Euclid distance between sites s_i and s_j, $i, j = 1,...,N$; $NN_i = \sum_{l=1}^{m} NN_i^l$, where NN_i^l are sites belonging to the Nearest Neighbourhood (NN) set of s_i in population Ω_l, $l = 1,...,m$. NN set can include only the closest, same fixed number, neighbours from the area with fixed distance or all training samples. NN set for the probability of class label, and class-conditional probability density (or mass) function can be the same or different.

The neighbourhood selection rule usually relates to the definition of the closeness between two samples that are used to find the nearest neighbours of the sample in question. The closeness is usually defined in terms of a distance or similarity function. For example, Euclidean distance is the most commonly used in any distance-based algorithm for numerical attributes. A spatial location s_j is defined as the neighbour of the spatial location s_i if conditional distribution of $Z(s_i)$, given all other site values, depends functionally on $Z(s_j)$, $s_i \neq s_j$. Also, define $NN_i \equiv \{s_j : s_j \text{ is a neighbour of } s_i\}$ to be the neighbourhood set of the spatial location s_i.

This probability of class label estimation function and RF models with conditional distributions together describe spatial contextual information

incorporation into the distribution of class label $P(Y)$ and the class-conditional probability distribution $P(Z|Y)$. In this work, the spatial contextual information notion is used to describe the spatial contextual information incorporation into the classificatory structure. The proposed classifier uses BDF that are formed using conditional density (probability) function ratio logarithm.

1.2.2. A Discriminant Function and an Actual Error Rate

The classification rule, given $Z = z$, is denoted as $D_z(\bullet): Z \to \{1,...,m\}$. Then the probability of misclassification of random observation Z_0, given *z*, from the population Ω_l, by prescribed classification rule is given by (see Anderson, 2003):

$$P_{0z}\left(D_z(\bullet)\right) = 1 - \sum_{l=1}^{m} \pi_0^l PC_{lz}\left(D_z(\bullet)\right), \tag{1.3}$$

where $PC_{lz}\left(D_z(\bullet)\right)$ corresponds to the conditional probability of correct observation $Z_0 \in \Omega_l$ classification, with a class-conditional PDF $p_l\left(Z_0|Z=z;\Psi\right)$ or class-conditional probability mass function $P_l\left(Z_0|Z=z;\Psi\right)$. The rule minimising the probability of misclassification is said to be the Bayes classification rule (McLachlan, 2004, Anderson, 2003) and is denoted as D_z^B and for the Z_0 it could be expressed as

$$D_z^B(Z_0) = \arg\max_{\{l=1,...,m\}} \left\{\pi_0^l p_l\left(Z_0|Z=z;\Psi\right)\right\}, \tag{1.4}$$

where Ψ is a set of unknown model parameters, $p_l\left(Z_0|Z=z;\Psi\right)$ is PDF for continuous distribution in discrete $P_l\left(Z_0|Z=z;\Psi\right)$ is class-conditional probability mass function. The basic rule minimizes the probability of misclassification by maximizing the posterior probability of class label *l*.

Discriminant functions are used in this thesis to represent classifiers. Let the pairwise Bayes discriminant function (BDF) based on log-ratios be denoted as $W_{lk}^B(Z_0,\Psi)$. In the following, the equivalent discriminant functions are used

$$W_{lk}^B(Z_0,\Psi) = \ln\left(\frac{p_l\left(Z_0|Z=z;\Psi\right)}{p_k\left(Z_0|Z=z;\Psi\right)}\right) + \gamma_{lk}, \tag{1.5}$$

where $\gamma_{lk} = \ln\left(\pi_0^l / \pi_0^k\right)$, $l,k = 1,...,m,\ k \neq l$. According to the expression (1.5) the Z_0 , given $Z = z$, is allocated to the population Ω_l if $W_{lk}^B\left(Z_0, \Psi\right) \geq 0$ for all $l,k = 1,...,m,\ k \neq l$.

Then the probability of misclassification or error rate (1.3) due to the Bayes classification rule (1.4) is called optimal or Bayes error rate (see Anderson, 2003)

$$P_{0z}^B\left(\Psi\right) = 1 - \sum_{l=1}^{m} \pi_0^l PC_{lz} \ , \tag{1.6}$$

where $PC_{lz} = P_{lz}\left(W_{lk}^B\left(Z_0, \Psi\right) \geq 0, k = \overline{1,m}, l \neq k\right)$ corresponds to the conditional probability of correct classification of the observation $Z_0 \in \Omega_l$ and P_{lz} is a probability measure with $p_l\left(Z_0 \middle| Z = z; \Psi\right)$ or $P_l\left(Z_0 \middle| Z = z; \Psi\right)$, Ψ is a set of known model parameters.

When actual values of the model parameters Ψ are known, based on the BR minimum classification error is obtained. But in practice, model parameters are usually not known; thus a set of unknown model parameters Ψ is replaced by the set of their estimates $\hat{\Psi}$. PBDF $W_{lk}^B\left(Z_0, \hat{\Psi}\right)$ is formed by plugging the estimators of the parameters into (1.5).

One of the objectives of evaluating a discriminant function is to determine its performance in the classification of future observations. When parameters Ψ are known, it is called optimal or Bayes error rate (1.6). AER is defined as the probability that a random observation form Ω_l is misallocated when the rule $W_{lk}^B\left(Z_0, \hat{\Psi}\right)$ is used. Then the actual misclassification probability or AER can be defined. The expression for AER is derived from the Bayes error rate (1.6) when BDF is incurred by PBDF:

$$P_{0z}^B\left(\hat{\Psi}\right) = 1 - \sum_{l=1}^{m} \pi_0^l P\hat{C}_{lz} \ , \tag{1.7}$$

where $P\hat{C}_{lz} = P_{lz}\left(W_{lk}^B\left(Z_0, \hat{\Psi}\right) \geq 0, k = \overline{1,m}, l \neq k\right)$. Note that this error rate is conditional on the estimated parameters, which in turn are determined by the training sample.

For the two-class case, the BDF and probability of misclassification are of the following form:

$$W^B\left(Z_0, \Psi\right) = \ln\left(\frac{p_1\left(Z_0 \middle| Z = z; \Psi\right)}{p_2\left(Z_0 \middle| Z = z; \Psi\right)}\right) + \gamma \ , \tag{1.8}$$

where $\gamma = \ln\left(\pi_0^1 / \pi_0^2\right)$, and Bayes error rate

$$P_{0z}^{B}(\Psi) = \sum_{l=1}^{2} \pi_0^l P_{lz} \,, \tag{1.9}$$

where for $l = 1,2 \quad P_{lz} = P_{lz}\left((-1)^l W^B(Z_0, \Psi) \geq 0\right)$, corresponds to the conditional probability of incorrect classification of the observation $Z_0 \in \Omega_l$. PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ is formed by plugging the estimators of the parameters into (1.8), and AER is formed by plugging the estimators of the parameters into (1.9).

The Expected Error Rate (EER) is obtained by averaging the AER with respect to the distribution of the training sample and is defined as

$$EER = E_z\left(P_{0z}^{B}\left(\hat{\Psi}\right)\right). \tag{1.10}$$

The AER is useful in providing a guide to the performance of the plug-in classification rule when it is formed from a training sample. It depends on observed values of training observations as well as on their locations. The EER is the performance measure of PBDF before a training sample is observed, and it depends on the STLs and the location of the observation to be classified s_0. It plays a similar role as the mean squared prediction error (MSPE) plays in evaluating the performance of the plug-in kriging predictor (Diggle et al., 2002). MSPE and its estimators are used for the spatial sampling design criterion for prediction (Zhu and Zhang, 2006, Zimmerman, 2006). Finding analytical expressions for EER is a complex mathematical task. For this reason, empirical EER estimates are calculated using AER expressions in this work.

General BDF and AER expressions are defined so that they can be used to construct the proposed supervised generative classification algorithm. The proposed algorithm differs from the ones already in use by forming BDF using conditional density expressions and class label probabilities. Conditional mass probabilities or density function expressions for classification points are specified by including feature values of neighbourhood spatial location. Class label distribution is specified using probabilities of class label. To estimate these probabilities, the contextual spatial information which is related to classification point spatial location and neighbourhood spatial locations, is used. Also in this work, spatial context and spatial dependence conceptions are used interchangeably. Analysed feature values random field is specified using multivariate distribution from elliptical or exponential families. In the

next part of this chapter, spatial data models are described in expressions of conditional densities and mass probability functions.

1.3. Spatial Data Models

The structure of a Bayes classifier is determined by the conditional probability functions $P\left(Z_0 \middle| Z = z; \Psi\right)$ and by the probabilities of the class label. Of the various density functions that have been investigated, none has received more attention than the multivariate Gaussian density. GRF have a dominant role in spatial statistics and especially in the traditional domain of geostatistics (Cressie, 1993, Diggle and Ribeiro 2007, Chiles and Delfiner, 2012).
Recall that $Z(s)$ is a RF with any set of locations $s_1, s_2, ..., s_N \in \mathbb{R}^d$, the joint distribution of $Z(s_1), ..., Z(s_N)$, that is equivalent to $Z_1, ..., Z_N$, is multivariate distribution belonging to families of elliptical or exponential distributions. Univariate conditional probability functions for FO Z_0 with training sample $T = t$ and different vectors of unknown population parameters Ψ are used to describe the different distributions. Spatial dependency can affect the performance of the allocation rule in two ways. Firstly, the observations in the training samples are not independent, and the Probability Density Function (PDF) incorporates the correlation between neighbouring locations. The second way is by incorporating spatial dependency into the probabilities of the class label. Modelling both class-conditional distribution $P\left(Z \middle| Y\right)$ and class label distribution $P(Y)$ becomes an essential task.

Modelling a class-conditional distribution $P\left(Z \middle| Y\right)$ is generally considered to be a spatial Gaussian distribution. Also, spatial GLMMs are used to model spatial non-Gaussian data. In these models, for any spatial location *s*, an unobservable spatial random process is modelled, which represents the random effect at spatial location *s* of unknown or unobservable causes unaccounted for by explanatory variables. This spatial random process is a Gaussian stationary process. Conditionally on this process, the feature values vector indicates an independent process (Zhang, 2002). In this work, class-conditional distributions $P\left(Z \middle| Y\right)$ are modelled using spatial Gaussian and spatial *t*-distribution and using auto-models for feature values with the exponential family distribution.

Moving away from Gaussian distribution to other models, a decision is made to choose the elliptical family *t*-distribution with useful analytical

properties: random values from TRF linear combination have the same distribution, and marginal and conditional distributions remain to obey the same principle. The multivariate *t*-distribution are suitable models to formulate and describe the random phenomenon involving high probability in the tails. In observing spatial phenomena, it is sometimes found that the TRF model is particularly useful whenever multiple, sparsely sampled realizations of the random field are available.

Moving away from symmetrical distributions to expand the scope, a decision is made to choose exponential family distributions that also describe asymmetrical data with heavy tails. Two distributions are chosen that are frequently used to model discrete values: Poisson and binomial. Also, two distributions for continuous values are chosen: gamma and beta. Expressions for classification point Z_0 from univariate conditional densities of the chosen distributions are presented further in this work.

1.3.1.Elliptical Family Distributions

The broad class of elliptical distributions gives the natural generalisation of the multivariate Gaussian distribution function. When a Gaussian process prior is combined with a Gaussian likelihood, the resulting marginal and conditional distributions have exact and straightforward expressions. But, if the underlying data is not Gaussian, the conditional distribution may be seriously misleading. A family of elliptical processes adapt to such situations. The elliptical processes subsume the Gaussian and *t*-processes (Shah et al., 2014). It is based on the elliptical distribution, a broad family of distributions that includes the Gaussian and *t*-distribution, which is attractive because it can describe fat-tailed distributions while retaining most of the Gaussian distribution's computational tractability (Fang et al., 1990). Boente et al. (2014) analysed characteristics of the elliptical distribution. It is shown that the class of elliptical distributions in the infinite-dimensional case is equivalent to the class of scale mixtures of Gaussian distributions on the space. Bånkestad et al. (2020) performed the elliptical processes analysis on a representation of elliptical distributions as a continuous mixture of Gaussian distributions.

Gaussian distribution. Gaussian distribution is a continuous probability distribution for a real-valued random variable. This distribution has two parameters: mean and variance. One attractive feature of GRFs is that the first two moments determine the complete distribution. This characteristic is one

of the most popular distributions for spatial data analysis due to its flexibility, interpretability, and probabilistic nature.

Suppose that FO Z_0 from class Ω_l, $l=\overline{1,m}$ has a conditional Gaussian distribution with mean μ_{0z}^l and variance σ_l^2. So class-conditional density function is

$$p_l\left(Z_0\left|Z=z;\Psi\right.\right)=\exp\left[-\frac{1}{2}\left(\left(Z_0-\mu_{0z}^l\right)/\sigma_z\right)^2\right]\Big/\sigma_z\sqrt{2\pi}\,, \qquad (1.11)$$

where conditional mean and conditional variance, for class Ω_l, $l=\overline{1,m}$:

$$\begin{aligned}&\mu_{0z}^l=E\left(Z_0\left|Z=z;\Psi\right.\right)=\mu_0^l+\varepsilon\,,\\&\sigma_z{}^2=\mathrm{var}\left(Z_0\left|Z=z;\Psi\right.\right)=\sigma^2\rho_0\,;\end{aligned} \qquad (1.12)$$

conditional standard deviation $\sigma_z=\sqrt{\sigma_z^2}$; $\mu_0^l=x'\left(s_0\right)\beta_l$, where $x'\left(s_0\right)$ is a $q\times 1$ vector of non-random regressors for FO Z_0, β_l is a $q\times 1$ vector of regression parameters for class Ω_l, $l=\overline{1,m}$; $\varepsilon=r_0'R^{-1}\left(z-X\beta\right)$, ε is zero-mean stationary GRF $\mathrm{cov}\left\{\varepsilon\left(s_i\right),\varepsilon\left(s_j\right)\right\}=\sigma^2 r\left(s_i-s_j\right)$, $s_i s_j\in D$, $r\left(\bullet\right)$ is a spatial correlation function, σ^2 is a marginal dispersion parameter; $\beta'=\left(\beta_1',...,\beta_m'\right)$ is a $1\times mq$ vector of regression parameters, q is a number of regressors, m is a number of classes; $\rho_0=1-r_0'R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and Z, R is a matrix of spatial correlation between components of Z. X is the $N\times mq$ design matrix for Z, it is specified by $\oplus_{l=1}^m X_l$, where symbol $\oplus$ denotes the direct sum of matrices and X_l is the $N_l\times q$ matrix of regressors for observations from Ω_l, $l=\overline{1,m}$. The model parameter set $\Psi=\left\{\beta',\sigma^2\right\}$.

Discriminant Function. Suppose that the FO Z_0 class Ω_l, $l=\overline{1,m}$ has a conditional Gaussian distribution specified in (1.11) with a conditional mean μ_{0z}^l and conditional variance σ_z^2 specified in (1.12). Consider a multiclass case $m>2$. The main goal is to solve the problem of classification of FO Z_0, given Z into several populations. The pairwise BDF specified in (1.5) in this case has the following expression:

$$W_{lk}^B\left(Z_0,\Psi\right)=\left(Z_0-\left(\mu_{0z}^l-\mu_{0z}^k\right)/2\right)\left(\mu_{0z}^l-\mu_{0z}^k\right)\Big/\sigma_z^2+\gamma_{lk}\,, \qquad (1.13)$$

where $\mu_{0z}^{l}, \mu_{0z}^{k}$ are conditional mean for populations Ω_l, Ω_k, $l \neq k,\ k,l = \overline{1,m}$, respectively; σ_z^2 is the conditional variance for class $\Omega_l, l = \overline{1,m}$; $\gamma_{lk} = \ln\left(\pi_0^l / \pi_0^k\right)$, π_0^l, π_0^k are probabilities of the class label for populations Ω_l, Ω_k $l \neq k,\ k,l = \overline{1,m}$, respectively. Classify FO Z_0 to the population Ω_l if $W_{lk}^{B}\left(Z_0, \Psi\right) \geq 0$, $l \neq k,\ l = \overline{1,m}$.

By replacing the conditional mean and variance into (1.13), the following formula for the pairwise BDF is achieved.

$$W_{lk}^{B}\left(Z_0, \Psi\right) = \left(Z_0 - \varepsilon - \mu^{lk}\right) d^{lk} \Big/ \left(\sigma\sqrt{\rho_0}\right) + \gamma_{lk}, \tag{1.14}$$

where $\mu^{lk} = x'\left(s_0\right)\left(\beta_l + \beta_k\right)/2$ and $\varepsilon = r_0' R^{-1}\left(z - X\beta\right)$, $\rho_0 = 1 - r_0' R^{-1} r_0$, $x'\left(s_0\right)$ is a $1 \times q$ vector of non-random regressors for FO Z_0; β_l, β_k are $q \times 1$ vectors of regression parameters for populations Ω_l, Ω_k $l,k = \overline{1,m},\ k \neq l$; ε is a zero-mean stationary GRF; r_0 is a vector of spatial correlations between Z_0 and Z; σ^2 is a dispersion parameter; R is a matrix of spatial correlation between components of Z; $d^{lk} = \left(\mu_{0z}^{k} - \mu_{0z}^{l}\right)/\sigma_z$ is a conditional Mahalanobis distance between Ω_l and Ω_k at spatial location s_0 conditional on Z; σ is a marginal standard deviation parameter.

Let $\varphi\left(x; \mu, \sigma^2\right)$ be the PDF of the Gaussian distribution with mean μ and variance σ^2 and set $\varphi\left(x; 0, 1\right) = \varphi\left(x\right)$.

Lemma 1. Due to the Bayes classification rule specified in (1.5), the Bayes error rate (1.6) for $m > 2$ is

$$P_{0z}^{B}\left(\Psi\right) = 1 - \sum_{l=1}^{m} \pi_0^l \int_{B_l} \varphi\left(u\right) du, \tag{1.15}$$

where $B_l = \left\{u : u \in R^1, d^{lk}u + \left(d^{lk}\right)^2 \Big/ 2 + \gamma_{lk} \geq 0;\ k = \overline{1,m},\ k \neq l\right\}$, d^{lk} is a conditional Mahalanobis distance between Ω_l and Ω_k at s_0 conditional on Z, with conditional mean and conditional variance specified in (1.12).

Lemma 2. The AER associated with the Bayes error rate specified in (1.15) has the following form for $m > 2$

$$P_{0z}^{B}\left(\hat{\Psi}\right) = 1 - \sum_{l=1}^{m} \pi_0^l \int_{B_l} \varphi\left(u\right) du,$$

where

$$B_l = \left\{ u : u \in R^1, \hat{d}^{lk} u + \frac{\left(\mu_0^l + r_0' R^{-1} X\left(\hat{\beta} - \beta\right) - \hat{\mu}^{lk}\right)\hat{d}^{lk}}{\sigma\sqrt{\rho_0}} + \frac{\gamma_{lk}\hat{\sigma}}{\sigma} \geq 0; k = \overline{1m}, l \neq k \right\},$$

$u \sim N(0,1)$, $\hat{d}^{lk}$ is an estimate of Mahalanobis distance between Ω_l and Ω_k at s_0 conditional on Z, $\mu^{lk} = x'(s_0)(\beta_l + \beta_k)/2$, $\rho_0 = 1 - r_0' R^{-1} r_0$, $\gamma_{lk} = \ln\left(\pi_0^l / \pi_0^k\right)$; r_0 is a vector of spatial correlations between Z_0 and Z; R is a matrix of spatial correlation between components of Z; X is a design matrix; $\hat{\beta}, \hat{\sigma}$ are estimates of parameters; β, σ are theoretical parameter values.

***t*-distribution.** Visually, the *t*-distribution looks much like a normal distribution; it is symmetric and bell-shaped. However, the *t*-distribution has heavier tails. Heavier tails allow for a higher dispersion of variables, meaning that it is more prone to producing values that fall far from its mean. Three parameters describe this distribution: mean, scale parameter, and degrees of freedom.

A random vector Z is multivariate *t*-distributed, denoted by $Z \sim T_N(\mu, \Pi, \nu)$, a mean vector $\mu \in R^N$, a positive definite $N \times N$ scaling matrix Π, and degrees of freedom $\nu > 0$ if its PDF is

$$p_l(Z) = \frac{\Gamma((\nu + N)/2)\left|\Pi\right|^{-\frac{1}{2}}\left[1 + (Z - \mu_l)'\Pi^{-1}(Z - \mu_l)/\nu\right]^{-\frac{N+\nu}{2}}}{(\Gamma(\nu/2)(\nu\pi)^{N/2})}, \qquad (1.16)$$

where $\Gamma(\bullet)$ is the gamma function. The expression (1.16) specifies a spherical-symmetric PDF centred at μ, with Π controlling scale and multivariate dependence while ν controls the tail behaviour (Mardia et al., 1979).

Denote by r_0 the vector of spatial correlations between Z_0 and Z. Let $R = \Sigma / \sigma^2 = \nu\Pi / (\sigma^2(\nu - 2))$ denotes a matrix of spatial correlation of Z. The conditional distribution of Z_0 given $Z = z$ in population Ω_l is $T_1(\mu_{0z}^l, \omega_{0z}, \nu + N)$ (see Roislien and Omre, 2006)

$$p_l\left(Z_0 \middle| Z = z; \Psi\right) = \frac{\Gamma((\nu + N)/2)\left[1 + (Z_0 - \mu_{0z}^l)^2 / \omega_{0z}\nu\right]^{-(\nu+N)/2}}{\sqrt{\omega_{0z}}\Gamma(\nu/2)(\nu\pi)^{N/2}}$$

with the conditional mean function, which is linear in the training sample observations:

$$\mu_{0z}^{l} = E\left(Z_0 \middle| Z = z; \Psi\right) = x_0'\beta_l + \alpha'\left(z - X\beta\right) \tag{1.17}$$

and the conditional scaling parameter

$$\begin{aligned} \omega_{0z} &= (\nu + N - 2)Var\left(Z_0 \mid Z = z; \Psi\right) / (\nu + N) = \\ &= \sigma^2 \rho_0 \zeta_z(\beta)(\nu - 2) / \nu, \end{aligned} \tag{1.18}$$

where

$$\zeta_z(\beta) = \left[1 + \left(z - X\beta\right)' \Pi^{-1}\left(z - X\beta\right)/\nu\right] \Big/ \left((1 + N)/\nu\right), \tag{1.19}$$

and $\rho_0 = 1 - r_0'\alpha$, $\alpha = R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and Z; R is a matrix of spatial correlation between components of Z; $\beta' = \left(\beta_1', ..., \beta_m'\right)$ is a $1 \times mq$ vector of regression parameters; X is the $N \times mq$ design matrix for Z; it is specified by $\oplus_{l=1}^{m} X_l$, X_l is the $N_l \times q$ matrix of regressors for observations from Ω_l, $l = \overline{1,m}$; σ^2 is a dispersion parameter; N is a training sample size. The set of model parameters is $\Psi = \left\{\beta', \sigma^2, \nu\right\}$.

Moving away from symmetrical distributions to expand the scope, a decision is made to choose exponential family distributions that also describe strictly bounded, asymmetrical data with heavy tails. GLMMs can accommodate spatial random effects and provide a flexible means of analysing spatially correlated observations. Valid assumptions are as follows: random effect at spatial location *s* is represented by an unobservable Gaussian random process and conditional on this Gaussian random process, and the property values are independent random variables Zhang (2002). Another class of models that are commonly used in spatial statistics are the exponential family auto-models. Using auto-models, independence assumption is refused, and models are formed by applying conditional specification Kaiser et al. (2002). For these reasons, they are chosen to model feature values from exponential family distributions.

1.3.2.Exponential Family Auto-Models

There is extensive literature on classification in normal populations. However, not much work has been done for the non-normal case. The exponential distributions family is another family of distributions that will be analysed. The Exponential family is a practically convenient and widely used unified family of distributions. The exponential distribution is perhaps one of the most interesting ones to be considered since it is pretty frequent to find it in practical contexts such as reliability or survival analysis.

Recall that a $Z(s)$ is a RF with any set of locations $s_1, s_2, ..., s_N \in \mathbb{R}^d$, the joint distribution of $Z(s_1), ..., Z(s_N)$ is multivariate distribution belonging to the exponential distributions family. Let $NN_i \equiv \{s_j : s_j \text{ is a neighbour of } s_i\}$ define neighbourhoods for spatial location s_i. In this work, an approach for the conditional specification of spatial statistical models applied to spatial lattices by Besag (1974) is considered. The method of deducting the joint probability structure is associated with a conditional probability method. The author examines the problems and implications of deriving the joint probability structure associated with the spatial location variables given their individual conditional distributions. The author's work is based on the theorem of Hammersley and Clifford (Cressie, 1993): a spatial location s_j, $s_j \neq s_i$ is said to be a neighbour of the spatial location s_i, $s_j \in NN_i$, if and only if the function form of $P\left(Z(s_i) \middle| Z(s_1), ..., Z(s_{i-1}), Z(s_{i+1}), ..., Z(s_N)\right)$ is dependent upon the variable $Z(s_j)$. Any set of spatial locations that either consists of a single spatial location or in which every spatial location is a neighbour of every other spatial location is called a clique. Thus, in the NN situation with pairwise dependency, there are cliques. The definition of a clique is critical to the construction of valid MRF (Besag, 1974, Cressie, 1993). The author's work also assumes: if $Z(s_1), ..., Z(s_N)$ can individually occur at sites $s_1, ..., s_N \in \mathbb{R}^d$, respectively, then they can occur together. Formally, if $P(Z(s_i)) > 0$ for each s_i, then $P(Z(s_1), ..., Z(s_N)) > 0$. This restriction is called the positivity condition.

Hammersley–Clifford theorem gives necessary and sufficient conditions under which a strictly positive probability distribution can be represented as *Z* distributed according to MRF. It states that a probability distribution with a strictly positive mass or density satisfies one of the Markov properties if and only if its density can be factorised over the cliques of the graph, where the neighbourhood structure defines the cliques. An important assumption, that is often made in MRF modelling, is the assumption of pairwise-only dependence. An MRF is said to have pairwise-only dependency (Cressie, 1993).

So the approach used by Besag (1974) in his development of auto-models begins with a specification of the conditional dependencies present among a finite set of random variables that result in a Markov random field (MRF).

These conditional dependencies define which of the entries of the multivariate random vector can be considered as neighbours of each other. Spatial auto-models are constructed under two assumptions: first, the dependence between feature values in spatial locations is pairwise. Second, the conditional probability distribution associated with each site belongs to the exponential family of distributions.

The univariate conditional probability distribution of random variable Z_0 from class $\Omega_l, l=\overline{1,m}$ is (Besag, 1974):

$$P_l\left(Z_0 \middle| Z_j;\Psi\right)=\exp\left\{A_{0h}^l\left(Z_j\right)B_h\left(Z_0\right)+C\left(Z_0\right)+D_0^l\left(Z_j\right)\right\}, \tag{1.20}$$

where $B_h(\bullet)$ and $C(\bullet)$ have specified forms, $A_{0h}^l(\bullet)$ and $D_0^l(\bullet)$ are functions of the values observed at neighbouring sites of s_0. Z_j is a feature value for neighbours of the focal location s_0. From these assumptions $A_0^l\left(Z_j\right)$

$$A_{0h}^l\left(Z_j\right)=\theta_{0h}^l+\sum_{Z_j\in NN_0}\theta_{0j}^l B_h\left(Z_j\right), \tag{1.21}$$

where $\theta_{0j}^l=\theta_{j0}^l$, and $\theta_{0j}=0,\ s_j\notin NN_0$. In general, parameters θ_{0h}^l and θ_{0j}^l differ for populations $\Omega_l, l=\overline{1,m}$. In this work, a situation, when θ_{0h}^l differ and θ_{0j} are the same, is analysed. The expressions of the functions depend on the properties of the distribution in question. A detailed analysis is provided by Arnold et al. (2001).

Moller (1999) presented simulation algorithms for several spatial one-parameter auto-models. The MRF approach is proposed in Kaiser and Cressie (2000), where the commonly used positivity condition on the joint distribution is relaxed. The authors analysed a spatial model using beta conditional distributions as an exponential family of distributions with two parameters in detail. In Kaiser et al. (2002), the authors introduce a class of spatial models with general multi-parameter exponential family conditional distributions, and raise the question of ensuring their compatibility with a joint distribution. Multi-parameter auto-models and their applications are accurately studied by Hardouin and Yao (2008). The authors analyse the competition and cooperation properties of neighbouring spatial locations. They determine the necessary form for multi-parameter exponential families in terms of the full conditionals.

Below are the conditional density functions of Poisson, binomial, gamma, and beta distributions with conditional model parameter expressions related to auto-models. These expressions are intended to construct models for spatial

feature values belonging to the exponential family. It is further used for BDF construction to solve classification problems.

Auto-Poisson model. Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space. This distribution has one parameter, which indicates the average number of events in the given time interval. A Poisson regression approach for modelling spatial autocorrelation between geographically referenced observations (Mohebbi et al., 2011).

Suppose that Z_0 for the population Ω_l, $l=\overline{1,m}$ has a conditional Poisson distribution with a conditional mean μ_{0z}^l dependent on the neighbouring site values. Hence from (1.20), the Poisson conditional probability mass function (Besag, 1974, Cressie, 1993) is

$$P_l\left(Z_0\middle|Z_j;\Psi\right)=\exp\left\{-\mu_0^l\left(Z_j\right)\right\}\mu_0^l\left(Z_j\right)^{Z_0}\Big/Z_0!, \tag{1.22}$$

where under (1.21) and with Poisson distribution sufficient statistic for FO $B_h\left(Z_0\right)=B\left(Z_0\right)=Z_0$ and conditional mean μ_{0z}^l is subject to the form:

$$\mu_{0z}^l\left(Z_j\right)=E\left(Z_0\middle|Z_j;\Psi\right)=\exp\left\{\theta_0^l+\sum_{Z_j\in NN_0}\theta_{0j}Z_j\right\}, \tag{1.23}$$

where $\Psi=\left\{\theta_0^l,\theta_{0j}\right\}$, θ_0^l is a parameter that represents large scale variation, that is different for every class *l*. θ_{0j} is a parameter that represents small scale variation, that determines spatial dependency between FO and its neighbours, with $\theta_{0j}=0,\ s_j\notin NN_0$.

Auto-binomial model. The binomial distribution is used when there are precisely two mutually exclusive outcomes of a trial. These outcomes are appropriately labelled "success" and "failure". The binomial distribution is used to obtain the probability of observing Z successes in *N* trials, with the probability of success on a single trial.

Conditional probability mass function for the auto-binomial distribution of Z_0, given *Z* in the populations Ω_l, $l=\overline{1,m}$, specification (Cressie, 1993) is

$$P_l\left(Z_0\middle|Z_j;\Psi\right)=\binom{n_{0z}}{Z_0}p_{0z}^l\left(Z_j\right)^{z_0}\left(1-p_{0z}^l\left(Z_j\right)\right)^{n_{0z}-Z_0}, \tag{1.24}$$

where n_{0z} is a number of trails, the number is fixed. Then, because (1.24) is of exponential family form, and assuming pairwise-only dependence between spatial locations from (1.20),

$$p_{0z}^{l}\left(Z_j\right)=\exp\left\{\theta_0^{l}+\sum_{Z_j\in NN_0}\theta_{0j}Z_j\right\}\Bigg/\left(1+\exp\left\{\theta_0^{l}+\sum_{Z_j\in NN_0}\theta_{0j}Z_j\right\}\right), \qquad (1.25)$$

where p_{0z}^{l} is a conditional success probability for each trail, different for every class l, θ_0^{l} is a parameter that represents large scale variation, assumed to be different for every class l. θ_{0j} is a parameter that represents small scale variation, that determines spatial dependency between FO and its neighbours, with $\theta_{0j}=0,\ s_j\notin NN_0$, $\Psi=\left\{\theta_0^{l},\theta_{0j}\right\}$. When $n_{0z}=1$ in (1.24) and (1.25), the relevant expressions are obtained for the autologistic model for binary data.

Auto-gamma model. Gamma distribution is a two-parameter family of continuous probability distributions with shape and scale parameters. Both parameters are positive real numbers.

Suppose that Z_0 from the population Ω_l, $l=\overline{1,m}$ has a conditional gamma distribution with a conditional scale parameter α_{0z}^{l} different for every class l, and a conditional shape parameter γ_{0z} dependent upon the neighbouring site values. Hence from (1.20), the auto-gamma conditional density function (Besag, 1974, Cressie, 1993) is

$$p_l\left(Z_0\middle|Z_j;\Psi\right)=\exp\left\{\begin{array}{l}-\left(\alpha_{0z}^{l}\right)^{-1}Z_0+\left(\gamma_{0z}\left(Z_j\right)-1\right)\ln Z_0+\\+\gamma_{0z}\left(Z_j\right)\ln\left(\alpha_{0z}^{l}\right)^{-1}-\ln\Gamma\left(\gamma_{0z}\left(Z_j\right)\right)\end{array}\right\} \qquad (1.26)$$

and is defined:

$$\gamma_{0z}\left(Z_j\right)=\theta_0+\sum_{Z_j\in NN_0}\theta_{0j}Z_j\ , \qquad (1.27)$$

where θ_0 is a parameter that represents large shape variation, and θ_{0j} is a parameter that represents small shape variation, that determines spatial dependency between FO and its neighbours, with $\theta_{0j}=0,\ s_j\notin NN_0$, $\Psi=\left\{\alpha_{0z}^{l},\theta_0,\theta_{0j}\right\}$.

Auto-beta model. Beta distribution is a continuous probability distribution defined on the interval (0,1). It can be parameterised by two positive shape

parameters, that control the shape of the distribution, and also by using mean and precision parameters (Ferrari and Cribari-Neto, 2004).

Defined initially on the unit interval (0,1) but easily extended to any finite interval, the beta distributions are very versatile, and they can usefully model a variety of uncertainties. Many of the finite range distributions encountered in practice can be easily transformed into the standard distribution. They can be fitted practically to any data representing a phenomenon in almost any field of application. Nadarajah and Kotz (2007) provided a review of the properties and the variations of beta distributions as well as their relationship to other distributions.

Suppose that Z_0 from the population Ω_l, $l=\overline{1,m}$ has a conditional beta distribution with a conditional mean parameter μ_{0z}^l and a conditional precision parameter ϕ_{0z}^l different for every class and dependent upon the neighbouring site values. So from (1.20), the beta conditional density function (Besag, 1974, Cressie, 1993) is

$$p_l\left(Z_0\left|Z_j;\Psi\right.\right)=\frac{Z_0^{\mu_{0z}^l\phi_{0z}^l-1}\left(1-Z_0\right)^{\left(1-\mu_{0z}^l\right)\phi_{0z}^l-1}}{Be\left(\mu_{0z}^l\phi_{0z}^l;\left(1-\mu_{0z}^l\right)\phi_{0z}^l\right)} \tag{1.28}$$

where $Be(\bullet)$ is Euler Beta function; conditional mean and conditional precision parameters expression with natural parameter is:

$$\begin{aligned}&\mu_{0z}^l=E\left(Z_0\left|Z_j;\Psi\right.\right)=\frac{1+A_{01}^l}{2+A_{01}^l+A_{02}^l},\\&\phi_{0z}^l=2+A_{01}^l+A_{02}^l.\end{aligned} \tag{1.29}$$

Beta distributions have two sufficient statistics: $B_1\left(Z_0\right)=\ln\left(1-Z_0\right)$ and $B_2\left(Z_0\right)=\ln\left(Z_0\right)$ so there are two natural parameters for different class *l*, respectively:

$$\begin{aligned}&A_{01}^l\left(Z_j\right)=\theta_{01}^l-\sum_{Z_j\in NN_0}\theta_{0j}\ln\left(1-Z_j\right),\\&A_{02}^l\left(Z_j\right)=\theta_{02}^l-\sum_{Z_j\in NN_0}\theta_{0j}\ln\left(Z_j\right),\end{aligned} \tag{1.30}$$

where θ_{0h}^l is a parameter that represents large shape variation; *h* is an index of sufficient statistics; θ_{0j} is a parameter that represents small shape variation, determines spatial dependency between FO and its neighbours, with $\theta_{0j}=0,\ s_j\notin NN_0,\ \Psi=\left\{\theta_{0h}^l,\theta_{0j}\right\}$.

Zero-inflated auto-beta model. This zero-inflation phenomenon is a particular type of overdispersion, and a zero-inflated regression model has been suggested for handling zero-inflated data. Different zero-inflated models can be constructed depending on the distribution function needs of the feature values. The applications of the Zero-inflated Negative Binomial (ZINB) and Zero-inflated Poisson (ZIP) model can be found in Ridout et al. (2001), in which these regression models are fitted to apple shoot propagation data and also provide score statistics for testing the ZIP against the ZINB. In Osei et al. (2022), authors fit ZIP and ZINB for cholera data and analyse the different models with random effects, using BA method for parameter estimation.

Inflated beta regression models are discussed in Cook et al. (2008), where the ZIB model is introduced with applications to corporate financial decisions. In Ospina and Ferrari (2012), the authors propose a general class of regression models for continuous proportions when the data contain zeros or ones. Liu and Eugenio (2018) provide an extensive review study and comparison of Bayesian and likelihood-based inferences in beta regression and ZOIB model regressions.

Many studies involve data in fractions, rates or proportions that are measured continuously in the open interval (0, 1) and contain zeros and ones. In such cases, continuous distributions are not suitable for modelling the data. Ospina and Ferrari (2010) propose mixed continuous-discrete distributions to model data observed on [0,1), (0,1] or [0,1]. These distributions capture the probability mass at 0, at 1 or both. For data observed on [0,1) or (0,1] a mixture of a continuous distribution on (0, 1) and a degenerate distribution that assigns a non-negative probability to 0 or 1 can be used. Suppose the response variable is observed on the closed interval [0, 1]. In that case, a mixture of a continuous distribution on (0, 1) and the Bernoulli distribution can be used, which yields non-negative probabilities to 0 and 1. These models are special cases of the inflated models class where the probability mass of some points exceeds what is expected by the proposed model (Tu, 2002). The focus is on cases where the dataset contains only zeros for practical reasons. Then it is natural to model the data using a mixture of two distributions: a beta distribution and a degenerate distribution in a value of 0.

Suppose that Z_0 from class Ω_l, $l=\overline{1,m}$ has a conditional ZIB distribution with conditional mean parameter μ_{0z}^{l} and conditional precision parameter ϕ_{0z}^{l} different for every class and dependent upon the neighbouring site values.

Corresponding ZIB PDF with respect to the mixture of Lebesgue and degenerative measure is given by (Osmina and Ferrari, 2010)

$$p_l\left(Z_0\middle|Z_j;\Psi\right)=c_{0z}^l I\left(Z_0=0\right)+ \\ +\left(1-c_{0z}^l\right)I\left(Z_0>0\right)\frac{Z_0^{\mu_{0z}^l\phi_{0z}^l-1}\left(1-Z_0\right)^{\left(1-\mu_{0z}^l\right)\phi_{0z}^l-1}}{Be\left(\mu_{0z}^l\phi_{0z}^l;\left(1-\mu_{0z}^l\right)\phi_{0z}^l\right)} \tag{1.31}$$

where $0<c_{0z}^l<1$ is the conditional mixture parameter, and $I(\bullet)$ is the indicator function. Thus ZIB distribution with exponential family structure is described by (1.31), where $A_0^l\left(Z_j\right)=\left(A_{01}^l\left(Z_j\right),...,A_{0h}^l\left(Z_j\right)\right)'$ and $B\left(Z_0\right)=\left(B_1\left(Z_0\right),...,B_h\left(Z_0\right)\right)'$ are dimensional vectors of natural parameters and sufficient statistics, and the following functions, depending on Z_0, feature values of neighbours and parameters:

$$C\left(Z_0\right)=\left(\log\left(1/\left(Z_0\left(1-Z_0\right)\right)\right)\right)^{\left(1-I\left(Z_0=0\right)\right)},$$

$$D_0^l\left(Z_j\right)=-\log\left\{1+\exp\left[A_{03}^l\left(Z_j\right)-Be\left(A_{01}^l\left(Z_j\right)+1;A_{02}^l\left(Z_j\right)+1\right)\right]\right\}- \\ -Be\left(A_{01}^l\left(Z_j\right)+1;A_{02}^l\left(Z_j\right)+1\right).$$

To fully describe feature values for Z_0 ZIB distribution in class *l*, which has its density function defined by (1.31), a set with such natural parameters is chosen:

$$A_{01}^l\left(Z_j\right)=\theta_{01}^l-\sum_{\substack{Z_j\in NN_0\\ Z_j\neq 0}}\theta_{0j}B_2\left(Z_j\right), \\ A_{02}^l\left(Z_j\right)=\theta_{02}^l-\sum_{\substack{Z_{j0}\in NN_0\\ Z_j\neq 0}}\theta_{0j}B_1\left(Z_j\right), \\ A_{03}^l\left(Z_j\right)=\left[\log\left(\frac{c_{0z}^l}{1-c_{0z}^l}\right)+Be\left(A_{01}^l\left(Z_j\right)+1,A_{02}^l\left(Z_j\right)+1\right)\right] \tag{1.32}$$

and corresponding sufficient statistics:

$$B_1\left(Z_0\right)=\ln\left(1-Z_0\right),\ B_2\left(Z_0\right)=\ln\left(Z_0\right),\ B_3\left(Z_0\right)=I\left(Z_0=0\right),$$

where θ_{0h}^l represents large shape variation, *h* is an index of sufficient statistics for beta distribution, that determines spatial dependency between FO and its neighbours, with $\theta_{0j}=0,\ s_j\notin NN_0,\ \Psi=\left\{c_{0z}^l,\theta_{0h}^l,\theta_{0j}\right\}$.

To properly define the model for feature values, it is necessary to choose the best-fitting distribution, and the parameters of the model have to be set. The parameters of the models are usually unknown, so their values in the model are changed by estimations. Estimates are obtained after applying some parameter estimation methods to the observation dataset. The next part of this chapter presents the parameter methods that are chosen for the evaluation of unknown parameters in empirical investigation.

1.3.3. Estimation of Unknown Parameters

The actual parameters are not usually known in practical applications, so they need to be evaluated using statistical sampling. For spatial data, the adapted classical methods, such as Maximum Likelihood (ML) and Least Squares (LS) methods, may be used to estimate unknown parameters. There are also specific methods, i.e., Maximum Pseudo-Likelihood (MPL) method (Cressie 1993), Bayesian method (Dučinskas and Šaltytė, 2003, Clark, 2005, King et al., 2010), which could be used for estimation of the unknown parameters. The methodology of these methods is presented in this part of the chapter.

Maximum Likelihood. ML estimation is a widely accepted statistical method, with well-known optimality properties in large samples. Under mild regularity conditions, the ML estimator is asymptotically normally distributed, unbiased, and fully efficient. Within the spatial context, the implementation of ML estimation is straightforward when the data is generated by a Gaussian model. But large-sample optimality properties of ML estimation hold much more generally than in the Gaussian setting. (Diggle and Ribeiro, 2007). This dissertation uses the ML method for the Gaussian and *t*-distribution spatial data models.

A likelihood function arises from a PDF considered as a function of its distributional parameterization argument. While the sample is taken as a given, the likelihood function of the parameters is often written as $L\left(\Psi \middle| Z\right)$. According to the likelihood principle, all of the information a given sample provides about Ψ is expressed in the likelihood function. In ML estimation, the values, which maximize the probability of observing the given sample, i.e.,

$$\hat{\Psi} = \arg\max_{\Psi} L\left(\Psi \middle| Z\right),$$

serve as a point estimate for the parameter of the distribution from which the sample is drawn (Diggle and Ribeiro, 2007).

When the spatial process is Gaussian, then the log-likelihood function for training sample vector Z is (Diggle and Ribeiro, 2007):

$$\log L\left(\Psi\left|Z\right.\right)=-\frac{N}{2}\log\left(2\pi\right)-\frac{1}{2}\log\left|\Sigma\right|-\frac{1}{2}\left(Z-X\beta\right)'\Sigma^{-1}\left(Z-X\beta\right),$$

maximization of which yields the ML estimates of the model parameters. When the spatial process is *t*-distribution, then the log-likelihood function for training sample vector Z is

$$\log L\left(\Psi\left|Z\right.\right)=\log\Gamma\left(\frac{\nu+N}{2}\right)-\log\Gamma\left(\frac{\nu}{2}\right)-\frac{\nu}{2}\log\left(\nu\right)-\frac{N}{2}\log\pi-$$

$$-\frac{1}{2}\log\left|\Pi\right|^{-1}-\frac{\nu+N}{2}\log\left(\nu+\left(Z-X\beta\right)\Pi^{-1}\left(Z-X\beta\right)\right).$$

To find the ML estimators, the log-likelihood function is differentiated with respect to the parameters.

Maximum Pseudo-Likelihood. Because the ML estimate is generally difficult to compute due to the normalizing factor of the probability function of spatial data models, several alternative estimates have been proposed. Besag (1975) proposed the MPL, which maximizes the direct product of conditional probabilities or conditional probability densities of the variable at each site on those at the rest of the sites

$$L_{MPL}\left(\Psi\left|Z\right.\right)\equiv\prod_{i=1}^{N}p_l\left(Z_i\left|Z_j;\Psi\right.\right),$$

namely maximizing the log pseudo-likelihood

$$\log L_{MPL}\left(\Psi\left|Z\right.\right)=\sum_{i=1}^{N}\log p_l\left(Z_i\left|Z_j;\Psi\right.\right).$$

An interpretation of MPL is that it finds the parameter $\hat{\Psi}$, such that the induced conditional distributions of feature values best match the empirical conditional distributions of feature values:

$$\hat{\Psi}=\arg\max_{\Psi}\left\{\log\left(L_{MPL}\left(\Psi\left|Z\right.\right)\right)\right\}.$$

Many researchers have provided that under suitable conditions MPL estimate is consistent and asymptotically normally distributed around the true parameter value for large samples of various spatial processes (Jensen and Moller, 1991; Mase, 1995) when examining observations from an exponential family of distributions.

Assume that $Z(s)$ is a RF with conditional beta distribution observation and $s \in D$ represents a location where the observations are taken. RF can be described by conditional density function (1.28) with conditional mean and conditional precision expressions (1.29), (1.30). Based on beta distribution properties, the shape parameters have to be positive. From this condition, it follows that $A_{0h}^{l}+1>0$, $h,l=1,2$. Also, it is necessary for all spatial locations s_i $Z(s_i)\in(0,1)$. So from the condition, $A_{0h}^{l}+1>0$, $h,l=1,2$, using the expressions, $1+\theta_{i1}^{l}-\sum_{Z_j\in NN_i}\theta_{ij}\log(1-Z_j)>0$ and $1+\theta_{i2}^{l}-\sum_{Z_j\in NN_0}\theta_{ij}\log(Z_j)>0$. If Z_j tends to 0_+ or 1_-, it necessarily follows that $\theta_{ij}\geq 0$. So pseudo-likelihood, which is formed by the sum of logarithmic conditional densities specified (1.28), i.e.:

$$\log\left(L_{PML}\left(\Psi|Z\right)\right)=\sum_{i=1}^{N_1}\log\left(p_1\left(Z_i|Z_j;\Psi\right)\right)+\sum_{i=N_1+1}^{N}\log\left(p_2\left(Z_i|Z_j;\Psi\right)\right)$$

$$\text{subject to: } 1+\theta_{i1}^{l}-\sum_{i\neq j}\theta_{ij}\log(1-Z_j)>0$$

$$1+\theta_{i2}^{l}-\sum_{i\neq j}\theta_{ij}\log(Z_j)>0$$

$$\theta_{ij}\geq 0$$

So using the optimization procedure, the estimation $\hat{\Psi}$ is obtained. Assume that $Z(s)$ is a RF with conditional ZIB distribution observation and $s\in D$ represents a location where the observations are taken. RF can be described by conditional density function (1.31) with conditional mean and conditional precision expressions (1.32). Beta distribution conditions apply to this model. The conditions related to the conditional mixture parameter are also valid, then it follows that $\lambda^{l}\leq 0$. So pseudo-likelihood, which is formed by the sum of logarithmic conditional densities specified (1.32), i.e.:

$$\log\left(L_{PML}\left(\Psi|Z\right)\right)=\sum_{i=1}^{N_1}\log\left(p_1\left(Z_i|Z_j;\Psi\right)\right)+\sum_{i=N_1+1}^{N}\log\left(p_2\left(Z_i|Z_j;\Psi\right)\right)$$

$$\text{subject to: } 1+\theta_{i1}^{l}-\sum_{i\neq j}\theta_{ij}\log\left(1-Z_j\right)>0$$

$$1+\theta_{i2}^{l}-\sum_{i\neq j}\theta_{ij}\log\left(Z_j\right)>0$$

$$\theta_{ij}\geq 0,\lambda^{l}\leq 0$$

So using the optimization procedure, the estimation $\hat{\Psi}$ is obtained.

Bayesian analysis. From a classical point of view, the parameters of the models to be evaluated are considered fixed values. From a Bayesian point of view, the model's parameters are considered random. This way, it is assumed that the parameters have a distribution that provides information (Clark, 2005; King et al., 2010). Unknown parameter values are derived from the posterior PDF, the general expression of which can be written using the BR (1.1) (Duda et al., 2001):

$$P\left(\Psi|Z\right)\propto P\left(\Psi\right)P\left(Z|\Psi\right), \tag{1.33}$$

where $P\left(\Psi|Z\right)$ is a posterior PDF of unknown parameters Ψ, $P\left(\Psi\right)$ is a prior probability of parameters Ψ $P\left(Z|\Psi\right)$ is a likelihood function (Gelman et al., 2003). Bayesian analysis yields an overall posterior distribution for all model parameters. If you are interested in one specific parameter, i.e., its marginal posterior distribution is desired, the overall posterior distribution needs to be integrated (King et al., 2010).

Many statistical applications involve many parameters that can be considered related or aggregated somehow due to the structure of the task. The overall probabilistic model for the parameters must reflect their interdependence. It is natural to model such problems hierarchically to obtain results conditionally according to specific parameters, consisting of probabilistic expressions of other parameters called hyperparameters. This hierarchical approach helps to understand multiparametric problems and is essential in developing computational strategies. (Le et al., 2006). In the hierarchical model, the likelihood of the sample for the observed data

$P(Z|\Psi)$ is defined for a given set of model parameters Ψ, assuming that there is a random variable with a prior distribution $P(\Psi|\varpi)$. Here ϖ is a hyperparameters vector that is unknown and has its own prior distribution $P(\varpi)$. The relationships between data and processes can depend on many things. These can be spatial or temporal aspects, suggesting that modelling can depend on where and when a process took place (Clark, 2005). Thus, the total prior distribution of a vector (Ψ,ϖ), according to formula (1.6), is defined as follows: $P(\Psi,\varpi)=P(\Psi|\varpi)P(\varpi)$, and the total posterior distribution according to formula (1.33) is:

$$P(\Psi,\varpi|Z)\propto P(\Psi,\varpi)P(Z|\Psi,\varpi)=P(Z|\Psi)P(\Psi,\varpi),$$

where the likelihood function of the sample $P(Z|\Psi)$ depends only on Ψ, the hyperparameters ϖ acting on the Z sample only through Ψ (Clark et al., 2006).

After analysing the parameter estimation method described above, it is chosen to use them in empirical research for the evaluation of the parameters of the analysed data model, as well as after performing a comparison using different parameter estimations in solving the classification problem.

1.4. Conclusions of the Section

This section provides an overview of the works analysed by other authors. It has been observed that the focus is on the Gaussian distribution in describing the spatial data models and also on spatial classification based on BDF using Gaussian discriminant functions (Stabingis, 2019, Wang et al., 2020, Dučinskas and Dreižienė, 2021a, 2021b). From the spatial data analysis, it can be concluded that other probability models, such as *t*-distribution, or exponential family distributions, are used to model data that do not satisfy GRF properties. New classification methods for non-Gaussian spatial data would allow expanding the possibilities of classification based on BDF.

Also, the general expression of BDF and the AER formulas are presented in this section. BDF expression related to the log ratio of conditional density or probability mass functions is proposed. Conditional probability functions analysis allows to assume that these functions can be used to create spatial data models, which provide opportunities to construct a BDF expression for solving classification problems. Based on these, BDF and AER formulas will be derived for individual data models.

Selected methods of context information incorporation into class label probabilities calculation and data model construction are provided in this section. That will make it possible to assess whether different class label calculation methods and feature value models with different amounts of constitutive information have an impact on classification accuracy.

2. DEVELOPMENT OF SUPERVISED GENERATIVE CLASSIFICATION ALGORITHM FOR FEATURE WITH SPATIAL NON-GAUSSIAN DISTRIBUTION, DISCRIMINANT FUNCTIONS AND AER EXPRESSIONS

The main focus of this work is on a supervised generative classification algorithm based on BR. This section presents the proposed supervised generative classifier also called the Bayesian classification rule, with class-conditional density (or mass) functions using BDF, which is based on log-ratios (1.5). A supervised generative classification algorithm is proposed based on this rule. Conditional density expressions are used to include contextual information for feature values. The entire rulemaking methodology is formulated for FO Z_0 with distribution from an elliptical or exponential family, using univariate conditional density (probability) functions described in subsection 1.3. For the sake of simplicity, separate cases of population selection have been examined. They can be used as basic usage cases of the proposed classification functions. There can also be generalizations for various parametric structures. Formulas for the probabilities of AER are provided as well. Some results are published in articles [A2], [A3].

2.1. BDF and Error Rates for Conditional *t*-distribution

***t*-distribution.** Suppose that the FO Z_0 class Ω_l, $l=1,2$ has a conditional *t*-distribution specified in (1.16) with a conditional mean μ_{0z}^l (1.17) and conditional scaling parameter ω_{0z} set in (1.18). Under the assumption that the populations are completely specified and for the known class label probabilities of the populations π_0^l, set $\mu_0^l = x_0'\beta_l$, $l=1,2$, x_0' is a $1\times q$ vector of non-random regressors for FO Z_0, β_l is a $q\times 1$ vector of regression parameters for class Ω_l, $l=1,2$ and assume $\pi_i^l = 0.5$ under insignificant loss of generality. For the pairwise case, the BDF specified in (1.8) in this case has the expression

$$W^B(Z_0,\Psi)=\left(Z_0-\frac{1}{2}\left(\mu_{0z}^1+\mu_{0z}^2\right)-\alpha'\left(z-X\beta\right)\right)\left(\mu_{0z}^1-\mu_{0z}^2\right), \qquad (2.1)$$

where, $\alpha = R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and *Z*, R is a matrix of spatial correlation between components of *Z*. *z* is a feature values of training sample realization, $Z=z$, X is a $N\times 2q$ design matrix for *Z*, $\beta'=\left(\beta_1',\beta_2'\right)$ is a $1\times 2q$ vector of regression parameters, q is a number of

regressors, N is a training sample size, $\Psi = \{\beta', \sigma^2, \nu\}$ is a set of model parameters, σ^2 is a dispersion parameter, ν denotes degrees of freedom.

Vector of regression parameters $\beta' = (\beta_1', \beta_2')$ is considered unknown. Other model parameters are considered fixed, so the unknown model parameters set Ψ matches the regression parameters vector. Define $S_N(\bullet)$ and $t_N(\bullet)$ as Cumulative Distribution Function (CDF) and PDF of $T_1(0,1,N)$ and let $\Delta\mu_0 = \mu_0^1 - \mu_0^2 > 0$ and $\Delta_0 = \Delta\mu_0 / (\sigma\sqrt{\rho_0})$ is conditional Mahalanobis distance between Ω_l and Ω_k at point s_0 conditional on Z (this notation is used only in this subsection), where σ is a standard deviation parameter, $\rho_0 = 1 - r_0'\alpha$, $\alpha = R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and Z, R is a matrix of spatial correlation between components of Z. Then the PBDF is obtained from the BDF by replacing the mean parameter β with $\hat{\beta}$ in the equation (2.1)

$$W^B(Z_0, \hat{\beta}) = \left(Z_0 - \alpha'\left(z - X\hat{\beta}\right) - \frac{1}{2}x_0'H\hat{\beta}\right)\left(x_0'G\hat{\beta}\right), \tag{2.2}$$

with $H = (I_q, I_q)$ and $G = (I_q, -I_q)$, where I_q denotes the identity matrix of order q; z is the training sample realization, $Z = z$; x_0' is a $1 \times q$ vector of non-random regressors for FO Z_0.

Lemma 3. The probability of misclassification (Bayes error rate) based on BDF $W^B(Z_0, \beta)$ specified in (2.1) is

$$P_{0z}^B(\beta) = S_{\nu+N}\left(-(\Delta_0/2)\sqrt{(\nu/((\nu-2)\zeta_z(\beta)))}\right), \tag{2.3}$$

where ν denotes degrees of freedom, $\zeta_z(\beta)$ specified in (1.19).

Proof. The probability of misclassification for $W^B(Z_0, \beta)$ is defined as $P_{0z}^B(\beta) = \sum_{l=1}^{2} \pi_0^l P_{lz}$, where, for $l = 1,2$, $P_{lz} = P_{lz}\left((-1)^l W^B(Z_0, \beta) > 0 \middle| \Omega_l\right)$ is the conditional probability that $W^B(Z_0, \beta)$ specified in equation (2.1) misclassifies Z_0, when it comes from Ω_l.

If a random variable has a *t*-distribution, a linear transformation of a *t*-distribution random variable also has a *t*-distribution. This property also holds

for the conditional *t*-distribution (Roislien and Omre, 2006). Based on this property and obtained $W^B\left(Z_0,\hat{\beta}\right)$ in (2.1) it follows that

$$W^B\left(Z_0,\beta\right)\left|\Omega_l\right. \sim T_1(E_l,\omega_l,\nu+N),$$

this expression means that conditionally from class Ω_l, $W^B\left(Z_0,\Psi\right)$ values are random variables with *t*-distribution, with conditional mean

$$E_l = E\left(W^B\left(Z_0,\beta\right)\left|\Omega_l\right.\right) = (-1)^{l+1}(\Delta\mu_0)^2/2,$$

conditional scaling parameter

$$\omega_l = (\nu+N-2)Var\left(W^B(Z_0,\beta)\left|\Omega_l\right.\right)/(\nu+N)$$
$$= \sigma^2(\Delta\mu_0)^2\zeta_z(\beta)\rho_0(\nu-2)/\nu.$$

where $\Delta\mu_0 = \mu_0^1 - \mu_0^2 > 0$, μ_0^l is mean for population Ω_l, $l=1,2$; ν denotes degrees of freedom; σ^2 is a dispersion parameter; *N* is a training sample size; $\rho_0 = 1 - r_0'\alpha$, $\alpha = R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and Z; *R* is a matrix of spatial correlation between components of *Z*; $\zeta_z\left(\beta\right)$ specified in (1.19).

From the properties of the multivariate *t*-distribution, it follows that in population $\Omega_l, l=1,2$:

$$(W^B\left(Z_0,\beta\right)-E_l)/\sqrt{\omega_l}\left|\Omega_l\right. \sim T_1(0,1,N+\nu),$$

where, E_l is a conditional mean, ω_l is a conditional scaling parameter, ν denotes degrees of freedom, *N* is a training sample size. Proof of the lemma is complete.

Lemma 4. The AER for the plug-in Bayes classification rule associated with PBDF specified in (2.2) has the following form:

$$P_{0z}^B\left(\hat{\beta}\right) = \sum_{l=1}^{2} S_{\nu+N}\left(\hat{Q}_l\right)/2, \tag{2.4}$$

where

$\hat{Q}_l = (-1)^l\left(\left(a_l + b\hat{\beta}\right)\mathrm{sgn}(x_0'G\hat{\beta})\Big/\sigma\sqrt{\rho_0\zeta_z\left(\beta\right)}\right)\sqrt{(\nu+N)/(\nu+N-2)}$,

$a_l = x_0'\beta_l - \alpha' X\beta$, $b = \alpha' X - x_0'H/2$, $\hat{\beta}' = \left(\hat{\beta}_1', \hat{\beta}_2'\right)$ is an estimation of $\beta' = \left(\beta_1', \beta_2'\right)$ is a $1\times 2q$ vector of regression parameters; $H = \left(I_q, I_q\right)$ and $G = \left(I_q, -I_q\right)$, I_q denotes the identity matrix of order q; x_0' is a $1\times q$ vector of non-random regressors for FO Z_0; σ^2 is a dispersion parameter;

$\rho_0 = 1 - r_0'\alpha$, $\alpha = R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and Z, R is a matrix of spatial correlation between components of Z; $\zeta_z\left(\hat{\beta}\right)$ specified in (1.19); ν denotes degrees of freedom; N is a training sample size; sgn is signum function, $\operatorname{sgn}(u) = \{-1 \text{ if } u < 0; 0 \text{ if } u = 0;\ 1 \text{ if } u > 0\}$.

Proof. AER for PBDF $W^B\left(Z_0, \hat{\beta}\right)$ is defined by $P_{0z}^B\left(\hat{\beta}\right) = \sum_{l=1}^{2} \pi_0^l \hat{P}_{lz}$ where for $l = 1,2$, $\hat{P}_{lz} = P_{lz}\left((-1)^l W^B\left(Z_0, \hat{\beta}\right) > 0 \middle| \Omega_l\right)$ is a conditional probability that $W^B\left(Z_0, \hat{\beta}\right)$ misclassifies Z_0 when it comes from Ω_l (conditional probability is based on the conditional distribution of Z_0 with conditional mean μ_{0z}^l and conditional scaling parameter ω_{0z}).

Based on the conditional *t*-distribution property and obtained $W^B\left(Z_0, \hat{\beta}\right)$ in (2.2) it follows that

$$W^B\left(Z_0, \hat{\beta}\right) \middle| \Omega_l \sim T_1\left(\hat{E}_l, \hat{\omega}_l, \nu + N\right),$$

where conditional mean and scaling parameter expressions with estimations of regressions parameters

$$\hat{E}_l = E\left(W^B\left(Z_0, \hat{\beta}\right) \middle| \Omega_l\right) = \left(a_l + b\hat{\beta}\right)\left(\Delta\hat{\mu}_0\right),$$

$$\begin{aligned}\hat{\omega}_l &= (\nu + N - 2)\operatorname{Var}\left(W^B\left(Z_0, \hat{\beta}\right) \middle| \Omega_l\right) \Big/ (\nu + N) \\ &= \sigma^2\left(\Delta\hat{\mu}_0\right)^2 \xi_z\left(\hat{\beta}\right)\rho_0(\nu - 2)\Big/\nu.\end{aligned},$$

where $a_l = x_0'\beta_l - \alpha' X\beta$, $b = \alpha' X - x_0' H/2$, $\hat{\beta}$ is an estimation of $\beta' = \left(\beta_1', \beta_2'\right)$ is a $1 \times 2q$ vector of regression parameters; $H = \left(I_q, I_q\right)$ and $G = \left(I_q, -I_q\right)$, I_q denotes the identity matrix of order q; σ^2 is a dispersion parameter; $\rho_0 = 1 - r_0'\alpha$, $\alpha = R^{-1}r_0$, r_0 is a vector of spatial correlations between Z_0 and Z, R is a matrix of spatial correlation between components of Z; $\Delta\mu_0 = \mu_0^1 - \mu_0^2 > 0$, μ_0^l is a mean for population Ω_l, $l = 1,2$; $\zeta_z\left(\hat{\beta}\right)$ specified in (1.19), ν denotes degrees of freedom, N is a training sample size.

From the properties of the multivariate *t*-distribution, it follows that for the population Ω_l, $l = 1,2$:

$$\left(W^B\left(Z_0, \hat{\beta}\right) - \hat{E}_l\right) \Big/ \sqrt{\hat{\omega}_l} \sim T_1(0, 1, N + \nu),$$

where $\hat{E}_l$ is a conditional mean, $\hat{\omega}_l$ is a conditional scaling parameter, with estimations of regressions parameters described above, ν denotes degrees of freedom, *N* is a training sample size. Proof of the lemma is complete.

When moving from symmetric distributions to non-symmetric distributions, the exponential family of distributions is selected and the expressions of the BDF and AER functions for the selected distributions are presented below.

2.2. BDF and AER for Conditional Exponential Family Distributions

Classification rules for one parameter exponential populations with restrictions that the mean of the second population is greater than the mean of the first population have been studied by Conde et al. (2005). Jana et al. (2014) proposed a class of classification rules that includes the rule submitted by Conde et al. (2005). Classification in a multi-class situation, cases of independent observations are examined by Jana et al. (2016). An example where such ordering among the scale parameters arises is considered. A limited MLE and Bayes methods are used for parameter estimation (Jana and Kumar, 2016).

Consider a 2 class case supervised classification problem. The main goal is to solve the problem of classification of FO Z_0, given the training sample *T*, into two populations. Assume that the training sample *T* distribution belongs to the family of exponential distributions.

Auto-Poisson. Suppose that the FO Z_0 population $\Omega_l, l=1,2$ has a conditional Poisson distribution specified in (1.22). The pairwise BDF $W^B\left(Z_0,\Psi\right)$ specified in (1.8) has the expression:

$$W^B\left(Z_0,\Psi\right)=\left(\theta_0^1-\theta_0^2\right)Z_0-\left(\mu_{0z}^1-\mu_{0z}^2\right)+\gamma, \tag{2.5}$$

where $\Psi=\left\{\theta_0^l,\mu_0^l\right\}$, μ_{0z}^l is a conditional mean for population Ω_l specified in (1.23), θ_0^l representing large scale variation, that is different for population $\Omega_l, l=1,2$, $\gamma=\ln\left(\pi_0^1/\pi_0^2\right)$. By replacing the estimators of the parameters into (1.8), PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ is obtained.

Lemma 5. The AER for the plug-in Bayes classification rule associated with PBDF specified in (2.5) has the following form:

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\pi_0^1\sum_{B_1}\exp\left\{Z_0\ln\hat{\mu}_{0z}^1-\hat{\mu}_{0z}^1-\ln\left(Z_0!\right)\right\}$$
$$+\pi_0^2\left(\sum_{B_2}\exp\left\{Z_0\ln\hat{\mu}_{0z}^2-\hat{\mu}_{0z}^2-\ln\left(Z_0!\right)\right\}\right),$$

where $B_1=\left\{Z_0: Z_0\in\mathbb{N}\cup\{0\}, W^B\left(Z_0,\Psi\right)<0\right\}$,

and $B_2=\left\{Z_0: Z_0\in\mathbb{N}\cup\{0\}, W^B\left(Z_0,\Psi\right)\geq 0\right\}$ are the scopes of sum, $\hat{\Psi}=\left\{\hat{\theta}_0^l,\hat{\theta}_{0j}\right\}$, $\hat{\theta}_0^l$ is an estimation of a parameter that represents large scale variation, that is different for the population Ω_l, $l=1,2$. $\hat{\theta}_{0j}$ is an estimation of the parameter representing small scale variation, with $\theta_{0j}=\theta_{j0}$ and $\theta_{0j}=0$, $j\notin NN_0$, $\hat{\mu}_0^l$ is an estimation of conditional mean defined in (1.23), π_0^l is the probability of class label for population Ω_l, $l=1,2$.

Proof. AER for PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ is defined by $P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}$ where, for $l=1,2$, $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W^B\left(Z_0,\hat{\Psi}\right)>0\middle|\Omega_l\right)$ is the conditional probability that $W^B\left(Z_0,\hat{\Psi}\right)$ misclassifies Z_0 when it comes from Ω_l (conditional probability is based on the conditional distribution of Z_0 with conditional mean μ_{0z}^l). If Z_0 is a random variable, $\left(\theta_0^1-\theta_0^2\right)Z_0-\left(\mu_0^1-\mu_0^2\right)+\gamma$ is a random variable as well, then AER

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_l\pi_0^l P_{lz}=\pi_0^1 P_{1z}\left(\left(-1\right)\left(\left(\theta_0^1-\theta_0^2\right)Z_0-\left(\mu_0^1-\mu_0^2\right)+\gamma\right)\geq 0\middle|\Omega_1\right)$$
$$+\pi_0^2 P_{2z}\left(\left(\left(\theta_0^1-\theta_0^2\right)Z_0-\left(\mu_0^1-\mu_0^2\right)+\gamma\right)>0\middle|\Omega_2\right)$$

Using properties of CDF for Poisson distribution, AER expression presented in lemma 5 is received. Proof of the lemma is complete.

Auto-binomial. Suppose that FO Z_0 population Ω_l, $l=1,2$ has a conditional binomial distribution specified in (1.24) with a fixed number of trails n_0 and success probability for each trail p_{0z}^l set in (1.25). The pairwise BDF $W^B\left(Z_0,\Psi\right)$ specified in (1.8) has the expression:

$$W^B\left(Z_0,\Psi\right)=Z_0\left(\theta_0^1-\theta_0^2\right)+n_0\left(\ln\left(1-p_{0z}^1\right)-\ln\left(1-p_{0z}^2\right)\right)+\gamma, \qquad (2.6)$$

where $\Psi = \left\{\theta_0^l, p_{0z}^l; l = 1,2\right\}$, θ_0^l representing large scale variation, that is different for population $\Omega_l, l = 1,2$, $\gamma = \ln\left(\pi_0^1/\pi_0^2\right)$. By replacing the estimators of the parameters into (1.8), PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ is obtained.

Lemma 6. The AER for the plug-in Bayes classification rule associated with PBDF specified in (2.6) has the following form:

$$P_{0t}^B\left(\hat{\Psi}\right) = \pi_0^1 \sum_{B_1} \exp\left\{\ln\left[\frac{n_0!}{Z_0!(n_0 - Z_0)!}\right] + Z_0 \ln \hat{p}_{0z}^1 + (n_0 - Z_0)\ln\left(1 - \hat{p}_{0z}^1\right)\right\}$$
$$+\pi_0^2\left(\sum_{B_2} \exp\left\{\ln\left[\frac{n_0!}{Z_0!(n_0 - Z_0)!}\right] + Z_0 \ln \hat{p}_{0z}^2 + (n_0 - Z_0)\ln\left(1 - \hat{p}_{0z}^2\right)\right\}\right)$$

where $B_1 = \left\{Z_0 : Z_0 \in \mathbb{N} \cup \{0\}, W^B\left(Z_0, \Psi\right) < 0\right\}$,

and $B_2 = \left\{Z_0 : Z_0 \in \mathbb{N} \cup \{0\}, W^B\left(Z_0, \Psi\right) \geq 0\right\}$ are the scopes of sum; $\hat{\Psi} = \left\{\hat{\theta}_0^l, \hat{\theta}_{0j}\right\}$, $\hat{\theta}_0^l$ is an estimation of a parameter that represents large scale variation, that is different for the population Ω_l, $l = 1,2$. $\hat{\theta}_{0j}$ is an estimation of the parameter that represents small scale variation, with $\theta_{0j} = \theta_{j0}$ and $\theta_{0j} = 0$, $j \notin NN_0$; $\hat{p}_{0z}^l$ is an estimation of conditional success probability for each trail, defined in (1.25); π_0^l is the probability of class label for population Ω_l, $l = 1,2$.

Proof. AER for PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ is defined by $P_{0z}^B\left(\hat{\Psi}\right) = \sum_{l=1}^{2} \pi_0^l \hat{P}_{lz}$ where, for $l = 1,2$, $\hat{P}_{lz} = P_{lz}\left((-1)^l W^B\left(Z_0, \hat{\Psi}\right) > 0 \middle| \Omega_l\right)$ is the conditional probability that $W^B\left(Z_0, \hat{\Psi}\right)$ misclassifies Z_0 when it comes from Ω_l (conditional probability is based on the conditional distribution of Z_0 with conditional success probability for each trail p_{0z}^l and fixed number of trails n_0).

If Z_0 is a random variable, then $Z_0\left(\theta_0^1 - \theta_0^2\right) + n_{0z}\left(\ln\left(1 - p_{0z}^1\right) - \ln\left(1 - p_{0z}^2\right)\right) + \gamma$ is also a random variable and AER

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l}\pi_0^l\hat{P}_{lz}$$
$$=\pi_0^1 P_{1z}\left(\left(-1\right)\left(Z_0\left(\theta_0^1-\theta_0^2\right)+n_{0z}\left(\ln\left(1-p_{0z}^1\right)-\ln\left(1-p_{0z}^2\right)\right)+\gamma\right)\geq 0\middle|\Omega_1\right)$$
$$+\pi_0^2 P_{2z}\left(\left(Z_0\left(\theta_0^1-\theta_0^2\right)+n_{0z}\left(\ln\left(1-p_{0z}^1\right)-\ln\left(1-p_{0z}^2\right)\right)+\gamma\right)>0\middle|\Omega_2\right)$$

using properties of CDF for binomial distribution AER expression presented in lemma 6 is received. Proof of the lemma is complete.

Auto-gamma. Suppose that the FO Z_0 population $\Omega_l, l=1,2$ has a conditional gamma distribution specified in (1.26) with a conditional scale parameter α_{0z}^l different for every population $\Omega_l, l=1,2$ and a conditional shape parameter γ_{0z} specified in (1.27). The pairwise BDF specified in (1.8) has the expression:

$$W^B\left(Z_0,\Psi\right)=\left(\frac{\alpha_{0z}^1-\alpha_{0z}^2}{\alpha_{0z}^1\alpha_{0z}^2}\right)Z_0+\gamma_{0z}\ln\left(\frac{\alpha_{0z}^2}{\alpha_{0z}^1}\right)+\gamma, \tag{2.7}$$

where $\Psi=\left\{\alpha_{0z}^l,\gamma_{0z}, l=1,2\right\}$, $\gamma=\ln\left(\pi_0^1/\pi_0^2\right)$. By replacing the estimators of the parameters into (1.8), the PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ is obtained.

Lemma 7. The AER for the plug-in Bayes classification rule associated with PBDF specified in (2.7) has the following form:

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}$$

where $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int\limits_{B_l}H\left(\left(-1\right)^l W\left(u,\hat{\Psi}\right)\right)p_l\left(u\right)du,$ for $l=1,2$ with $B_l=\left\{u:u\in\left(0;\infty\right)\right\}$ is a scope of integration, $H\left(\bullet\right)$ denoted Heaviside step function and probability measure P_{lz} based on conditional gamma distribution with PDF p_l specified in (1.26).

Auto-beta. A classifier based on the beta mixture models for strictly bounded and asymmetrically distributed data is proposed (Ma and Leijon, 2009).

Suppose that Z_0 class $\Omega_l, l=1,2$ has a conditional beta distribution specified in (1.28) with a mean μ_{0z}^l and precision parameter ϕ_{0z}^l, defined in (1.29) using natural parameter expression form (1.30). The pairwise BDF $W^B\left(Z_0,\Psi\right)$ specified in (1.8) has the expression:

$$W^B(Z_0,\Psi)=\left(A_{01}^1-A_{01}^2\right)\ln(Z_0)+\left(A_{02}^1-A_{02}^2\right)\ln(1-Z_0)+\gamma_0(\Psi), \qquad (2.8)$$

where A_{0h}^l are the natural parameters for different population Ω_l, $l=1,2$, defined in (1.30) with a set of parameters $\Psi=\left\{\theta_{0h}^l,\theta_{0j}\right\}$, h is a number of sufficient statistics, $\gamma_0(\Psi)=\ln\left\{\pi_0^1 B(a_{02},b_{02})/\pi_0^2 B(a_{01},b_{01})\right\}$, $B(a_{0l},b_{0l})$ is Euler Beta function with scaling parameters a_{0l} and b_{0l} form beta distribution $a_{0l}=A_{01}^l+1$, $b_{0l}=A_{02}^l+1$, $l=1,2$. By replacing the estimators of the parameters into (1.8), the PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ is obtained.

Lemma 8. The AER for the plug-in Bayes classification rule associated with PBDF specified in (2.8) for $m=2$ has the following form:

$$P_{0z}^B\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}\,, \qquad (2.9)$$

where $\hat{P}_{lz}=P_{lz}\left((-1)^l W\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int H\left((-1)^l W\left(u,\hat{\Psi}\right)\right)p_l(u)\,du,$ for $l=1,2$ with $B_l=\left\{u:u\in(0;1)\right\}$ is a scope of integration and probability measure P_{lz} based on conditional beta distribution with PDF p_l specified in (1.28).

Zero-inflated auto-beta model. Suppose that the FO Z_0 class Ω_l, $l=\overline{1,2}$ has a conditional ZIB distribution specified in (1.31) with a conditional mixture parameter c_{0z}^l and conditional mean μ_{0z}^l and precision parameter ϕ_{0z}^l from PDF of beta distribution, defined in (1.28). The pairwise BDF $W^B(Z_0,\Psi)$ specified in (1.8) has the expression:

$$\begin{aligned}W^B(Z_0,\Psi)=\ln\left(\frac{\pi_0^1 p_1\left(Z_0\middle|Z_j;\Psi\right)}{\pi_0^2 p_2\left(Z_0\middle|Z_j;\Psi\right)}\right)&=\ln\left(\frac{\pi_0^1}{\pi_0^2}\right)+\ln\left(\frac{c_{0z}^1}{c_{0z}^2}\right)I(Z_0=0)+\\&+\left(1-I(Z_0=0)\right)\ln\left(\frac{\left(1-c_{0z}^1\right)}{\left(1-c_{0z}^2\right)}\right)+\\&+\left(1-I(Z_0=0)\right)\left(A_{01}^1-A_{01}^2\right)\ln(Z_0)+\left(A_{02}^1-A_{02}^2\right)\ln(1-Z_0),\end{aligned} \qquad (2.10)$$

where $I(\bullet)$ is the indicator function. By replacing the estimators of the parameters into (1.8), the PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ is obtained.

Lemma 9. The AER for the plug-in Bayes classification rule associated with the PBDF specified in (2.10) has the following form:

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz} \tag{2.11}$$

where $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int_{B_l}H\left(\left(-1\right)^l W\left(u,\hat{\Psi}\right)\right)p_l\left(u\right)du,$ for $l=1,2$ with $B_l=\left\{u:u\in\left[0;1\right)\right\}$ is a scope of integration and probability measure P_{lz} based on conditional ZIB distribution with PDF p_l specified in (1.31).

The following subsection presents a pseudo code of the proposed supervised generative classification algorithm. This algorithm is based on BDF expressions defined in the subsection above for features with spatial contextual information distributions belonging to exponential and elliptic families.

2.3. Algorithm of Supervised Generative Classifier

In subsection 1.2. a general Bayesian classification rule and how a FO can be classified into one of *m* populations using BDF are presented. The basic idea in the supervised generative classifier is to classify FO into one of the *m* populations when the decision is made according to the BDF function given the training sample that is present. The BDF construction methodology uses a log ratio of conditional density (or mass) functions and probabilities of class labels. A general BDF expression is denoted in (1.5). They are formed by incorporating the spatial contextual information of the feature values and computing the probabilities of the class labels, where the spatial information is incorporated using a neighbourhood scheme between the elements of the training sample. That solves the problem of statistical spatial dependence in the data. The obtained BDF expressions for the *t*-distribution and exponential family's Poisson, binomial, gamma, beta distribution and ZIB distribution are presented in subsections (2.1) and (2.2). These expressions are obtained by examining the pairwise classification problem. This subsection presents a proposed supervised generative spatial contextual classification algorithm based on BDF expressions.

Let *T* be a training sample with feature values *Z*, which form RF $\left\{Z\left(s\right):s\in D\subset\mathbb{R}^d\right\}$ and class label *Y*. Suppose that FO Z_0 is a univariate RF observation that can be described by univariate distribution from elliptical or exponential families distributions. Situations with the *t*-distribution of the

elliptical family and the exponential family: Poisson, binomial, gamma and beta distributions are considered, as well as the separate case of the ZIB distribution described in more detail in subsection (1.3). The primary purpose is to solve the problem of classifying a univariate RF observation $Z_0 = Z(s_0)$, $s_0 \in D$ into one of m populations Ω_l. For simplicity, the supervised generative classification algorithm is constructed for 2 class cases. It is chosen to perform the classification decision using BDF, $W^B(Z_0, \Psi)$ (1.8), and their expressions for each of the considered discriminants are presented in subsections (2.1) and (2.2). To obtain $W^B(Z_0, \Psi)$ expressions, a set of model parameters, as well as the probability of class label values, are required. They are formed by including spatial contextual information of feature values.

First of all, the input variables are defined, namely, the data array of the training sample, with which it is necessary to define the feature values vector Z and the class labels vector Y. In addition to these two vectors forming the array, there can also be a design matrix *X* defining the needs of the explanatory variables. For example, spatial coordinates can be treated as explanatory variables. Another array is the testing sample defining FOs, which consists of a vector of feature values and also a design matrix *X* defining the needs of explanatory variables for each FO. The probability of the class label calculation method is also specified. The distribution is specified, which is selected to describe the feature values Z.

The next stage is the calculation of probabilities of class labels. The probabilities of the class label are estimated according to the selected method. One of the options is a case where the probabilities between the classes are equal. The second option is a case where they are evaluated according to the sample size. The third option is that the probabilities are evaluated by taking into account the neighbouring learning set observations, and their position in space detail in subsection (1.2). Next, knowing the distribution of the attribute values, the unknown parameters of the populations are estimated. ML, LS, BA or MPL detail in subsection (1.3) can be used to estimate unknown parameters. The estimated population parameters are used to calculate the BDF $W^B(Z_0, \Psi)$. Having BDF expressions, we move to the FO classification stage. Classification is performed at all points of the training sample; as a result, the vector *Y* of probabilities of the class label for FOs is output. The pseudo algorithm for the proposed supervised generative classifier is as follows:

	Algorithm 1. Supervised generative classifier	
(1)	**INITIALIZE** *i = 0, k = 0, m = 2, T,* π_{method} , S_N , T_{test} , T_{type} , d_0 , $p_{distribution}$	
(2)	**DECLARE**	
	l	*# a class index*
	i	*# a FO index*
	k	*# an order of NN for lattice data*
	Y	*#* $N_{Test} \times 1$ *vector of class labels for testing sample*
	Z	*#* $N_{Test} \times 1$ *vector of feature values for testing sample*
	$\hat{\Psi}$	*# a set of the estimates of the model parameters*
	$\hat{\Psi}_{t-distribution}$	*# a set of the estimates of the model parameters for t-distribution*
	$W^B(Z_i, \Psi)$	*# BDF*
	N_l	*# a number of observations in population l in the training sample*
	N	*# a number of observations in the training sample*
	N_{Test}	*# a number of observations in the testing sample*
	Ω_l	*# a population of class* $l = 1, 2.$
	NN_i	*# sites belonging to the NN of* s_i
	NN_i^l	*# sites belonging to the NN of* s_i *for population* Ω_l
	s_j	*# site belonging to the* NN_i *of* s_i , $s_i \neq s_j$
	$S_i^{(l)}$	# a subset of S_N that contains NN_i^l NN_i^l locations of Z_i *from population* Ω_l .
(3)	**READ**	
	T	*# a training sample*
	T_{type}	*# a spatial data type (lattice or geostatistical (with continuous spatial support))*
	π_{method}	*# a class label probability method (equal probabilities, sample size, inverse distance)*
	S_N	*# STL*
	T_{test}	*# a testing sample*
	d_0	*# an initial distance of* NN_i *for geostatistical data*
	$p_{distribution}$	*# a training sample distribution (t-distribution, Poisson, binomial, gamma, beta, ZIB)*
(4)	$N \leftarrow f_N(T)$	

(5) $N_{Test} \leftarrow f_N\left(T_{test}\right)$

(6) **FOR** $l \leftarrow 1$ *to* m

(7) $N_l \leftarrow f_{N_l}\left(l, T\right)$

(8) **END FOR**

(9) **FOR** $i \leftarrow 1$ *to* N_{Test}

(10) **IF** π_{method} = *"inverse distance"*

(11) **REPEAT**

(12) $k \leftarrow k+1$

(13) $d_{\max\min} \leftarrow f_d\left(T_{test}, T\right)$

(14) **IF** T_{type} = *"lattice"* **THEN**

(15) $NN_i \leftarrow f_{NN_i}\left(k, d_{\max\min} = NULL, T, T_{test}, T_{type}, Z_i\right)$

(16) **ELSE**

(17) **IF** $d_0 \geq d_{\max\min}$

(18) $NN_i \leftarrow f_{NN_i}\left(k = NULL, d_0, T, T_{test}, T_{type}, Z_i\right)$

(19) **ELSE**

(20) $NN_i \leftarrow f_{NN_i}\left(k = NULL, d_{\max\min}, T, T_{test}, T_{type}, Z_i\right)$

(21) **END IF**

(22) **END IF**

(23) $l \leftarrow 1$

(24) **FOR** $l \leftarrow 1$ *to* m

(25) $S_i^{(l)} \leftarrow f_{S_i^{(l)}}\left(S_N, NN_i, l\right)$

(26) **END FOR**

(27) **WHILE** $s_j \in NN_i,\ s_j \in S_i^{(l)},\ S_i^{(l)} = \varnothing$

(28) **END IF**

(29) $l \leftarrow 1$

(30) **FOR** $l \leftarrow 1$ *to* m

(31) **IF** π_{method} = *"equal probabilities"* **THEN**

(32) π_i^l = *1/m*

(33) **ELSE IF** π_{method} = *"sample size"* **THEN**

(34) $\pi_i^l = N_l / N$

(35) **ELSE IF** π_{method} = *"inverse distance"*

(36) $NN_i^l \leftarrow f_{NN_i^l} = \left(NN_i, T, T_{test}, Z_i\right)$

(37) $\pi_i^l \leftarrow f\left(NN_i^l, NN_i, T\right)$ *#(eq. 1.2)*

(38) **END FOR**
(39) *# Distribution of feature values is expertly evaluated*
(40) **IF** $p_{distribution}$ = *"t-distribution"* **THEN**
(41) *compute* $\hat{\Psi} = \hat{\Psi}_{t-distribution}$ *# using ML, LS in this work*
(42) *compute* $W^B\left(Z_i, \hat{\Psi}\right)$ *# (eq. 2.2)*
(43) **ELSE IF** $p_{distribution}$ = *"Poisson"* **THEN**
(44) *compute* $\hat{\Psi} = \hat{\Psi}_{Poisson}$ *# using MPL in this work*
(45) *compute* $W^B\left(Z_i, \hat{\Psi}\right)$ *# (eq. 2.5)*
(46) **ELSE IF** $p_{distribution}$ = *"binomial"* **THEN**
(47) *compute* $\hat{\Psi} = \hat{\Psi}_{binomial}$ *# using MPL in this work*
(48) *compute* $W^B\left(Z_i, \hat{\Psi}\right)$ *# (eq. 2.6)*
(49) **ELSE IF** $p_{distribution}$ = *"gamma"* **THEN**
(50) *compute* $\hat{\Psi} = \hat{\Psi}_{gamma}$ *# using MPL in this work*
(51) *compute* $W^B\left(Z_i, \hat{\Psi}\right)$ *# (eq. 2.7)*
(52) **ELSE IF** $p_{distribution}$ = *"beta"* **THEN**
(53) *compute* $\hat{\Psi} = \hat{\Psi}_{beta}$ *# using MPL in this work*
(54) *compute* $W^B\left(Z_i, \hat{\Psi}\right)$ *# (eq. 2.8)*
(55) **ELSE IF** $p_{distribution}$ = *"ZIB"* **THEN**
(56) *compute* $\hat{\Psi} = \hat{\Psi}_{ZIB}$ *# using MPL in this work*
(57) *compute* $W^B\left(Z_i, \hat{\Psi}\right)$ *# (eq. 2.10)*
(58) **END IF**
(59) **IF** $W^B\left(Z_i, \hat{\Psi}\right) \geq 0$ **THEN**
(60) Y_i = *1,* *#* $Z_i \in \Omega_1$
(61) **ELSE**
(62) Y_i =*2,* *#* $Z_i \in \Omega_2$
(63) **END IF**
(64) **END FOR**
(65) **RETURN** Y

Where $f_N(\bullet)$ calculates the dimension of the array and assigns the number of rows in the array as the result, $f_{N_l}(\bullet)$ divides the data array into *l* parts and

counts the number of rows in each part. $f_{d_i}\left(T_{test}, T, Z_i\right)$ calculates the minimum distance for each FO point to the NN from each population $\Omega_l, l=1,2$ and selects the maximum from all the minimum distances so that the NN area defined for each FO includes observations from each population, $f_{NN_i}\left(k, d_{\max\min}, T, T_{test}, T_{type}, Z_i\right)$ according to the data type T_{type}, whether the NN area is defined by distance or neighbourhood order, the NN area NN_i is created for FO, which must contain observations from each population $\Omega_l, l=1,2$, $f_{S_i^{(l)}}\left(S_N, NN_i, l\right)$ forms a set of STL for FO for the population $\Omega_l, l=1,2$. $f_{NN_i^l}=\left(NN_i, T, T_{test}, Z_i\right)$ forms NN area for FO for the population $\Omega_l, l=1,2$. In the step (39) the array of the training sample is explored. According to the properties of the distributions described in subsection (1.3), the assessment is made, and the chosen from the examined ones is the best for the description of the training sample data.

The pseudo code in Algorithm 1 describes the proposed supervised generative classification algorithm for pairwise case based on BDF. This algorithm can be used for the classification of features with spatial contextual information. This algorithm can be used when features are modelled using TRF and also when features are modelled using Poisson, binomial, gamma, beta and ZIB distributions. Algorithm 1 allows conjoining the analysed situations at the theoretical level and practically performing spatial data classification in the empirical investigation.

2.4. Conclusions of the Section

This section presents BDF expressions that allow the classification of spatial data described by non-Gaussian distributions. That is implemented by using the proposed generative supervised classification algorithm based on BDF expressions. This algorithm allows solving pairwise classification tasks for data with spatial dependence, described by *t*-distribution from distributions of the elliptic family, Poisson, binomial, gamma and beta, and ZIB distributions from the exponential family. Formulas for the probabilities of AER are also provided. They allow evaluation of the accuracy of the proposed BDF. The following section will apply the proposed algorithm with BDF $W^B\left(Z_0, \Psi\right)$ in an empirical study.

3. EMPIRICAL INVESTIGATION AND DISCUSSIONS

Performance study of the proposed supervised generative classification algorithm is conducted in this section. The section consists of four parts. The first part presents a construction part of the empirical investigation. The second part contains empirical research conducted for simulated data model by Gaussian distribution, *t*-distribution and beta distributions. In the third part, black carrageen algae coverage in the southeastern Baltic Sea research is performed, and the bottom identification problem is solved. The last part contains the description of the obtained results and conclusions. Results of empirical analysis in this thesis are published in [A1], [A2], [A3], [A4], [A5].

3.1. Construct Empirical Investigation

The research objectives are to extend supervised generative classification based on BDF for GRF observations and to investigate the proposed supervised generative classification algorithm, whose code is provided in Algorithm1.

To reach the first goal, it is chosen to extend GRF classification when the BA method is used for population parameters estimation. Also, GRF observations classification is performed in a multiclass case by analysing the situation when the population number $m = 3$. In order to obtain the second goal, a classification problem is solved for features described by *t*-distribution and beta distribution belonging to the exponential family.

Bayes error rate and AER estimates are used as measures to evaluate the performance of the discriminant function. In this work, classification error rates are empirically estimated by simulating different initial situations and performing a comparison. Empirical research should be divided into possible stages: data preparation, model parameters estimation, classification problem solving, accuracy evaluation using error rates, and analysis of the obtained results.

Empirical research is firstly performed using simulated data. STL structures are selected freely to satisfy specific properties: symmetry, asymmetry, groupness or mixedness, to cover only nearest neighbours or also include points further in the space. Then FL classification is analysed using different parametrical structures. The parametrical structures are chosen freely to satisfy specific trends: increasing distance between population means, and trendy variation of covariate function parameters. Data arrays are simulated for the chosen STL structure and population parameter sets. In the first part, black carrageen samples statistical analysis is performed by solving the

identification problem of the Baltic Sea bottom in the Lithuanian coastal waters between Palanga and Sventoji. The dependence of the number of algae with respect to the regressors is analysed.

In this section, data simulation from GRF and TRF is performed using functions in R software. Spatial data described by beta distribution is not yet implemented in R. For this goal, Algorithm 2 is suggested to generate spatially dependent variable simulation for selected parametrical structures.

	Algorithm 2. Get simulate spatial data sample with Beta distribution	
(1)	**INITIALIZE** D_{init}, N, M, $\Psi = \left\{\beta_1^l, ..., \beta_4^l, \eta_1, \eta_2\right\}$	
(2)	**DECLARE**	
	A_{lh}^l	# a natural parameter, $l, h = 1, 2$
	Z	*# an $M \times N$ array of simulated data*
	Z_{new}	*# a new simulated value*
	Z_{je}	*# eastern neighbour feature value*
	Z_{jw}	*# western neighbour feature value*
	Z_{jn}	*# northern neighbour feature value*
	Z_{js}	*# southern neighbour feature value*
	$Z_{i-1\,js}$	*# southern neighbour in the simulation i-1*
	Y_i	*# class label*
(3)	**READ**	
	D_{init}	*# initial data array with spatial training and testing sample locations also class labels for training locations. An initial dataset with point coordinates x_1^l and x_2^l.*
	N	*# a simulated sample size (training sample and testing sample size)*
	M	*# a number of simulations*
	$\Psi = \left\{\beta_1^l, ..., \beta_4^l, \eta_1, \eta_2\right\}$	*# a set of model parameters, $l = 1, 2$*
(4)	$Z \leftarrow f_{array}(M, N)$ *# construct* an *empty $M \times N$ array*	
(5)	$Y_j \leftarrow f_{Y_j}(D_{init})$ *# generates a random value: 1 and 2, with a probability 0.5*	
(6)	**FOR** $i \leftarrow 0$ TO M	
(7)	**FOR** $j \leftarrow 1$ TO N	
(8)	**IF** $i = 0$ **THEN**	

(9) $Z_{ij} \leftarrow f_{Beta}(1,1)$ *# generate random value from* $Beta(1,1)$

(10) **ELSE**

(11) $Z_{je} \leftarrow f_{nbg}\left(Z_{ij}, Z_{i-1\,je} D_{init}\right)$

(12) $Z_{jw} \leftarrow f_{nbg}\left(Z_{ij}, Z_{i-1\,jw} D_{init}\right)$

(13) $Z_{jn} \leftarrow f_{nbg}\left(Z_{ij}, Z_{i-1\,jn} D_{init}\right)$

(14) $Z_{js} \leftarrow f_{nbg}\left(Z_{ij}, Z_{i-1\,js} D_{init}\right)$

(15) **IF** $Y_j = 1$ **THEN**

(16) $A_{j1}^1 = f_{A_{j1}^1}\left(D_{init}, Z_{je}, Z_{jw}, Z_{jn}, Z_{js}, \Psi\right)$ *#(eq. 3.2)*

(17) $A_{j2}^1 = f_{A_{j2}^1}\left(D_{init}, Z_{je}, Z_{jw}, Z_{jn}, Z_{js}, \Psi\right)$ *#(eq. 3.3)*

(18) $Z_{new} \leftarrow f_{Beta}(A_{j1}^1, A_{j2}^1)$ *# generate value from* $Beta\left(A_{j1}^1 + 1, A_{j2}^1 + 1\right)$

(19) **ELSE**

(20) $A_{j1}^2 = f_{A_{j1}^2}\left(D_{init}, Z_{je}, Z_{jw}, Z_{jn}, Z_{js}, \Psi\right)$ *#(eq. 3.2)*

(21) $A_{j2}^2 = f_{A_{j2}^2}\left(D_{init}, Z_{je}, Z_{jw}, Z_{jn}, Z_{js}, \Psi\right)$ *#(eq. 3.3)*

(22) $Z_{new} \leftarrow f_{Beta}(A_{j1}^2, A_{j2}^2)$ *# generate value from* $Beta\left(A_{j1}^2 + 1, A_{j2}^2 + 1\right)$

(23) **END IF**

(24) $Z_{ij} \leftarrow Z_{new}$

(25) **END IF**

(26) **END FOR**

(27) **END FOR**

(28) **RETURN** Z

Typically, the data model parameters are unknown. When working with a simulated dataset the model parameters are assumed to be unknown and therefore need to be estimated. The available simulated data arrays are divided into training and testing samples according to the defined STL configurations. The population parameters of the training sample models are evaluated by selectively applying the LS, BA, ML and MPL methods described in the (1.3) subsection. The resulting model parameter estimates $\hat{\Psi}$ are further used in the expressions of the PBDF $W^B\left(Z_0, \hat{\Psi}\right)$, that are used to solve the classification problem.

When solving the classification problem, the probabilities of class label evaluation choices are defined: whether the probabilities are considered equal

for each class, whether they are calculated by sample, or whether they are evaluated by including spatial information: the distance between the FO and the neighbouring points of the training sample *T*. FO classification is performed by applying Algorithm 1, if the distribution of feature values are non-Gaussian. If feature values are modelled by GRF, classification is performed by applying BDF defined in (1.13).

Bayes error rate and AER estimates are used to evaluate the accuracy of the discriminant function in classifying future observations. The AER helps to guide the performance of the plug-in classification rule when it is actually formed from the training sample. It depends on observed values of training observations as well as their locations. The EER specified in (1.10) is obtained by averaging the AER with respect to the distribution of the training sample. The empirical estimation of the EER:

$$\overline{EER}^{(\delta)} = \sum_{i=1}^{M} P_{0z(i)}^{\delta}\left(\hat{\Psi}\right) \Big/ M \tag{3.1}$$

where $P_{0z(i)}^{\delta}\left(\hat{\Psi}\right)$ are the estimates of AER, which expression depends on *Z* distribution, in general, specified in (1.7), δ indicate discriminant function method (*B* denotes Bayes, *L* denotes linear), M is a simulations number. It plays a similar role to the mean squared prediction error in evaluating the performance of the plug-in kriging predictor (Diggle et al., 2002). These facts strengthen the motivation for deriving estimators of the AER and the EER associated with the PBDF. For this goal, Algorithm 3 is suggested to get AER and empirical EER estimates.

	Algorithm 3. Get AER and empirical EER	
(1)	**INITIALIZE** T, T_{test} , δ , $p_{distribution}$	
(2)	**DECLARE**	
	i	*# a FO index*
	j	*# a simulation index*
	Z_{ij}	*# feature value for* FO *i, in summation j*
	$\hat{\Psi}$	*# a set of estimates of the model parameters*
	Ψ	*# a set of model parameters*
	$W^{L}\left(Z_i, \Psi\right)$	*# modified LDF*
	$W^{B}\left(Z_i, \Psi\right)$	*# BDF*
	$\overline{EER_i}^{\delta}$	*# an empirical estimate of EER for FO i.*
	$\overline{EER}^{\delta}$	*# an empirical estimate of EER*

	$\hat{P}_{ij}^{\delta}$	*#AER*

(3)	**READ**	
	T	*# a training sample*
	δ	*# AER method (Bayes, Linear)*
	$p_{distribution}$	*# a training sample distribution (t-distribution, Poisson, binomial, gamma, beta, ZIB)*
	T_{test}	*# a testing sample*

(4) $N_{test} \leftarrow f_N(T_{test})$

(5) $M \leftarrow f_M(T_{test})$

(6) $\hat{P}_{ij}^{\delta} \leftarrow f_{array}(M, N_{test})$ *# construct* an *empty* $M \times N_{test}$ *array*

(7) *compute* π_i^l *# using Algorithm 1 from 9 to 38 code line.*

(8) *compute* $\hat{\Psi} \leftarrow f_{\Psi}(T)$

compute $W^B\left(Z_{ij}, \hat{\Psi}\right) \leftarrow f_{W^B}\left(Z_{ij}, \hat{\Psi}, T, T_{test}\right)$

 # using Algorithm 1 from 39 to 58 code line.

(9) **IF** $p_{distribution}$ *= "beta"* **AND** δ *= "Linear"* **THEN**

(10) *compute* $W^L\left(Z_{ij}, \hat{\Psi}\right) \leftarrow f_{W^L}\left(Z_{ij}, \hat{\Psi}, T, T_{test}\right)$ *# (eq. 3.4)*

(11) **END IF**

(12) **FOR** $i \leftarrow 1$ *to* N_{Test}

(13) **FOR** $j \leftarrow 1$ *to* M

(14) **IF** $p_{distribution}$ *= "t-distribution"* **THEN**

(15) $\hat{P}_{ij}^B \leftarrow f_{P_{0z}^B}\left(\pi_i^l, W^B\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$ *# (eq. 2.4)*

(16) **ELSE IF** $p_{distribution}$ *= "Poisson* **THEN**

(17) $\hat{P}_{ij}^B \leftarrow f_{P_{0z}^B}\left(\pi_i^l, W^B\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$

(18) **ELSE IF** $p_{distribution}$ *= "binomial"* **THEN**

(19) $\hat{P}_{ij}^B \leftarrow f_{P_{0z}^B}\left(\pi_i^l, W^B\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$

(20) **ELSE IF** *gamma* **THEN**

(21) $\hat{P}_{ij}^B \leftarrow f_{P_{0z}^B}\left(\pi_i^l, W^B\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$

(22) **ELSE IF** $p_{distribution}$ *= "beta"* **THEN**

(23) **IF** δ *= "Bayes"* **THEN**

(24) $\hat{P}_{ij}^B \leftarrow f_{P_{0z}^B}\left(\pi_i^l, W^B\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$ *# (eq. 2.9)*

(25) **ELSE IF** δ = *"Linear"* **THEN**

(26) $\hat{P}_{ij}^{L} \leftarrow f_{P_{0z}^{L}}\left(\pi_i^{l}, W^{L}\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$ *# (eq. 3.5)*

(27) **END IF**

(28) **ELSE IF** $p_{distribution}$ = *"ZIB"* **THEN**

(29) $\hat{P}_{ij}^{B} \leftarrow f_{P_{0z}^{B}}\left(\pi_i^{l}, W^{B}\left(Z_{ij}, \hat{\Psi}\right), \Psi, T_{test}\right)$ *# (eq. 2.11)*

(30) **END IF**

(31) **END FOR**

(32) **IF** δ = *"Bayes"* **THEN**

(33) $\overline{EER_i}^{B} \leftarrow f\left(\hat{P}_{ij}^{B}, M\right)$ *# (eq. 3.1)*

(34) **ELSE IF** δ = *"Linear"* **THEN**

(35) $\overline{EER_i}^{L} \leftarrow f\left(\hat{P}_{ij}^{L}, M\right)$ *# (eq. 3.1)*

(36) **END IF**

(37) $\overline{EER}^{\delta} \leftarrow f\left(\overline{EER_i}^{\delta}, N_{test}\right)$ *# an average for all* FO

(38) **END FOR**

(39) **RETURN** $\overline{EER}^{\delta}$

The rest of this chapter describes the empirical research to be conducted in more detail. In subsection 3.2, empirical studies for simulated data are described in more detail according to the research stages described above. First, the study with GRF observations is described. Second, the classification of symmetric *t*-distribution observations is performed. The last subsection describes the study on the classification of observations of the exponential family beta distribution.

3.2. Simulated Data

Simulated datasets are used for evaluating algorithm reliability estimation. The classification problem is created using a dataset for the Gaussian, *t*-distribution and beta distributions models. Simulated datasets are divided into training samples denoted STL and testing samples of FO with known spatial locations and feature values. During the training stage, model parameters are evaluated. During the testing stage, the classification based on the BDF (1.14) for Gaussian distribution and Algorithm 1 for *t*-distribution and beta distribution are used on the testing sample. The probability of misclassification estimators is also empirically measured. The simulations are

performed by the package for geostatistical analysis included in statistical computing software R.

Extension of Gaussian Models. At first, GRFs are analysed in this work to get more familiar with the classification procedures based on BDF in use. At this stage of the study, it is chosen to solve the classification problem for a univariate observation from GRF using BDF (1.14), when the BA method is used for estimating parameters of unknown populations. Also, it is compared to the classification of FOs using BDF (1.14) when the ML method is used to estimate the parameters of unknown populations.

Assume that $Z(s)$ is a GRF observation and $s \in D$ represents a location where the observations are taken. Let $D \subset \mathbb{R}^2$ that denotes a spatial domain of interest is a regular 2-dimensional lattice with unit spacing. A selected STL S_8 consists of eight second-order neighbours. STL is partitioned into a union of two disjoint subsets, i.e., $S_8 = S^{(1)} \cup S^{(2)}$, where $S^{(l)}$ is a subset of S_8 that includes N_l locations of the feature observations from Ω_l, $l = 1,2$. Each STL with marked labels determines the Training Labels onfiguration (TLC), denoted as $\xi(S_8)$. It is assumed that STL and TLC are fixed and TLC configurations used in this stage are shown in Figure 3.1. Two TLC configurations are chosen: symmetric $\xi_1(S_8)$ and asymmetric $\xi_2(S_8)$, to evaluate whether the TLC configuration influences error rate estimations.

FO $Z_0 = Z(s_0)$, where s_0 is the observation from GRF, and can be described by GRF class-conditional density function (1.11) with conditional mean and conditional variance expressions (1.12). Conditional mean expression depends on the (marginal or non-conditional) mean $\mu_0^l = x'(s_0)\beta_l$. for class *l*. The case where the mean is constant is chosen for the analysis. Conditional variance expression depends on marginal variance σ^2 and ρ_0 denotes a ratio between conditional and marginal variances. The case is chosen for analysis when $\sigma^2 = 1$ with the isotropic exponential covariance function specified by $c(d_{ij}) = \exp\{-d_{ij}/\phi\}$, where d_{ij} is the Euclid distance between spatial locations, $\phi = 3$ is a range parameter.

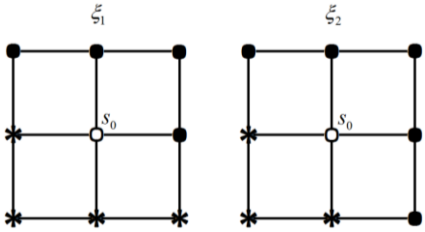


Figure 3.1 Two TCLs with points marked as dots and asterisks, marking different classes.

Two training samples for different TCL (Figure 3.1) are simulated using multivariate Gaussian distribution with the selected parameter values: with different mean μ_0^l and the same variance $\sigma^2 = 1$, and fixed range parameter $\phi = 3$ in class $l, l = 1,2$. The marginal squared Mahalanobis distance $\Delta^2 = \left(\mu_0^1 - \mu_0^2\right)^2 \Big/ \sigma^2$ is used to estimate the distance between the considered populations. Using marginal Mahalanobis, the conditional Mahalanobis $d = \Delta \Big/ \sqrt{\rho_0}$, ρ_0 denotes a ratio between conditional and marginal variances. The considered variation of the marginal Mahalanobis distance is chosen freely $\Delta = (0.5,\ 0.7,\ ...,\ 1.9)$. Also, due to the accuracy of the estimate, two cases of the simulated realizations *M* of the training sample are chosen for comparison here, $M = 10^2$ and $M = 10^4$.

Brief description of the data simulation stage: GRF observations are considered. 2 different TLC configurations are studied; for each TLC configuration different situations are described by the marginal squared Mahalanobis distance and simulated realizations *M* are selected for each population. Once the realizations of the simulated data are available, the next stage can be implemented: the estimation of the population parameters. Since the parameters of the population are unknown when studying real data, for this reason, in the expressions of the conditional mean and conditional variance described above for all studied simulated populations, it is further chosen to assume that the parameters of the mean β_l, $l = 1,2$ and the variance parameter σ^2 are considered unknown.

It is chosen to estimate population parameters using BA and ML methods. Using the ML method, parameter estimates are obtained: β and σ^2 based on *Z* are $\hat{\beta}_{ML} = \left(X' R^{-1} X\right) X R^{-1} Z$, $\sigma^2 = \left(Z - X\hat{\beta}_{ML}\right)' R^{-1} \left(Z - X\hat{\beta}_{ML}\right) \Big/ N$, where *X* is a $N \times 3q$ design matrix for *Z*, *R* is a spatial correlation matrix and *N* is an

observation number. This estimation of σ^2 is biased. The bias-adjusted ML estimator of σ^2 is $\hat{\sigma}^2 = \hat{\sigma}^2_{ML} N/(N-2q)$. Using the properties of the multivariate Gaussian distribution, it is known that $\hat{\beta}_{ML} \sim N_{2q}(\beta, \Sigma_\beta)$, and $\hat{\sigma}^2_{ML} \sim \sigma^2 \chi^2_{N-2q}/(N-2q)$, with $\Sigma_\beta = \sigma^2 (X' R^{-1} X)^{-1}$, and q is a number of regressors. The ML estimators of β and σ^2 are used in the discriminant function evaluation, i.e., $\hat{\beta} = \hat{\beta}_{ML}$, $\hat{\sigma}^2 = \hat{\sigma}^2_{ML} N/(N-2q)$ (Dučinskas, 2009).

In the BA method, the parameters are considered random, and unknown parameter values are derived from the posterior PDF (1.33). These expressions are calculated using the likelihood function and prior probability. In the case under consideration, the likelihood is given by $Z|\beta,\sigma^2 \sim N_N(\beta, \sigma^2 R)$. There are two unknown parameters, and the conjugate priors are chosen for the parameters so $p(\beta,\sigma^2) = p(\beta|\sigma^2)\, p(\sigma^2)$. Based on (Diggle, Ribeiro and Christensen, 2002) the corresponding prior distributions of the unknown parameters are chosen: $p(\beta|\sigma^2) \sim N_{2q}(\beta^{(0)}, \sigma^2 \Sigma^{(0)})$ is the Gaussian prior distribution for β conditional on σ^2 with initial parameter values $\beta^{(0)}$ and covariance matrix $\Sigma^{(0)}$, $p(\sigma^2) \sim IG(u^{(0)}, v^{(0)})$ is the prior density for σ^2 with shape and scale parameters $u^{(0)}, v^{(0)}$, where $u^{(0)}, v^{(0)} > 0$. The conjugate prior is the Gaussian inverse-gamma and is denoted as $NIG(\beta^{(0)}, \Sigma^{(0)}, u^{(0)}, v^{(0)})$, with initial parameter values, that can be chosen freely. In the case under consideration the following initial values are chosen: $\beta^{(0)} = (2; -2)$, $\Sigma^{(0)} = I_2$, $u^{(0)} = 50$ and $v^{(0)} = 49$. Combining the prior with the likelihood gives a joint normal–inverse-gamma posterior (Diggle, Ribeiro and Christensen, 2002):

$$p(\beta,\sigma^2|Z) = \frac{p(\beta,\sigma^2)\, p(Z|\beta,\sigma^2)}{p(Z)} = NIG(\beta^{(1)}, \Sigma^{(1)}, u^{(1)}, v^{(1)})$$

where $\beta^{(1)} = \left(X^T R^{-1} X + \left(\Sigma^{(0)} \right)^{-1} \right)^{-1} \left(X^T R^{-1} Z + \left(\Sigma^{(0)} \right)^{-1} \beta^{(0)} \right)$,

$\Sigma^{(1)} = \left(X^T R^{-1} X + \left(\Sigma^{(0)} \right)^{-1} \right)^{-1}$, $u^{(1)} = u^{(0)} + \frac{N}{2}$,

$v^{(1)} = v^{(0)} + \frac{1}{2} \left(\left(\beta^{(0)} \right)' \left(\Sigma^{(0)} \right)^{-1} \beta^{(0)} + Z' R^{-1} Z - \left(\beta^{(1)} \right)' \left(\Sigma^{(1)} \right)^{-1} \beta^{(1)} \right)$.

The marginal posterior for β in integrating out σ^2 is a multivariate *t*-distribution $p\left(\beta | Z\right) \sim t_p \left(\beta^{(1)}, \Sigma^* \right)$, where $\Sigma^* = \left(v^{(1)} / u^{(1)} \right) \Sigma^{(1)}$. The marginal posterior for σ^2 is $p\left(\sigma^2 \middle| Z\right) \sim IG\left(u^{(1)}, v^{(1)}\right)$, where $IG(\cdot,\cdot)$ is an inverse-gamma distribution. So the BA of β and σ^2 are $\hat{\beta}_{BA} = \beta^{(1)}$ and $\hat{\sigma}^2_{BA} = v^{(1)} \Big/ \left(u^{(1)} - 1 \right)$, respectively.

Population parameters estimates obtained using BA and ML methods in each simulation will be used to solve the supervised learning classification problem. The task is to classify Z_0 into one of two populations when $x'(s_0)\hat{\beta}_1 \neq x'(s_0)\hat{\beta}_2$ with $\hat{\sigma}^2$, the classification is performed using PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ (1.14). These expressions depend on the estimates of the population parameters and on the probabilities of the class labels. In the case under consideration, class label probabilities are estimated from the training sample. In different TLCs (Figure 3.1) for ξ_1 (symmetric) $\pi_0^l = 0.5$, $l = 1,2$ and for ξ_2 (asymmetric) $\pi_0^1 = 0.625, \pi_0^2 = 0.375$.

After classification using PBDF $W^B\left(Z_0, \hat{\Psi}\right)$, (1.14) obtain empirical estimators of AER specified in (1.9) for each simulation. An empirical estimator of the EER obtained by averaging AER over-runs is proposed as a measure for comparison. The values of $\overline{EER}$ for ξ_1 and for ξ_2 and different marginal Mahalanobis distances for the simulated realizations *M* are presented in Table 3.1. In all of the analysed situations with two different simulated realizations of the training sample 10^2 and 10^4, the error rates are achieved smaller with bigger training samples. These results are further analysed with an expectation that they are more accurate error estimates. Figure 3.2 represents the obtained $\overline{EER}$ values with ML and BA methods for different configurations (10^4 simulations).

Table 3.1 Values of $\overline{EER}$ for the different estimators and the TLC.

TLC	ξ_1				ξ_2			
Method	ML	BA	ML	BA	ML	BA	ML	BA
$\Delta \setminus M$	10^2		10^4		10^2		10^4	
0.5	0.352	**0.335**	0.365	0.349	0.359	0.340	0.328	**0.315**
0.7	0.275	0.262	0.277	**0.256**	0.310	0.280	0.257	**0.239**
0.9	0.195	**0.177**	0.201	0.180	0.196	0.182	0.189	**0.171**
1.1	0.143	0.123	0.140	**0.122**	0.143	0.119	0.136	**0.118**
1.3	0.094	**0.083**	0.098	**0.083**	0.097	0.085	0.097	**0.081**
1.5	0.067	**0.053**	0.067	0.055	0.061	**0.051**	0.066	0.053
1.7	0.049	**0.035**	0.047	0.036	0.044	**0.034**	0.047	0.035
1.9	0.034	**0.024**	0.033	**0.024**	0.033	0.024	0.032	**0.023**

The graph shows that the $\overline{EER}$ estimates decrease as the distance between populations increases. Comparing the obtained estimates with respect to TLC configurations, the estimates obtained are lower in the case of the asymmetrical configuration, most clearly observed in the analysed situations where the distance between populations varies from 0.5 to 1.1.

Analysing the estimates of simulated realisations of the training sample in Table 3.1, it can be seen that for all $\Delta = (0.5, 0.7, ..., 1.9)$ values $\overline{EER}^{ML} \geq \overline{EER}^{BA}$. So it can be concluded that the BA case has an advantage over the ML case by the EER minimum criterion. Comparing the obtained estimates in Figure 3.2 according to parameter estimation methods, the advantage of the BA method over the advantage of the ML method is also visible.

The quantitative comparison of the two cases of the parameter estimators is also made using the index $\kappa_{\xi_k} = \overline{EER}^{ML} / \overline{EER}^{BA}$ values. Index values for $M = 10^4$ case and ξ_k, $k = 1,2$ are shown in Figure 3.3. For both TLC cases κ_{ξ_1} and κ_{ξ_2} increases when Δ increases. Two growth curves with identical behaviour are achieved.

Another interesting aspect of the study is the inclusion of information about class label probabilities. Two situations are compared: ξ_1 notes a situation where the probabilities of a class label are equal and ξ_2 notes a situation where the probabilities of the class label are unequal.

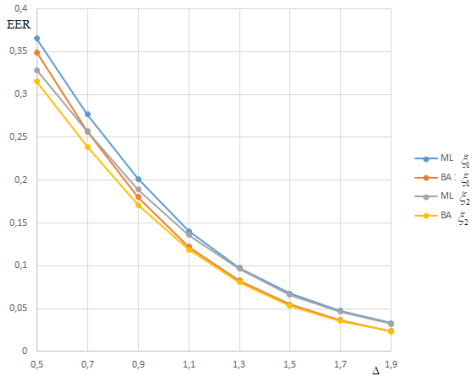


Figure 3. 2 Values of $\overline{EER}$ for different TLC and parameter estimates $M = 10^4$.

Below is a graph of the comparison in Figure 3.4, comparing the precision of computation between ML and BA estimates. The percentage precision is calculated according to the formula: $\tau^{\delta} = \left| \left(\overline{EER}^{\delta}(\xi_2) - \overline{EER}^{\delta}(\xi_1) \right) \Big/ \overline{EER}^{\delta}(\xi_2) \right| \cdot 100\%$, where δ denotes ML or BA estimates.

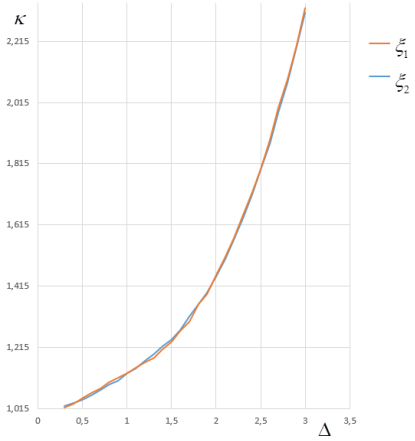


Figure 3.3 Values of κ for different TLC $M = 10^4$.

When the classes are close to each other: Δ varies from 0.3 to 0.9, and precision values vary from 13.97% to 6.45%. Estimates of calculated classification errors decrease in size by τ^{δ}. As the parameter Δ values increase from 1 to 3, the classes become more distant from each other, in which case the precision parameter estimates vary from 0.88% to 6.16%. In this case, the curves representing the different parameter estimates behave

very similarly. Independent of the method of parameter estimation, it can be concluded from the obtained results that the inclusion of prior information in the models has estimates of interference classification errors.

The comparison of the two approaches to parameter estimation is made based on the values of the EER incurred by the classification rule based on the PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ (1.14). The proposed optimality criterion is based on the derived formula of the AER. The simulation experiment shows the advantage of the BA estimation approach against the frequentist approach (ML). This advantage is more significant for strongly separated populations (larger values of Δ) than for close populations. These conclusions are valid for the symmetric TLC and the asymmetric one.

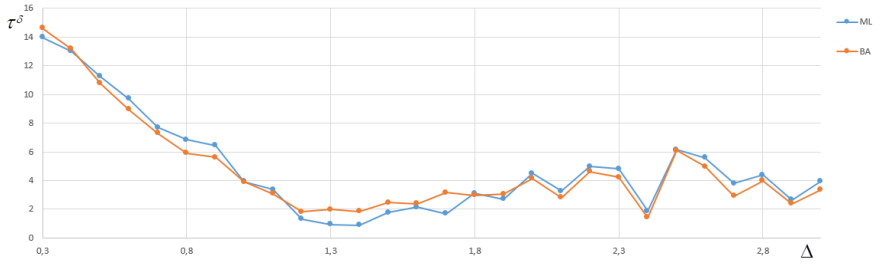


Figure 3.4 Values of τ^δ for different parameter estimation methods.

The comparison of the two approaches to parameter estimation is made based on the values of the EER incurred by the classification rule based on the PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ (1.14). The proposed optimality criterion is based on the derived formula of the AER. The simulation experiment shows the advantage of the BA estimation approach against the frequentist approach (ML). This advantage is more significant for strongly separated populations (larger values of Δ) than for close populations. These conclusions are valid for the symmetric TLC and the asymmetric one. The results give us strong arguments to expect that BA estimators of spatial population parameters could be effectively used in spatial Gaussian data classification incurred by PBDF. Based on the results, it can be concluded that the inclusion of prior information in the models influences the error rate estimates, and this influence increases significantly when the considered classes are closer to each other.

Also, GRF observations classification is performed in a multiclass case by analysing the situation when the population number $m=3$. Assume that $Z(s)$ is a GRF observation and $s\in D$ represents a location where the observations are taken. Let $D\subset\mathbb{R}^2$ that denotes a spatial domain of interest

is a regular 2-dimensional lattice with unit spacing. A selected STL S_{12}, consists of twelve third-order neighbours. STL is partitioned into a union of two disjoint subsets, i.e., $S_{12} = \bigcup_{l=1}^{3} S^{(l)}$, where $S^{(l)}$ is a subset of S_{12} that includes N_l locations of the feature observations from Ω_l, $l = \overline{1,3}$. All of the simulations consider small training sample sizes, $N_l = 4$, $l = 1,2,3$. In Figure 3.5, the points indicated by *A* belong to $S^{(1)}$, the points indicated by *B*, belong to $S^{(2)}$, and the points indicated by *C*, belong to $S^{(3)}$. A sign x denotes the focal location s_0. In this stage of the research, it is chosen to examine the dependence of AER estimates on the structure of STL: STL with the grouped label (STLG) and the mixed label (STLM). It is assumed that STLG and STLM are fixed, and configurations used in this stage are shown in Figure 3.5.

FO $Z_0 = Z(s_0)$, where $s_0 = (0,0)$ is the observation from GRF and can be described by GRF class-conditional density function (1.11) with conditional mean and conditional variance expressions (1.12). Conditional mean expression depends on the marginal (or non-conditional) mean $\mu_0^l = x'(s_0)\beta_l$ for class *l*. The case where the mean is constant is chosen for analysis. Conditional variance expression depends on marginal variance σ^2 and ρ_0. The case is chosen for analysis when $\sigma^2 = 1$ with the isotropic exponential covariance function specified by $c(d_{ij}) = \exp\{-d_{ij}/\phi\}$, where d_{ij} is the distance between spatial location, ϕ is a range parameter.

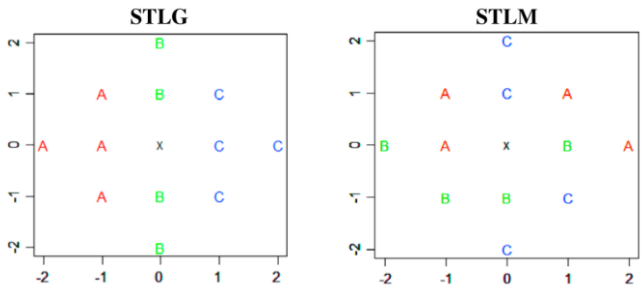


Figure 3.5 STL with different label distributions.

Two training samples for different STL (Figure 3.5) are simulated using multivariate Gaussian distribution with the selected parameter values: with different mean μ_0^l and the same variance $\sigma^2 = 1$, and freely selectable range parameter ϕ equal 1, 2 or 3, in class l, $l = 1,2$. A parameter *b* is used to define

constant mean: $\mu_0^1 = b$, $\mu_0^2 = 0$ and $\mu_0^3 = -3b$. The marginal squared Mahalanobis distance $\Delta_{lk}^2 = \left(\mu_0^l - \mu_0^k\right)/\sigma^2$ is used to estimate the distance between the populations, in the case under consideration Mahalanobis distance: $\Delta_{12} = b$, $\Delta_{13} = 4b$ and $\Delta_{23} = 3b$. The conditional Mahalanobis distance: $d^{lk} = \Delta_{lk}/\sqrt{\rho_0}$, ρ_0 denoted a ratio between conditional and marginal variances. Here *b* represents the separation between classes and is called the separation step. The separation between classes increases with increasing of *b*. Each case is simulated 1000 times.

Using the ML method, parameter estimates are obtained: β and σ^2 based on Z are $\hat{\beta}_{ML} = \left(X'R^{-1}X\right)XR^{-1}Z$, $\sigma^2 = \left(Z - X\hat{\beta}_{ML}\right)' R^{-1}\left(Z - X\hat{\beta}_{ML}\right)/N$, where *X* is a $N \times 3q$ design matrix for *Z*, *R* is a spatial correlation matrix and *N* is an observation number. The bias-adjusted ML estimator of σ^2 is $\hat{\sigma}^2 = \hat{\sigma}_{ML}^2 N/(N-3q)$. The ML estimators of β and σ^2 are used in the discriminant function evaluation, i.e., $\hat{\beta} = \hat{\beta}_{ML}$, $\hat{\sigma}^2 = \hat{\sigma}_{ML}{}^2 N/(N-3q)$.

Population parameters estimates obtained using the ML method for each simulation will be used to solve the supervised learning classification problem. The task is to classify Z_0 into one of three populations, the classification is performed using PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ (1.14). These expressions depend on the estimates of the population parameters as well as on the probabilities of the class labels. In the case under consideration class label probabilities are equal: $\pi_0^l = 1/3$, $l = 1, 2, 3$.

After classification using PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ (1.14) empirical estimators of AER specified in (1.15) are obtained for each simulation. An empirical estimator of the EER obtained by averaging AER over-runs is proposed as a measure for comparison. The values of $\overline{EER}$ are presented for STLG and STLM in Table 3.2.

Table 3.2 contains the $\overline{EER}$ values, which are calculated with different levels of spatial correlation and different levels of class separation. They show that $\overline{EER}$ decreases as the separation step increases for both STL structure cases. Also, these values are represented graphically.

Table 3.2 $\overline{EER}$ values for STL with different label distributions.

STL		STLG			STLM	
b	$\phi=1$	$\phi=2$	$\phi=3$	$\phi=1$	$\phi=2$	$\phi=3$
0.5	0.414882	0.352129	**0.303666**	0.425847	0.409503	**0.398642**
1	0.228513	0.167294	**0.131867**	0.249523	0.228437	**0.221165**
1.5	0.132704	0.08409	**0.056355**	0.152461	**0.137678**	0.139104
2	0.079256	0.040844	**0,021153**	0.100267	0.08558	**0.082012**
2.5	0.046043	0.017309	**0.006761**	0.063318	0.055167	**0.051326**
3	0.024973	0.00669	**0.001876**	0.040019	**0.029397**	0.029753
3.5	0.01263	0.002277	**0.000398**	0.022736	0.017271	**0.01673**
4	0.005914	0.000649	**7.03E-05**	0.012825	**0.007675**	0.008792
4.5	0.002551	0.000168	**9.3E-06**	0.005996	0.003556	**0.003194**

In Figure 3.6, the variation of $\overline{EER}$ depends on the parameters b, and also the curves represent the situations defined by the parameter ϕ values. In this case, the $\overline{EER}$ graphs show that as the distance between populations increases, the error estimates obtained decrease. The different curves reflect the different covariance functions used, differing in the range parameter ϕ. Comparing the $\overline{EER}$ graphs for different tag structures, it can be seen that the $\overline{EER}$ for the grouped tag structure of classes decreases as the spatial correlation range increases.

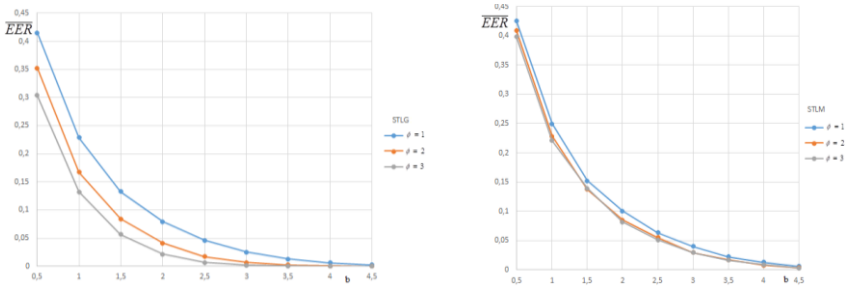


Figure 3.6 Values of the $\overline{EER}$ for different STL and parameter estimates.

For comparison, the ratio $\kappa = \overline{EER}^{STLG} / \overline{EER}^{STLM}$ is calculated in Figure 3.7. Different covariance functions are used for the curve graphs, differing by the range parameter ϕ. As the distance between populations increases, so does the advantage of the clustered structure of class markers.

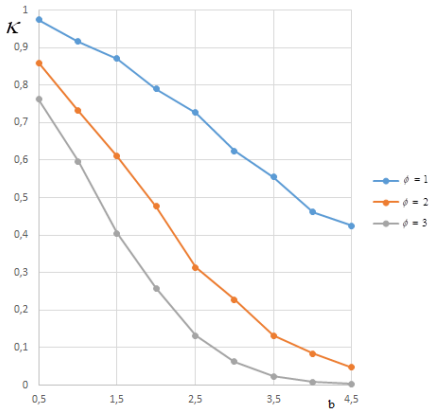


Figure 3.7 Curves of κ values with different ϕ.

A feature values vector Z model by GRF is investigated. During the experiment, the classification is performed for three classes. STL with the grouped label (STLG) and the mixed label (STLM) distributions are investigated. The comparison of STL structures is made based on the values of the $\overline{EER}$ incurred by the classification rule based on the PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ (1.14). The simulation experiment shows the advantage of STLG approach over STLM. This advantage is more significant for strongly separated populations (larger values of *b*) than for close populations. The results give us strong arguments to expect that STLG structure of spatial population could be effectively used in spatial Gaussian data classification incurred by PBDF.

T-distributed Random Field. After observations from GRF classification, the TRF observations models are analysed. Assume that $Z(s)$ is a TRF observation and $s \in D$ represents a location where the observations are taken.

Let $D \subset \mathbb{R}^2$ that denotes a spatial domain of interest is a regular 2-dimensional lattice with unit spacing. A selected STLs S_8, S_{24}, contain 8 and 24 neighbours of focal location s_0. Different STL is selected to evaluate whether the number of neighbours influences error rate estimations. Each STL is partitioned into a union of two disjoint subsets, i.e., $S_N = S^{(1)} \cup S^{(2)}$, where *N* is a training sample size (in this situation equal to the number of neighbours) $S^{(l)}$ is a subset of S_N that includes N_l locations of the feature observations from Ω_l, $l = 1, 2$. It is assumed that STLs are fixed and configurations used in this stage are shown in Figure 3.8.

FO $Z_0 = Z(s_0)$, where $s_0 = (1,1)$ is the observation from TRF and can be described by TRF class-conditional density function with conditional mean and conditional scaling parameter expressions (1.17) and (1.18). Conditional mean expression depends on the (marginal or non-conditional) mean $\mu_0^l = x'(s_0)\beta_l$. for class *l*. In the case where the mean is constant, variance is known, $\sigma^2 = 1$ with the isotropic exponential covariance function specified by $c(d_{ij}) = \exp\{-d_{ij}/\phi\}$, where d_{ij} is the Euclid distance between spatial points, ϕ is a range parameter that is chosen for analysis.

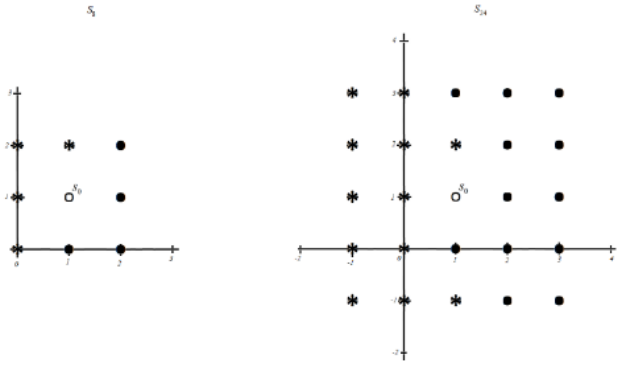


Figure 3. 8 S_8 (left) and S_{24} (right) with $S^{(1)}$, $S^{(2)}$ points that are marked as asterisks and dots.

Two training samples for different STL (Figure 3.8) are simulated using multivariate *t*-distribution with the selected parameter values: with different mean μ_0^l which is chosen with the assumption that $\Delta = \mu_0^1 - \mu_0^2 > 0$, chosen freely $\Delta = (0.5; 0.7; ...; 1.5)$, different range parameter values chosen freely: $\phi = (0.1; 0.5; ...; 3.3)$ and the same variance $\sigma^2 = 1$ in class l, $l = 1,2$. This set of parameters is chosen freely to study the dependence of AER estimates on the considered population parameters. Each case is simulated 1000 times, $M = 1000$.

In the considered case, it is assumed that the mean parameters β_l, $l = 1,2$ are unknown. Population parameters are estimated using ML and LS methods, to compare whether the parameter estimation method affects the classification accuracy.

Using the ML and LS methods, parameter $\beta' = (\beta_1', \beta_2')$ estimates based on Z are $\hat{\beta}_{ML} = \left(X'R^{-1}X\right)^{-1} X'R^{-1}Z$ and $\hat{\beta}_{LS} = \left(X'X\right)^{-1} X'Z$. From the properties of the multivariate *t*-distribution, it follows that $\hat{\beta}_{ML} \sim T_{2q}\left(\beta, \sigma^2 R_{ML}(v-2)/v, v)\right)$ and $\hat{\beta}_{LS} \sim T_{2q}\left(\beta, \sigma^2 R_{LS}(v-2)/v, v)\right)$, with ν denoting degrees of freedom, where $R_{ML} = \left(X'R^{-1}X\right)^{-1}$ and $R_{LS} = (X'X)^{-1}X'RX(X'X)^{-1}$ with R being a spatial correlation matrix, X being a design matrix for Z.

Population parameters estimates obtained using ML and LS methods in each simulation will be used to solve the supervised learning classification problem. The task is to classify Z_0 into one of two populations when $x'(s_0)\hat{\beta}_1 \neq x'(s_0)\hat{\beta}_2$. The classification is performed using Algorithm 1. This algorithm depends on the estimates of the population parameters as well as on the probabilities of the class labels. In the case under consideration, class label probabilities are estimated from the training sample. In different STLs (Figure 3.8) $\pi_0^l = 0.5$.

After classification, the empirical estimators of AER, obtained using Algorithm 3, are specified in (2.4) for each simulation. An empirical estimator of the EER obtained by averaging AER over-runs is proposed as a measure for comparison. Also, Bayes error rate estimates are received, when real parameter values are used. The values of error rates are presented for S_8 and for S_{24}, for different marginal Mahalanobis distances and for different range parameters ϕ in Table 3.3.

By analysing the estimates in Table 3.3, it can be concluded that both estimators of the $\overline{EER}$ monotonically decrease when Δ and ϕ decreases. For all cases $\overline{EER}^{LS} \geq \overline{EER}^{ML}$ for S_8 and S_{24}. For all cases, both $\overline{EER}^{ML}$ and $\overline{EER}^{LS}$ are more significant than P_0^B.

From Table 3.3, it can be concluded that the ML case has an advantage over the LS case in the sense of the minimal value of the $\overline{EER}$. It can also be said that the inclusion of more neighbouring points reduced the classification error estimates.

Table 3.3 Values of P_0^B, $\overline{\overline{EER}}^{ML}$ and $\overline{\overline{EER}}^{LS}$ for various Δ and ϕ.

	Δ	0.5		0.7		0.9		1.1		1.3		1.5	
ϕ	Error type	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}
0.1	P_0^B	0.3467	0.3467	0.2931	0.2930	0.2457	0.2456	0.2049	0.2048	0.1705	0.1704	0.1418	0.1417
	$\overline{\overline{EER}}^{ML}$	0.3539	0.3501	0.3025	0.2971	0.2565	0.2501	0.2164	0.2095	0.1820	0.1750	0.1530	0.1461
	$\overline{\overline{EER}}^{LS}$	0.3539	0.3501	0.3025	0.2971	0.2565	0.2501	0.2164	0.2095	0.1820	0.1750	0.1530	0.1461
0.5	P_0^B	0.3421	0.3426	0.2872	0.2879	0.2391	0.2399	0.1980	0.1989	0.1636	0.1645	0.1351	0.1360
	$\overline{\overline{EER}}^{ML}$	0.3508	**0.3470**	0.2985	**0.2932**	0.2520	**0.2457**	0.2115	**0.2048**	0.1770	**0.1702**	0.1480	**0.1414**
	$\overline{\overline{EER}}^{LS}$	0.3510	0.3472	0.2988	0.2935	0.2523	0.2460	0.2118	0.2051	0.1773	0.1705	0.1483	0.1417
0.9	P_0^B	0.3213	0.3210	0.2617	0.2615	0.2114	0.2111	0.1700	0.1698	0.1366	0.1366	0.1101	0.1101
	$\overline{\overline{EER}}^{ML}$	0.3327	**0.3276**	0.2762	**0.2694**	0.2273	**0.2196**	0.1862	**0.1782**	0.1522	**0.1444**	0.1246	**0.1173**
	$\overline{\overline{EER}}^{LS}$	0.3338	0.3298	0.2775	0.2719	0.2288	0.2223	0.1876	0.1809	0.1536	0.1470	0.1259	0.1197
1.3	P_0^B	0.2989	0.2992	0.2353	0.2357	0.1837	0.1842	0.1432	0.1436	0.1119	0.1124	0.0880	0.0885
	$\overline{\overline{EER}}^{ML}$	0.3139	**0.3093**	0.2535	**0.2473**	0.2030	**0.1960**	0.1619	**0.1549**	0.1292	**0.1226**	0.1035	**0.0975**
	$\overline{\overline{EER}}^{LS}$	0.3167	0.3160	0.2567	0.2546	0.2063	0.2034	0.1652	0.1619	0.1324	0.1290	0.1064	0.1032
1.7	P_0^B	0.2789	0.2792	0.2127	0.2130	0.1612	0.1614	0.1223	0.1224	0.0935	0.0934	0.0722	0.0720
	$\overline{\overline{EER}}^{ML}$	0.2983	**0.2915**	0.2348	**0.2265**	0.1833	**0.1747**	0.1429	**0.1347**	0.1118	**0.1043**	0.0880	**0.0814**
	$\overline{\overline{EER}}^{LS}$	0.3032	0.3000	0.2401	0.2362	0.1886	0.1845	0.1479	0.1439	0.1162	0.1125	0.0918	0.0885

Continued on next page

Table 3.3 Continued from previous page.

	Δ	0.5		0.7		0.9		1.1		1.3		1.5	
ϕ	Error type	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}
2.1	P_0^B	0.2592	0.2598	0.1910	0.1916	0.1400	0.1407	0.1032	0.1037	0.0768	0.0773	0.0580	0.0584
	$\overline{\overline{EER}}^{ML}$	0.2822	**0.2719**	0.2161	**0.2047**	0.1643	**0.1531**	0.1249	**0.1147**	0.0955	**0.0866**	0.0737	**0.0662**
	$\overline{\overline{EER}}^{LS}$	0.2878	0.2831	0.2223	0.2174	0.1704	0.1657	0.1305	0.1263	0.1005	0.0968	0.0780	0.0749
2.5	P_0^B	0.2447	0.2456	0.1759	0.1770	0.1261	0.1273	0.0911	0.0923	0.0668	0.0679	0.0498	0.0508
	$\overline{\overline{EER}}^{ML}$	0.2670	**0.2585**	0.2001	**0.1903**	0.1490	**0.1396**	0.1113	**0.1029**	0.0837	**0.0767**	0.0637	**0.0579**
	$\overline{\overline{EER}}^{LS}$	0.2724	0.2740	0.2061	0.2069	0.1550	0.1553	0.1167	0.1168	0.0884	0.0884	0.0677	0.0676
2.9	P_0^B	0.2311	0.2280	0.1621	0.1588	0.1137	0.1106	0.0808	0.0780	0.0584	0.0560	0.0431	0.0411
	$\overline{\overline{EER}}^{ML}$	0.2549	**0.2382**	0.1867	**0.1697**	0.1362	**0.1207**	0.0999	**0.0866**	0.0742	**0.0630**	0.0560	**0.0467**
	$\overline{\overline{EER}}^{LS}$	0.2619	0.2566	0.1944	0.1885	0.1436	0.1376	0.1065	0.1008	0.0798	0.0745	0.0606	0.0558
3.3	P_0^B	0.2188	0.2169	0.1499	0.1482	0.1032	0.1017	0.0722	0.0709	0.0516	0.0506	0.0377	0.0369
	$\overline{\overline{EER}}^{ML}$	0.2460	**0.2315**	0.1771	**0.1623**	0.1273	**0.1139**	0.0923	**0.0809**	0.0679	**0.0584**	0.0508	**0.0430**
	$\overline{\overline{EER}}^{LS}$	0.2536	0.2478	0.1850	0.1800	0.1345	0.1304	0.0984	0.0951	0.0728	0.0703	0.0547	0.0527

Concluded

The visual comparison of these two cases of parameter estimators is also made by plotting the index $\kappa = \overline{EER}^{ML} \big/ \overline{EER}^{LS}$ values. The dependence of the values of κ the index for S_{24} on the distance between populations Δ for the parameter ϕ is shown in Figure 3.9.

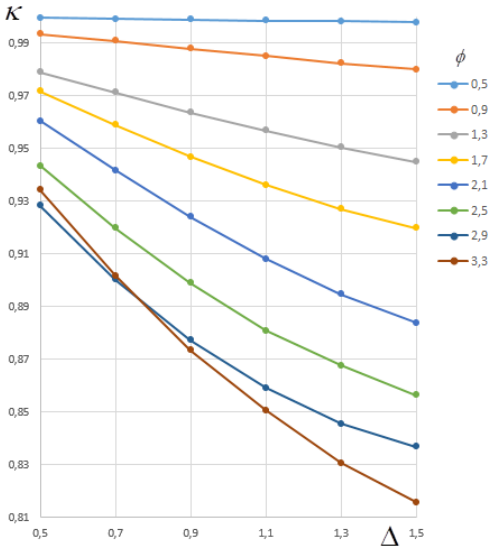


Figure 3.9 Curves of κ values for ϕ and S_{24}.

When the parameter ϕ is constant, the values of the index κ decreases when the distance between populations increases. This trend is more obvious when the analysed parameter ϕ values increase in the interval [0.5, 3.3] by 0.4. The dependence of the values of κ the index on the parameter ϕ for Δ is shown in Figure 3.10. When the parameter Δ is constant then values of the index κ decrease when the parameter of the covariance function ϕ increases. This trend is more obvious when the analysed distance between populations Δ values increases in the interval [0.5, 1.5] by 0.2.

The simulation experiment shows the advantage of the PBDF based on the ML estimator against the one based on the LS estimator. This advantage is more significant for the cases with stronger spatial dependence between observations (i. e., larger values of ϕ). This conclusion is valid for the different distances between populations Δ. Hence the results give us strong arguments that often intractable ML estimators of spatial mean parameters should be used in highly correlated spatial data modelled by TRF, and the simpler LS estimators could replace these estimators for weakly correlated spatial data without significant loss of the PBDF performance.

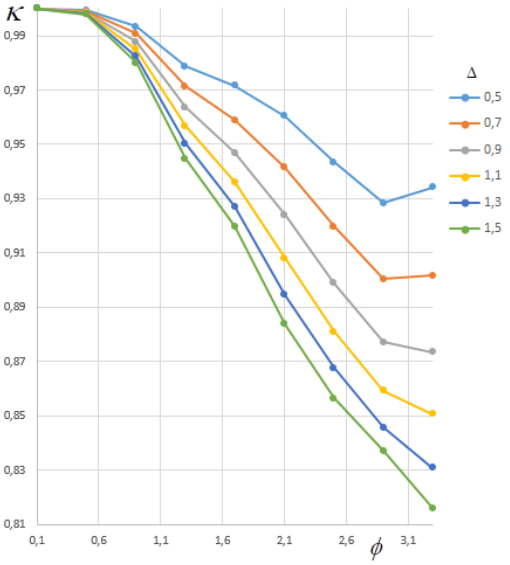


Figure 3.10 Curves of κ values for Δ and S_{24}.

Spatial auto-beta model (SABE). The next stage of work is applying the classification procedure to observations from an exponential family of distributions. At this stage of the study, it is chosen to solve the classification problem for a univariate observation from beta distribution using Algorithm 1, when the auto-models are used for spatial information incorporation into the data model.

Assume that $Z(s)$ is a RF with conditional beta distribution observation and $s \in D$ represents a location where the observations are taken. Let $D \subset \mathbb{R}^2$ that denotes a spatial domain of interest be a regular 2-dimensional lattice 16×16 with unit spacing. Assume that feature values belong to the interval (0,1) and the class label takes only the value 1 or 2. A class label for spatial locations is assigned at random with probability 0.5. The area comprises training and testing location sets (80% and 20%, respectively). A selected STL S_{205} consists of 205 spatial locations. STL is partitioned into a union of two disjoint subsets of locations, i.e., $S_{205} = S^{(1)} \cup S^{(2)}$, where $S^{(l)}$ is a subset of S_{205} that includes N_l locations of the feature observations from Ω_l, $l = 1,2$. It is assumed that STL of fixed training locations with known class labels and feature observations and testing sample with 51 spatial locations is created. A configuration used in this stage is shown in Figure 3.11.

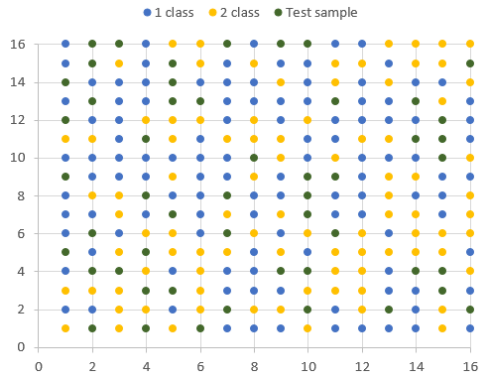


Figure 3.11 STL and testing sample for the auto-beta model.

FO $Z_0 = Z(s_0)$, where s_0 is a focal location, is the observation from RF with conditional beta distribution, and can be described by conditional density function (1.28) with conditional mean and conditional precision expressions (1.29), (1.30). Conditional mean and conditional precision expressions depend on the natural parameter A_{0h}^{l} for class l, $l=1,2$ where h is a sufficient statistics index. The beta distribution has two sufficient statistics, so there are two A_{0h}^{l} expressions (1.30) for different classes l, $l=1,2$. Two types of parameters are used in the A_{0h}^{l}, parameters represented large shape variation, these parameters are different for population Ω_l, $l=1,2$ and parameters represent small shape variation and spatial information between Z_0 and training sample observations, which are common for population Ω_l, $l=1,2$. At this stage, large shape variation is chosen to be modelled using spatial location coordinates. Values of coordinates x_1, x_2 are included in the model as regressors with corresponding parameters, for the population Ω_l, $l=1,2$. Since for beta distribution, there are h, $h=1,2$ natural parameter expressions A_{0h}^{l} for class $l, l=1,2$. it is chosen to include spatial coordinates with different parameters in each of them i.e., $\beta_l^{(h1)}, \beta_l^{(h2)}$ where l is a class index, $l=1,2$, h denotes an index of sufficient statistics, $h=1,2$, and an upper second superscript denotes a regressor index (coordinates).

In this empirical investigation, spatial dependence parameters with first-order neighbours are used to describe small shape variations with spatial information. First-order neighbours are chosen to define the NN set for FO, required for the creation of auto-models (detail in subsection 1.3). First-order neighbour scheme: each spatial location $s_i \in D$ has four nearest neighbours

spatial locations denoted as NN set: $NN_i = \{s_{ie} = i + (1,0),\ s_{iw} = i - (1,0),\ s_{in} = i + (0,1),\ s_{is} = i - (0,1)\}$ with neighbour adjustments near the boundary and s_{ie} is a neighbour of s_i form east, *n* denotes north, *w* denotes west, *s* denotes south. Two parameters η_1, η_2 are chosen to describe spatial dependence: η_1 denotes spatial dependence between Z_0 and east-west direction first-order neighbours from training sample observations and η_2 denotes spatial dependence between Z_0 and north-south direction first-order neighbours from training sample observations. The dependence of the feature values on the spatial location coordinates and the neighbouring points' feature values are investigated when natural parameter values A_{0h}^l specified in (1.30) get the expressions:

$$A_{01}^l = \beta_l^{(11)} x_1(s_0) + \beta_l^{(12)} x_2(s_0) - \\ - \eta_1 \left(\ln(1 - Z_{je}) + \ln(1 - Z_{jw}) \right) - \eta_2 \left(\ln(1 - Z_{jn}) + \ln(1 - Z_{js}) \right), \tag{3.2}$$

$$A_{02}^l = \beta_l^{(21)} x_1(s_0) + \beta_l^{(22)} x_2(s_0) \\ - \eta_1 \left(\ln(Z_{je}) + \ln(Z_{jw}) \right) - \eta_2 \left(\ln(Z_{jn}) + \ln(Z_{js}) \right), \tag{3.3}$$

where $x_1(s_0), x_2(s_0)$ are the coordinates of spatial location for Z_0; $Z_{je} = Z(s_{je})$ is a training sample feature value that belongs to the NN set of FO Z_0. In this case a set of parameters $\Psi = \left\{ \beta_l^{(h1)}, \beta_l^{(h2)}, \eta_1, \eta_2,\ h, l = 1,2 \right\}$. Using (3.2) and (3.3), the conditional mean (1.29) and conditional precision (1.30) expressions are obtained and used to define the conditional density function (1.28) and RF observations can be simulated, choosing the appropriate values of the model parameters. In this empirical study, RF with conditional beta distribution observations is simulated using Algorithm 2 with natural parameter expressions (3.2) and (3.3). A few different scenarios are chosen that differ in model shape defined by parameter values Ψ, and parameter values are chosen freely. Two types of parametrical structures are chosen: when all parameters are fixed except the class 2 regression parameters and when the spatial dependency parameter that describes the effects of the north-south neighbourhood points is changing. Selected parameter values are presented in Table 3.4.

Table 3.4 Parametric structures.

Parameter	Structure	
β_1	A, B	$\beta_1 = \mathbf{1}_4$
β_2	A	$\beta_2' = \left(1,1,\beta_2^{(21)},\beta_2^{(22)}\right), \beta_2^{(21)} = \beta_2^{(22)}, \beta^* = \left\lvert\beta_2^{(22)} - \beta_1^{(22)}\right\rvert$ $\beta^* = 0.5;1;2;3;4;5;6;7;8;9;19;29;39;49$
	B	$\beta_2' = \left(1,1,2,2\right)$
η	A	$\eta_1 = 1, \eta_2 = 2$
	B	$\eta_1 = 1, \eta^* = \left\lvert\eta_2 - \eta_1\right\rvert$ $\eta^* = 0;1;2;3;4;5;6;7;8;18;28;38;48$

For each parameter set using Algorithm 2, feature value Z is simulated for each point from STL, $M = 100$. Visualisation of univariate densities for different scenarios is presented in Figure 3.12. The class densities are defined by structure A; with it, the classes become more distinct as the difference between regression parameters denoted by β^* increases. The densities of the classes are determined by structure B, and with it, the classes become less divergent as the difference between spatial dependency parameters denoted by η^* increases.

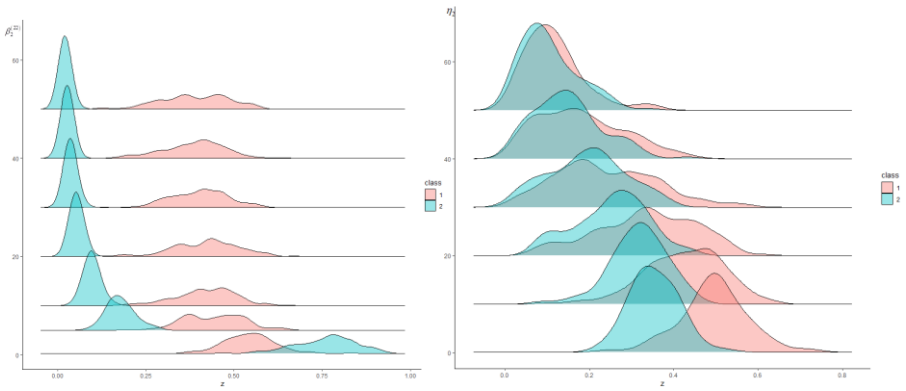


Figure 3.12 Univariate densities for different scenarios: structure A (left) and structure B (right).

Population parameters estimates obtained using the MPL method in each simulation will be used to solve the supervised learning classification problem. This work focuses on the SABE and supervised classification problem with fixed STL when training samples T are given. The task is to classify Z_0 into one of two populations, the classification is performed using Algorithm 1. PBDF $W^B\left(Z_0,\hat{\Psi}\right)$ (2.8), used in Algorithm 1, depends on the

estimates of the populations parameters and on the probabilities of the class labels. Also, classification results are compared to the classification of FOs using modified Linear Discriminant Function (LDF). The modified LDF function is used where class conditional means and conditional dispersions are used for the estimation. The modified LDF function:

$$W^L\left(Z_0,\Psi\right)=\frac{\left(Z_0-\frac{\mu_{0z}^1+\mu_{0z}^2}{2}\right)\left(\mu_{0z}^1-\mu_{0z}^2\right)}{\left(\pi_0^1\sigma_{0z1}^2+\pi_0^2\sigma_{0z2}^2\right)}+\gamma_0\left(\Psi\right), \tag{3.4}$$

with conditional variance σ_{0zl}^2, $l=1,2$:

$$\sigma_{0l}^2=\mathrm{var}\left(Z_0\left|Z_j,\Psi\right.\right)=\frac{\mu_{0z}^l\left(1-\mu_{0z}^l\right)}{1+\phi_{0z}^l},$$

where μ_{0z}^l,ϕ_{0z}^l, $l=1,2$ are conditional mean and conditional precision for the beta distribution specified in (1.29), $\gamma_0\left(\Psi\right)=\ln\left(\pi_0^1/\pi_0^2\right)$. When a set of model parameters Ψ is unknown, it is replaced by a set of its estimates $\hat{\Psi}$. Plug-in LDF (PLDF) $W^L\left(Z_0,\hat{\Psi}\right)$ is formed by plugging the estimators of the parameters into (3.4). The AER associated with PLDF specified in (3.4) has the following form:

$$P_{0z}^L\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}\ , \tag{3.5}$$

where $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W^L\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int H\left(\left(-1\right)^l W^L\left(u,\hat{\Psi}\right)\right)p_l\left(u\right)du,$ with $B_l=\left\{u:u\in\left(0;1\right)\right\}$ is a scope of integration, probability measure $\hat{P}_{lz}$ based on conditional beta distribution with PDF p_l specified in (1.28), $l=1,2$.

In the case under consideration class label probabilities are evaluated in three different ways. First, the simplest probabilities are considered equal: $\pi_0^l=0.5, l=1,2$. The second way is when the probabilities are calculated using the inverse distance function (1.2) with all training sample observations. The third is when probabilities are calculated using inverse distance function (1.2) for neighbour training sample observations of up to fourth-order.

After classification using Algorithm 1, the empirical estimators of AER, obtained using Algorithm 3, specified in (2.9) for each simulation. An empirical estimator of the EER, obtained by averaging AER over-runs is proposed as a measure for comparison, $\overline{EER}$ (3.1). The $\overline{EER}^B$ values based on BDF for different class label probability evaluated methods and different

population parameter structures with the simulated realizations $M = 100$ are presented in Figure 3.13.

In Figure 3.13 different curves reflect class label probabilities calculation methods: structure A, where β^* increases, and structure B, where spatial dependency parameter η^* increases. According to Figure 3.13, the graph on the left side shows that when β^* increase, $\overline{EER}^B$ values decrease. It can be concluded that populations become more separable. But the difference between class label probability calculation methods is difficult to notice.

Figure 3.14 are presented ratios of $\overline{EER}^B$ values denoted by κ, the 3rd method with 1st and 2nd. This graph shows that when the β^* increases, ratio values decrease. It can be concluded that using the 3rd method, the classification is more accurate. According to Figure 3.13, the graph on the right side shows that when spatial dependency η^* increases, $\overline{EER}^B$ values increase too and classification accuracy decreases.

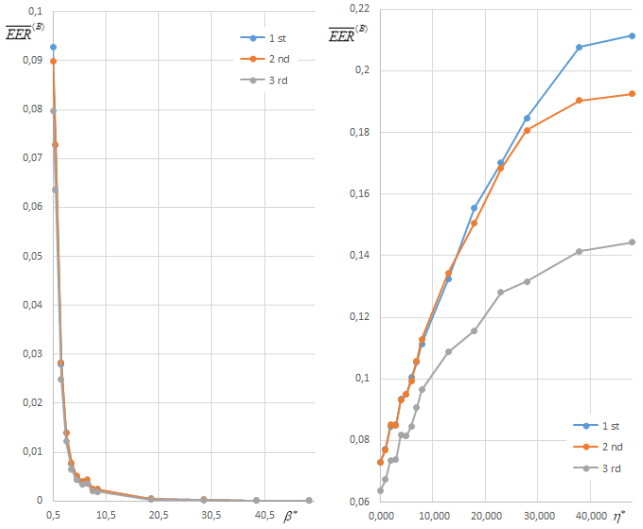


Figure 3.13 $\overline{EER}^B$ values curves with different class label probability calculations: structure A (left), structure B (right).

Figure 3.13 distinguishes a situation with the 3rd method, where class label probabilities are evaluated using an inverse distance function (1.2) for neighbour training sample observations of up to fourth-order. $\overline{EER}^B$ values are smaller than when probabilities are considered equal, or all training sample observations are incorporated into the estimation.

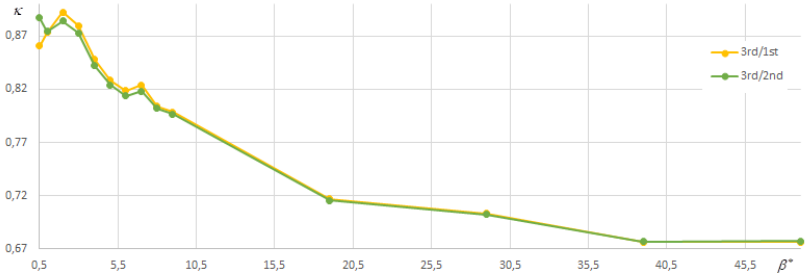


Figure 3.14 κ values, curves with different class label probability calculations for structure A.

FOs Z_0 classification using $W^L\left(Z_0,\hat{\Psi}\right)$ also is performed. The quantitative comparison of the two cases of the parameter estimators is also made using the index κ, $\kappa=\overline{EER}^B\big/\overline{EER}^L$ values, which are shown in Figure 3.15. In Figure 3.15. different curves reflect class label probabilities calculation methods: Structure A, where β^* increases and structure B, where η^* increases.

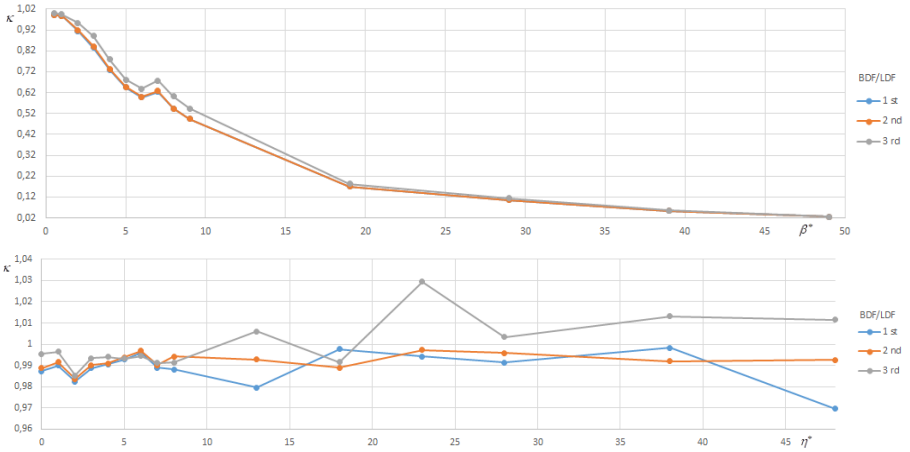


Figure 3.15 κ curves concerning β^* and η^* values, 1st, 2nd, 3rd marks method for class label probability estimation: structure A (first), structure B (second).

In structure A, when $\beta^*=0.5$, the $\kappa=\overline{EER}^B\big/\overline{EER}^L$ is greater than 1, and LDF based classification rule $W^L\left(Z_0,\hat{\Psi}\right)$ performs better. When $\beta^*>1$, the κ decreases and BDF based classification rule $W^B\left(Z_0,\hat{\Psi}\right)$ (2.8) gains an

advantage. In the structure A, when η^* is chosen less than 8, the κ is less than 1, and BDF based classification rule performs better. When the η^* value is selected 18 or greater, there are situations where LDF based classification rule gains an advantage.

The Wilcoxon signed rank test is chosen to evaluate the statistical comparison of the classification rules. This test is performed using $\overline{EER}$ values. Wilcoxon signed rank test uses magnitude and sign of the paired difference ranks. This procedure computes the differences between the performance scores of two classifiers on different population parameters sets. The results are presented in Table 3.5.

In Table 3.5, *V* is Wilcoxon signed rank test statistics. The null hypothesis of this method: LDF and BDF $\overline{EER}$ values do not differ. The null hypothesis can be rejected if the p-value is less than the selected significance level α.

Table 3.5 Wilcoxon signed rank test.

	Class label probability		
Structure	Equal probabilities	Inverse distance with all training sample obs.	Inverse distance for neighbour obs. of up to fourth-order
A	*V* = 136, (3.052e-05)	*V* = 120, (0.0007265)	*V* = 126, (0.0 02942)
B	*V* = 153 (0.0003204)	*V* = 103.5 (0.07023)	*V* = 47.5, (0.1769)

According to the results in Table 3.5, it can be said that in structure A with a rather small significance level (0.002942) LDF and BDF $\overline{EER}$ values differ for different class label probability calculation methods. In structure B LDF and BDF $\overline{EER}$ values differ for equal class label probability method and when class label probabilities are estimated by inverse distance with all training sample observations (0.07023).

The comparison of the two approaches to FO classification is made based on the values of the $\overline{EER}$. The proposed optimality criterion is based on the AERs specified in (2.9) and (3.5). Based on the obtained results it can be concluded that the inclusion of prior information related to focal locations and training sample locations in the models influences the error rate estimates. This influence increases significantly when the class label probabilities are estimated using a selected NN set of training sample observations. The simulation experiment shows the advantage of the approach based on BDF $W^B(Z_0, \Psi)$ (2.8) against the approach based on modified LDF $W^L(Z_0, \Psi)$

(3.4). This advantage is statistically significant for the examined parameter sets structure A, with a difference between regression parameters β^* and also for models where spatial context information is included only in the class-conditional probability distribution function expression.

In the next part of this chapter, the proposed classification Algorithm 1 is applied to solve the problem with real data. The task of determining the type of the bottom of the Baltic Sea is being solved, having algae coverage data sample. The coverage observed values are expressed as a percentage. For this reason, the SABE model is chosen for modelling these data.

3.3. Clasification Algorithm Application to Algae Coverage

The family of beta distributions offers a large variety of densities on bounded intervals $[a,b]$ or (a,b), which makes the beta models a potentially important class of data models. These models can be applied to the analysis of rates, proportions or concentration indices. Auto-models proposed by Besag (1974) allow applying the properties of beta models in the study of spatial data. The auto-beta models discussed in subsection 3.2 for simulated data at this stage of the empirical investigation are applicable for modelling algae coverage and solving the classification problem. The problem that is being solved is identifying the bottom of the sea using black carrageen concentration data. The classification problem is formed for two classes case: class 1 denotes no presence of boulders; class 2 denotes presence of boulders.

A spatially sampled algae coverage ratio database of sampling locations scattered in the Baltic Sea in the Lithuanian coastal waters between Palanga and Sventoji is analysed. One of the examined algae species – black carrageen (*Furcellaria lumbricalis*) – occurs in the Baltic Sea. The black carrageen forms extensive zones of dense meadows attached to different substrates of the underwater slope, such as gravel, cobble or boulders. The database is provided by the Marine Research Institute of Klaipėda University. Black carrageen concentration data used in this work is sampled at 641 spatial locations in the southeastern Baltic Sea. The size of each area is 1 m^2. Each algae coverage collecting spatial location (geographical coordinates of the site, wave generated orbital near-bottom velocity, Secchi depth, distance to sand) is also extracted directly from this database. The array of data received by the researchers is divided into two samples according to their expert knowledge: training and testing samples. Training and testing samples are selected to distribute the spatial locations throughout the study area. The data is divided into training and testing samples of $N = 448$, $N_1 = 354$, $N_2 = 94$

and $L=193$, $L_1=149$, $L_2=44$ spatial locations, respectively. Figure 3.16 shows the spatial positions of the sampling location.

The values of the feature of interest are the coverage ratio expressed as a percentage. In the database, there are no feature values that have a value of 100%, so dividing the values by 100% gives observations that have values from the interval $[0,1)$. In this case, $Z(s)$ is the spatial continuous observation on a bounded interval $[0,1)$, $s \in D$ represents a spatial location, where the observation is taken. Due to these characteristics, the beta distribution is chosen for modelling feature values. The first stage of the work consists of feature values modelling, analysing the dependence of the following values on regressors: wave-generated orbital near-bottom velocity, Secchi depth, and distance to the sand. During this stage for analysis of $Z(s)$, the following beta models are selected: GLM, Generalized Additive Model (GAM) and GAM with Spatial Random Field (GAM SRF). These models differ from each other in the chosen mean functions for beta distribution.

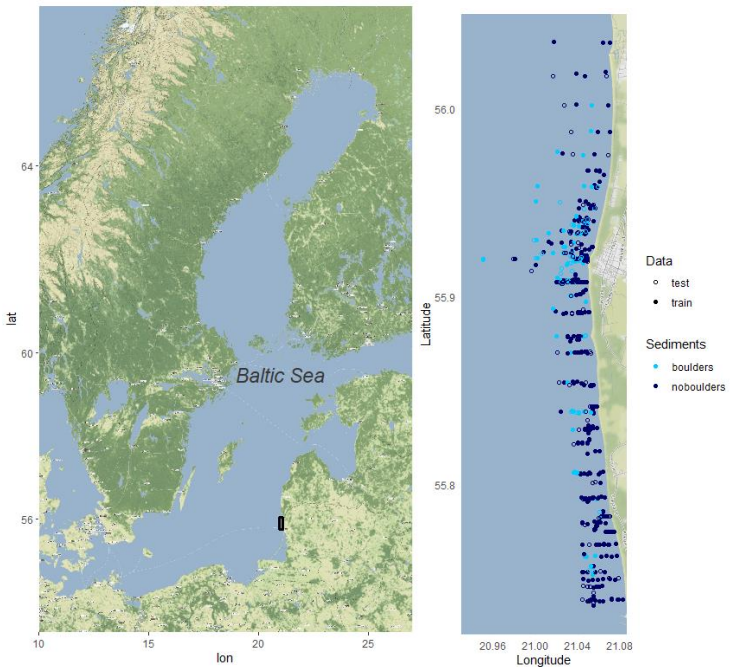


Figure 3.16 Spatial position of sampling location in the Baltic Sea.

Data Modelling. Several beta regression models are developed in the initial phase of data analysis. GLM and GAM (Zhang, 2004) have become standard

tools for analysing the impact of covariates on possibly non-Gaussian response variables. The only difference between GAM and GLM is that the former permits including nonlinear smooth functions in the model. When real data is at hand, the locations where it is picked are available. So the location of sampling units affects the response variable, and thus there would be a spatial correlation. A crucial question in setting up a spatial GAM for a particular application is how to model various spatial covariate effects. The spatial GAM is in the focus of this data modelling part. The spatial dependence in this type of model can be considered by introducing the covariance functions of location-specific random effects or/and implementing the smooth spatial functions in the mean structure (Nadrabadi et al., 2018). The latter is developed. The attention is restricted to the linear effects represented by regressors and nonlinear smoothed spatial effects.

In the database, there are more than 70% zero values acquired by feature values Z. In the R software, beta distribution models are created for random variables that acquire values from an open interval $(0,1)$, data transformation is necessary to be able to apply the already implemented function. For these reasons, a simple transformation is made to modify the data by a small amount $Z^*(s)=\left(Z(s)(N+M-1)/100+0,5\right)/(N+M)$. Further modelling uses the transformed data and returns to the initial denotation: $Z(s)=Z^*(s)$, $s\in D$. A beta regression model with different mean functions is chosen to be examined when the logit link function is used to model the mean as a function of covariates. The purpose of the logit link is to take a linear combination of the covariates values of the function of covariates and convert those values to the scale of a beta distribution mean, between 0 and 1. Consider that conditionally on μ, the Z is an independent random field with the marginal beta distribution. The model for Z, $s\in D$:

$$Z(s)\sim B\left(\mu(s),\phi(s)\right),$$

with marginal mean $\mu(s)$ and marginal precision $\psi(s)$. For the logit link function:

$$\text{logit}\left(\mu(s)\right)=x'(s)\beta+\psi(s),$$

where $x(s)$ is a $q\times 1$ vector of non-random regressors; β is a $q\times 1$ vector of regression parameters, $\psi(s)$ is an unpenalised cubic spline $\psi(s)=\sum_{j=1}^{K}B_j(s,k_j)$, where B_j is a basis of a smoother, k_j denotes a spline

knot, K is a number of knots. This chosen model without spatial correlation is called beta GAM. For comparison, a model with the same components is chosen that includes spatial correlation. The model for Z, $s \in D$:

$$Z(s) \sim B(\mu(s), \phi(s)),$$

for the logit link function:

$$\text{logit}(\mu(s)) = x'(s)\beta + \psi(s) + \upsilon(s),$$

where all components are equal to the model described above except the random effect part $\upsilon(s)$. $\{\upsilon(s) : s \in D\}$ is unobservable zero-mean stationary GRF representing the random effects with $\text{cov}(\upsilon(s+h), \upsilon(s)) = c(h)$ for all $s, h \in D$ So $\upsilon(s) \sim N(0, \Sigma)$, where Σ is a non-diagonal covariance matrix. This model is beta GAM SRF.

The models described above are created for the analysed data to assess whether the inclusion of spatial information affects the assessment of the model's feasibility. To achieve this goal, Pearson residuals variograms, fitted values and model selection criteria are analysed. Figure 3.17. shows the sample variograms of Pearson residuals of beta GAM and beta GAM SRF. Variogram of the model without spatial correlation indicated the presence of spatial correlation in the residuals.

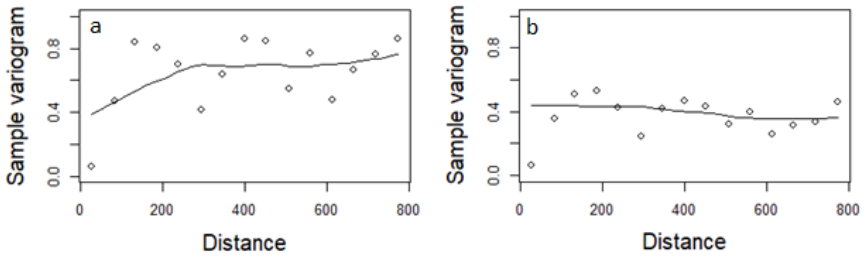


Figure 3.17 Variograms of Pearson residuals: beta GAM without spatial correlation (left), and beta GAM with spatial correlation (right).

The empirical variogram of the Pearson's residuals from the non-spatial model as discussed in Pfeiffer et al. (2008) indicates the presence of some unexplained spatial variation in the residuals (Figure 3.17 left). The residual spatial autocorrelation from the spatial model (Figure 3.17 right) gives a flatter variogram with a smaller amount of residual variation than the non-spatial model indicating that the spatial model has accounted for a larger amount of the spatially correlated variation in the prevalence data than the non-spatial model. Figure 3.18 shows fitted values versus the observed data. If the model

performs well and the scatter is precisely on a 1:1 line, then an exact fit is received, and the model is overfitting the data. Correlation coefficient values for models are presented here and indicate that the model with spatial correlation is considerably better.

This criterion is used for measuring the fit model with the given data. The Deviance Information Criterion (DIC) is similar to Watanabe Akaike's Information Criterion (WAIC). DICs and WAICs presented in Table 3.6 are compared; beta GAM SRF has lower DIC and AIC values.

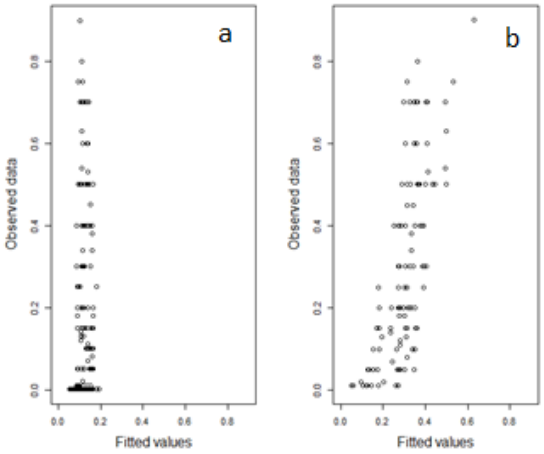


Figure 3.18 Observed algae coverage plotted versus fitted values obtained by beta GAM (left), and beta GAM SRF (right).

Table 3.6 DICs and WAICs criteria.

	DIC	**WAIC**
Beta GAM	-1438.166	-1443.391
Beta GAM SRF	-1521.901	-1531.754

Beta regression models without and with spatial effects are applied for the black carrageen dataset. Validation results show that analysing algae coverage utilising a beta regression model with random spatial random effect results in a better fit. For this reason, when solving the classification problem, i.e., identifying the bottom of the Baltic sea using black carrageen concentration data, it is decided to use spatial data models. Although GAM SRF hold one assumption, conditionally on μ, the set Z is an independent random field. After analysing the auto-models, it is decided to abandon this assumption and to use SABEs and spatial zero-inflated auto-beta models for the description of the feature values $Z(s)$, $s \in D$.

Classification. Auto-beta model. At this stage of the real data analysis, it is chosen to solve the spatial classification problem: identifying the bottom of the Baltic sea using black carrageen concentration as feature values $Z(s)$, $s \in D$. The classification problem is formed for two classes case: class 1 denotes no boulders; class 2 denotes boulders. Based on the fact that $Z(s)$ satisfies the properties of the beta distribution and also auto-models properties it is considered that $Z(s)$ is a RF observations with the conditional beta distribution. So FO $Z_0 = Z(s_0)$, where s_0 is a focal location, is the observation from RF with conditional beta distribution, and can be described by conditional density function (1.28) with conditional mean and conditional precision expressions (1.29), (1.30). Conditional mean and conditional precision expressions depend on the natural parameter A_{0h}^l for class *l*, $l = 1,2$ where *h* is a sufficient statistics index. The beta distribution has two sufficient statistics, so there are two A_{0h}^l expressions (1.30) for different classes *l*, $l = 1,2$. When compiling the models, it has been decided to include geographical coordinates of the spatial locations, as covariates, just like SABE for summarized data. The spatial dependence parameter η describes spatial relationships between FO and training sample observations belonging to NN of FO. The area where the data is collected $D \subset \mathbb{R}^2$ is continuous and the NN area for FO is defined by the distance. This distance is chosen based on how far training sample observations will be considered as FO neighbours. The dependence of the feature values on the spatial location coordinates and the neighbouring points' feature values are investigated when natural parameter values A_{0h}^l specified in (1.30) get the expressions:

$$A_{01}^l = \beta_l^{(11)} x_l(s_0) + \beta_l^{(12)} x_2(s_0) - \sum_{\substack{Z_j \in NN_0 \\ j \neq 0}} \eta \ln(1 - Z_j),$$

$$A_{02}^l = \beta_l^{(21)} x_1(s_0) + \beta_l^{(22)} x_2(s_0) - \sum_{\substack{Z_j \in NN_0 \\ j \neq 0}} \eta \ln(Z_j),$$

where $x_1(s_0), x_2(s_0)$ are coordinates for Z_0; $Z_j = Z(s_j)$ is a training sample feature value that belongs to the NN set of FO Z_0; $Z_j \in NN_0$, $j \neq 0$. A set of parameters is $\Psi = \left\{ \beta_l^{(h1)}, \beta_l^{(h2)}, \eta;\ h, l = 1,2 \right\}$.

The NN areas are selected freely to analyse several situations of interest and areas are defined at a selectable distance from the FO. In cases where the area is 20 m, the points to be classified have no or very few neighbours. The

distance is gradually increased, and situations, where more neighbouring spatial location information is included in the FOs model, are examined. Distances have been chosen by starting from 20 m and gradually increasing to 1 km. The other three distances are selected as follows. A distance is found with which the NN field contains at least one observation from each class. In this work, this distance is called the maxmin distance, i.e., 2852.204 m. The other two distances are double maxmin distance and half the maxmin distance. While analysing the data, the Euclidean distance is calculated, and the nearest neighbour areas are shown in Figure 3.19.

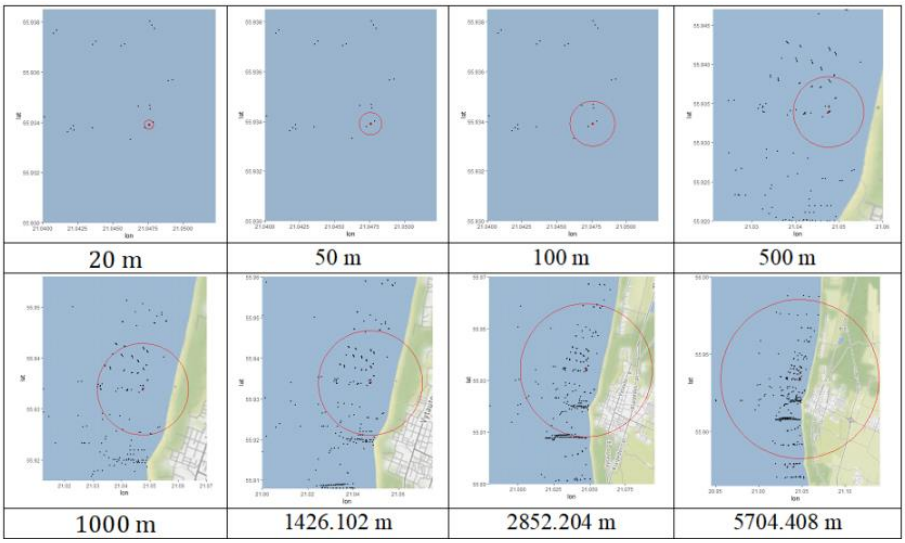


Figure 3.19 Nearest neighbour areas with different distances for model parameter estimation.

Eight SABE models and different NN areas of FOs for black carrageen concentration modelling is compiled. Population parameters estimates are obtained using the MPL method for the training sample in each model. The estimates are provided in Table 3.7. They will be used to solve the supervised classification problem, i.e. identifying the bottom of the Baltic sea.

The task is to classify Z_0 into one of two populations. The classification is performed using Algorithm 1. Also, the classification results are compared to the classification of FOs using modified LDF (3.4). In Algorithm 1 using PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ (2.8) and modified LDF (3.4) depend on the estimates of the population parameters as well as on the probabilities of the class labels. In the case under consideration class label probabilities are evaluated in three different ways. First, probabilities are estimated based on the sample size:

$\pi_0^l = N_l / N$, $l = 1,2$. The second way is when the probabilities are calculated using the inverse distance function (1.2) with all training sample observations. The third option is when probabilities are calculated using inverse distance function (1.2) for neighbour training sample observations of NN area for FO. This NN area is defined by a maximum minimum distance, that is the distance with which NN areas are formed for each FO, which contains at least one training sample observation from each population $\Omega_l, l = 1,2$.

Table 3.7 Estimates of parameters.

Nearest neighbour areas				
Parameter	20	50	100	500
$\beta_1^{(11)}$	-7.82E-02	-6.70E-02	-5.49E-02	-0.0632
$\beta_1^{(12)}$	6.42E-03	5.51E-03	4.52E-03	0.0052
$\beta_1^{(21)}$	2.74E+00	2.20E+00	3.34E+00	2.4278
$\beta_1^{(22)}$	-2.20E-01	-1.76E-01	-2.69E-01	-0.1946
$\beta_1^{(11)}$	2.41E-01	2.28E-01	2.18E-01	0.2232
$\beta_2^{(12)}$	-1.93E-02	-1.83E-02	-1.75E-02	`-0.0179
$\beta_2^{(21)}$	3.40E+00	3.11E+00	4.90E+00	3.9572
$\beta_2^{(22)}$	-2.72E-01	-2.49E-01	-3.94E-01	-0.3172
η	-2.98E+04	-4.60E+03	-1.11E+03	-39.8437
Nearest neighbour areas				
Parameter	1000	1426.102	2852.204	5704.408
$\beta_1^{(11)}$	-0.067	-0.0773	-9.71E-02	-0.097
$\beta_1^{(12)}$	0.0055	0.0063	7.96E-03	0.0079
$\beta_1^{(21)}$	2.4177	2.4879	2.67E+00	2.6778
$\beta_1^{(22)}$	-0.1938	-0.1994	-2.14E-01	-0.2146
$\beta_1^{(11)}$	0.2297	0.2308	2.34E-01	0.2424
$\beta_2^{(12)}$	-0.0184	-0.0185	-1.87E-02	-0.0194
$\beta_2^{(21)}$	3.954	4.0537	4.17E+00	4.3703
$\beta_2^{(22)}$	-0.3169	-0.3249	-3.35E-01	-0.3503
η	-13.9029	-5.677	-1.07E-05	-0.0005

The maximum minimal distance has the following form:

$$d_{\max\min} = \max_{s_i \in M} \min_{s_j \in N, l=1,2} \left(d_{ij}^{l} \right),$$

where d_{ij}^{l} is a distance between FO s_i, $s_i \in M$ and training sample observation $s_j, s_j \in N$.

After classification using Algorithm 1, the empirical estimators of AER, obtained using Algorithm 3, are specified in (2.9) for each SABE model. $\overline{EER}$ (3.1) values are proposed as a measure for comparison of classification based on BDF (2.8) and modified LDF (3.4). The $\overline{EER}$ values for different class label probability evaluated methods, and SABE models with different NN for FOs are presented in Table 3.8.

Table 3.8 $\overline{EER}$ values for SABE models.

Class label probability						
Distance of NN	Sample size		Inv. distance, all training sample		Inv. distance, max min distance	
	BDF	LDF	BDF	LDF	BDF	LDF
20	**8.52E-09**	9.75E-09	**1.09E-08**	1.21E-08	**1.60E-08**	1.76E-08
50	0.001749	0.001715	**0.001859**	0.001861	**0.001272**	0.001322
100	1.87E-07	1.63E-07	1.83E-07	1.54E-07	1.90E-07	1.59E-07
500	0.001235	0.001218	0.001549	0.001402	**0.001323**	0.001542
1000	0.00109	0.001028	**0.001041**	0.001108	**0.000958**	0.001095
1426.102	**0.000545**	0.00056	**0.000558**	0.000586	**0.00027**	0.000608
2852.204	**0.001254**	0.001288	**0.001312**	0.001369	**0.001514**	0.00153
5704.408	**0.001314**	0.001364	**0.001389**	0.001423	**0.001588**	0.00162

The $\overline{EER}$ values in Table 3.8 marked in bold indicate the advantage of the classification rule using BDF. It can be seen that the inclusion of contextual information in the calculation of class label probabilities related to the distance between the classified path and neighbouring locations from the training sample highlights the advantage of using the classification rule BDF specified in (5.13). A comparison of the ratio for the rules BDF and LDF values of $\overline{EER}$ is given in Figure 3.20. Based on the $\overline{EER}$ values, when the distance between the neighbours is between 20 and 50 m, the probability ratio approaches 1. When the distance is 100 m, the ratio is more than 1, and thus the classification rules based on LDF have the advantage. When the distance increases, the probability ratio decreases and BDF based classification rules gain the advantage. Also, when comparing different class label probability situations, when the distance increases, the BDF advantage can be seen when class label probabilities are calculated based on inverse distance, and the nearest

neighbours are chosen on the maximum minimal distance to the different class neighbours.

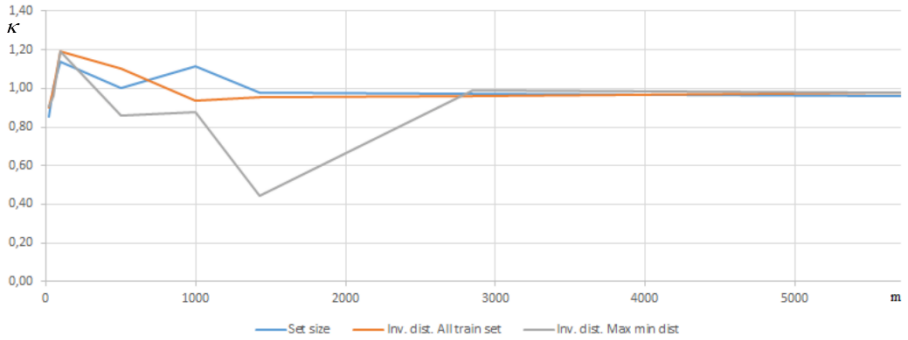


Figure 3.20 Curves of $\kappa = \overline{EER}^{B} / \overline{EER}^{L}$ with respect to distance.

Based on the obtained results it can be concluded that the inclusion of prior information related to focal locations and training sample locations in the models influences the error rate estimates, and this influence increases significantly when the class label probabilities are estimated using a selected NN set of training sample observations. The simulation experiment shows the advantage of the approach based on BDF $W^{B}(Z_0, \Psi)$ (2.8) against the approach based on modified LDF $W^{L}(Z_0, \Psi)$ (3.4).

These results are obtained for the transformed feature values $Z^{*}(s)$, to accommodate beta distribution models. However, the models examined above have not considered that there are more than 70% zero values in the data under consideration. For this reason, the zero-inflated auto-beta model is chosen for modelling the black carrageen dataset.

Classification. Zero-inflated auto-beta model. The initial values of the feature range from 0 to 100. After dividing by 100, the values can be summed up to [0,1), with more than 70% zero values. In this study, considering this property of the data, it is chosen to apply zero-inflated regression models. For data observed on [0,1) to model the data using a mixture of two distributions: a beta distribution and a degenerate distribution in a value 0 are selected.

It is considered that $Z(s)$ are RF observations with the conditional ZIB distribution. So FO $Z_0 = Z(s_0)$, where s_0 is a focal location, is the observation from RF with conditional ZIB distribution, and can be described

by conditional density function (1.31) with conditional mixture parameter and conditional PDF of beta distribution (1.28).

In order to specify the selection model for FO, it is enough to specify the natural parameters expression A_{0h}^{l} (1.32) for class l, $l=1,2$, where h is an index of sufficient statistics of ZIB distribution index.

When compiling zero-inflated auto-beta models for black carrageen concentration, it has been decided to include geographical coordinates of the spatial locations, as covariates, just like SABE for black carrageen data. The spatial dependence parameter η describes spatial relationships between FO and training sample observations of non-zero values from the population $\Omega_l, l=1,2$ belonging to NN of FO $Z_j = Z\left(s_j\right) \in NN_0^l,\ \ Z_j = Z\left(s_j\right) \neq 0,$ $s_j \neq s_0$. The NN area for FO is defined by the distance just like for SABE models. When modelling the conditional mixture parameter (the probability of 0), a logistic function is chosen to include parameters λ^l that represent large scale variation to be different for every class l, ρ^l represents spatial relationships between FO and training sample observations from population $\Omega_l, l=1,2$ belonging to NN of FO $Z_j = Z\left(s_j\right) \in NN_0^l$, $s_j \neq s_0$, $l=1,2$. The natural parameter values A_{0h}^{l} specified in (1.32) get the expressions:

$$A_{01}^{l} = \beta_l^{(11)} x_1\left(s_0\right) + \beta_l^{(12)} x_2\left(s_0\right) - \sum_{\substack{Z_j \in NN_0^l \\ Z_j = 0,\, j \neq 0}} \eta \ln\left(1 - Z_j\right),$$

$$A_{02}^{l} = \beta_l^{(21)} x_1\left(s_0\right) + \beta_l^{(22)} x_2\left(s_0\right) - \sum_{\substack{Z_j \in NN_0^l \\ Z_j = 0,\, j \neq 0}} \eta \ln\left(Z_j\right),$$

$$A_{03}^{l} = \left[\log\left(\frac{c_{0z}^{l}}{1 - c_{0z}^{l}} \right) + Be\left(A_{01}^{l} + 1, A_{02}^{l} + 1\right) \right]$$

and also mixture parameter is specified as

$$c_{0z}^{l} = \frac{\exp\left\{ \lambda^l + \sum\limits_{Z_j \in NN_0^l,\, j \neq 0} \rho^l B_3\left(Z_j\right) \right\}}{\left(1 + \exp\left\{ \lambda^l + \sum\limits_{Z_j \in NN_0^l,\, j \neq 0} \rho^l B_3\left(Z_j\right) \right\} \right)},$$

where $x_1\left(s_0\right), x_2\left(s_0\right)$ are the coordinates of the spatial location for Z_0 the class $l, l=1,2$, $Z_j = Z\left(s_j\right)$ is a training sample feature value that belongs to

the NN set of FO Z_0, $Z_j \in NN_0$, $j \neq 0$, $B_3\left(Z_j\right) = I\left(Z_j = 0\right)$. In this case, a set of parameters $\Psi = \left\{\beta_{hq}^l, \eta, \lambda^l, \rho^l, h, l, q = 1,2\right\}$.

The NN areas are selected freely to analyse several situations of interest and areas are defined at a selectable distance from the FO. Distances have been chosen at 1 km, 3 km, and 7 km. While analysing the data, the Euclidean distance is calculated, and the nearest neighbour areas are shown in Figure 3.21.

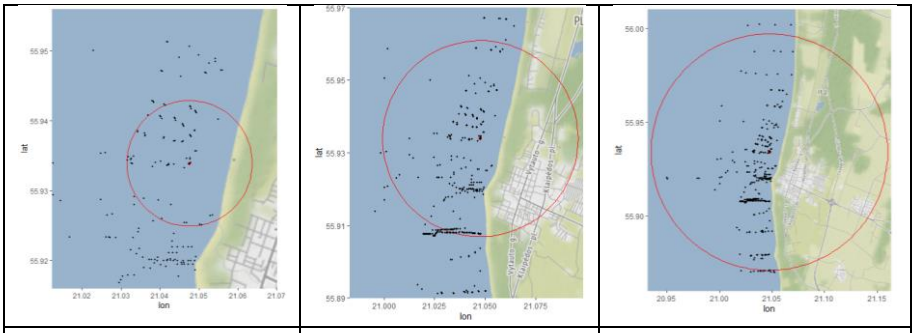


Figure 3.21 Nearest neighbour areas with different distances for model parameter estimation.

Compile three zero-inflated auto-beta models and different NN areas of FOs for black carrageen concentration modelling. Population parameters estimates are obtained using the MPL method for the training sample in each model. The estimates are provided in Table 3.9. A more significant difference between parameter values that differ by neighbour area can be seen between η parameters that describe the influence of neighbouring points to the classification point and between parameters that describe mixture parameters c_{0z}^l. These parameter estimates are used in the PBDF $W\left(Z_0, \hat{\Psi}\right)$ specified in (2.10). They will be used to solve the supervised classification problem, i.e. identifying the bottom of the Baltic sea.

The task is to classify Z_0 into one of two populations. The classification is performed using Algorithm 1. In Algorithm 1 using PBDF $W^B\left(Z_0, \hat{\Psi}\right)$ (2.8) depends on the estimates of the population parameters as well as on the probabilities of the class labels. In the considered case class label probabilities are evaluated in three different ways, the same as in the SABE model.

Table 3.9 Estimates of parameters.

Nearest neighbour areas				
Parameters	1 km	3 km	7 km	all training set
$\beta_1^{(11)}$	-0.998	-0.990	-0.987	-0.970
$\beta_1^{(12)}$	0.084	0.083	0.083	0.080
$\beta_1^{(21)}$	-0.974	-0.974	-0.979	-0.994
$\beta_1^{(22)}$	0.090	0.091	0.088	0.088
$\beta_1^{(11)}$	-0.987	-0.998	-0.982	-0.994
$\beta_2^{(12)}$	0.086	0.086	0.084	0.084
$\beta_2^{(21)}$	-0.971	-0.993	-0.988	-0.992
$\beta_2^{(22)}$	0.096	0.096	0.091	0.087
η	-0.035	-0.002	-0.012	-0.012
λ^1	-0.057	-0.792	-0.012	-0.108
ρ^1	0.079	0.039	0.010	0.005
λ^2	-0.123	-0.099	-0.838	-0.996
ρ^2	0.184	0.041	0.045	0.026

The values for the testing sample, when AER estimates are calculated according to the BDF sign, with different class label probability calculation methods, and zero-inflated auto-beta models with different NN areas for feature values, are presented in Table 3.10.

Table 3.10 $\overline{EER}$ values for zero-inflated auto-beta models.

Class label probability	Nearest neighbour areas			
	1 km	3 km	7 km	all sample
Sample size	**0.2435**	0.2642	0.2487	**0.2435**
Inv. dist. max min	**0.2021**	0.2383	0.2228	0.2176
Inv. dist. all training sample	**0.2228**	0.2487	0.2435	0.2383

Classification error rates can be compared between neighbour areas using the results presented in Table 3.10. When class label probabilities are calculated using the inverse distance function with maximum minimal distance and the whole training sample, the classification error rates are smaller when compared to the situation at different NN areas, and are the lowest for the 1 km NN area. When class probabilities are calculated using the training sample size, the lowest classification error rates are achieved using the NN area with 1 km and the whole training sample. A study of calculating class label probabilities using contextual information has shown that

classification is better when taking a portion of the nearest neighbourhood spatial location information. The lowest error rate is achieved using 1 km as the distance for the nearest neighbour area and maximum minimal distance as the class label probability inverse distance function. From here it can be concluded that the NN area for class label probability calculation influences classification errors in the studied cases.

3.4. Conclusions of the Section

This section presents the results of numerical experiments with simulated data and the application to the real data. At first, a classification problem using a classification rule based on GRF is analysed in a binary case. The comparison of two approaches for parameter estimation is made based on the values of the $\overline{EER}$ incurred by the classification rule based on the PBDF. The results give us strong arguments to expect that BA estimators of spatial population parameters can be effectively used in spatial Gaussian data classification incurred by PBDF.

Also, a classification problem is analysed in a multi-class case using GRF models. The simulation experiment shows the advantage of the STLG approach over the STLM. This advantage is more significant for strongly separated populations (larger values of b) than for close populations. The results give us strong arguments to expect that the STLG structure of spatial population could be effectively used in spatial Gaussian data classification incurred by PBDF.

TRF are analysed, comparison of the two PBDF based ML and LS estimators of the mean parameters is performed on the simulated values of the empirical estimators of the ER. Hence the results give us strong arguments that often intractable ML estimators of spatial mean parameters should be used in highly correlated spatial data modelled by TRF, and the simpler LS estimators can replace these estimators for weakly correlated spatial data without significant loss of the PBDF performance.

A simulation data study is conducted to estimate and empirically compare the BDF classifier with the LDF classifier for various parametric structures and class label probability models. While considering the situations with different prior probabilities, better results are achieved by including fourth-order neighbours in calculating prior probabilities in cases when prior probabilities are equal and when prior probabilities include all training points.

Real data analysis is performed to evaluate and empirically compare the BDF classifier with LDF for different nearest neighbour sets of the classified point for the SABE model and prior class probabilities models. Better results

are obtained with minimal nearest neighbour area or nearest neighbour area with all training set. Comparing different prior probability situations shows that prior probability calculation using inverse distance function reduces the classification error rate. Also, calculating prior probabilities using the inverse distance function for a selected number of neighbours reduced the classification error rate in contrast to using all training set points.

GENERAL CONCLUSIONS

This work extends the application of supervised generative classification methods for features with distributions of spatial contextual information described by GRF and non-GRF models. An algorithm based on BDF has been developed, which allows using supervised generative classification models for feature values with distribution belonging to the exponential (Poisson, binomial, gamma, beta, ZIB) and elliptical (t-distribution) families. The conclusions obtained during the conducted research are presented below.

1. A comparison of EER estimates, made to expand the use of supervised generative models for the classification of GRF observations and to investigate their effectiveness, revealed the superiority of the BA for populations parameters estimation approach over the ML approach when GRF observations are classified in the case of 2 classes. This advantage is more significant for strongly separated populations than for close populations, for the symmetric TLC as well as for the asymmetric one. In the asymmetric TLC, the ratio $\overline{EER}^{ML} / \overline{EER}^{BA}$ values monotonically increase from 1.0221 to 1.3938, when the distance between populations increases from 0.5 to 1.9. The results give strong arguments to expect that BA estimators of spatial population parameters could be effectively used in spatial Gaussian data classification incurred by PBDF.
2. Extending the application of supervised generative models to the classification of GRF observations in the case of 3 classes and investigating their effectiveness by comparing the EER estimates showed the advantage of STLG over STLM. This advantage is more significant for strongly separated populations than for close populations in the studied situations with covariance functions differing in the range parameter ϕ, which takes the values 1, 2, or 3. When $\phi = 3$, the ratio $\overline{EER}^{STLG} / \overline{EER}^{STLM}$ values decrease from 0.7618 to 0.0029, with increasing the distance between populations. The results give strong arguments to expect that the STLG structure of spatial population could be effectively used in the GRF observation classification incurred by PBDF.
3. In order to expand the use of supervised generative models for the classification of TRF observations in the case of 2 classes and to investigate their efficiency, the comparison of EER estimates obtained during the application of the constructed model realization

algorithm revealed the advantage of the ML population parameter estimation method over the LS method. The increase of advantage is recorded with increasing the distance between populations, as well as when the range parameter ϕ of the covariance function increases. When $\Delta = 1.5$, the ratio $\overline{EER}^{ML} / \overline{EER}^{LS}$ values decrease from 1 to 0.8159, when the range parameter ϕ of the covariance function grows from 0.1 to 3.3. The results give strong arguments that often intractable ML estimators of mean parameters for spatial data should be used in highly correlated spatial data modelled by TRF, and the simpler LS estimators can replace these estimators for weakly correlated spatial data without significant loss of the PBDF performance.

4. Setting side by side the EER estimates obtained during the application of the constructed model realization algorithm to expand the use of supervised generative models for the classification of features, whose spatial distribution is described by the auto-Poisson, the auto-binomial, the auto-gamma or the auto-beta model in the case of 2 classes and investigating their effectiveness, revealed that the model realization algorithm based on BDF has the advantage over the algorithm based on the modified LDF when the feature distribution is described by the auto-beta model. This difference is statistically significant when the classes are more separated. The ratio $\overline{EER}^{BDF} / \overline{EER}^{LDF}$ values decrease from 1.0018 to 0.0262, where the difference between the regression parameters for different classes increases from 0.5 to 49. The results provide arguments that allow us to expect that the proposed algorithm based on BDF can be effectively used in the classification of features whose spatial distribution is described by the auto-beta model.
5. In order to evaluate the effect of including spatial information during the calculation of class label probabilities, the comparison of EER estimates revealed the advantage of the method where the class label probabilities are calculated using the inverse distance function with a defined NN area (Method 3) over the method where the probabilities are considered equal (Method 1). The ratio $\overline{EER}^{3rd} / \overline{EER}^{1st}$ values decrease from 0,8608 to 0,6766. It is also confirmed by the study of the bottom coverage data of black carrageenan algae. After constructing the zero-inflated auto-beta model, the EER estimates are

the lowest when neighbouring feature values within 1 km are included in the FO distribution model. When the class label probabilities are calculated using the inverse distance function with a defined NN area, the EER estimate is 0.2021, and when the class label probabilities are calculated using the training sample size, it is 0.2435. The results provide an argument that including prior information related to FO and training sample locations in the models reduces the classification error.

BIBLIOGRAPHY

1. Anderson, T. W. 2003. An Introduction to Multivariate Statistical Analysis. Wiley, New York.
2. Andrews J. L., McNicholas P. D., Subedi S. 2011. Model-based classification via mixtures of multivariate *t*-distributions. Computational Statistics and Data Analysis, vol. 55(1), p. 520–529.
3. Arnold B. C., Castillo E., Sarabia J. M. 2001. Conditionally Specified Distributions: An Introduction. Statistical Science. Vol. 16, No. 3, 249-274.
4. Atkinson, P.M., Lewis, P. 2000. Geostatistical classification for remote sensing: an introduction. Computers and geosciences. 26, 361-371.
5. Baltmiškytė E., Dučinskas K. 2013. Baltijos jūros analizė naudojant Bajesinės erdvinės statistikos metodus. Computational science and techniques. Special issue for Research Innovations Fundamentals. Vol. 1, no. 1, p. 1-10.
6. Banerjee S., Gelfand A. E. 2003. On smoothness properties of spatial processes. Journal of Multivariate Analysis vol. 84 (1), 85-100. https://doi.org/10.1016/S0047-259X(02)00016-7.
7. Bånkestad M., Sjölund J., Taghia J., Schön T. 2020. The Elliptical Processes: a Family of Fat-tailed Stochastic Processes.
8. Batsidis A., Zografos K. 2011. Errors of misclassification in discrimination of dimensional coherent elliptic random field observations. Statistica Neerlandica, vol. 65, p. 446–461.
9. Batsidis A. 2012. Errors of misclassification in discrimination with data from truncated t populations. Stat Papers 53, 281–298.
10. Berret C., Calder, C. A. 2016. Bayesian spatial binary classification. Spatial Statistics, vol. 16, p. 72–102.
11. Besag J. 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society, vol. 36, p. 192-236.
12. Besag, J. 1975. Statistical analysis of non-lattice data. The Statistician 24, 179–95.
13. Boente G., Barrera M. S. Tyer E. T. 2014 A characterization of elliptical distributions and some optimality properties of principal components for functional data. Journal of Multivariate Analysis 131, 254–264.
14. Breslow N. E., Clayton D. G. 1993. Approximate inference in generalized linear mixed models. J Am Stat Assoc, 88:9-25.
15. Bishop Ch. M, Lasserre J. 2007. Generative or discriminative? Getting the best of both worlds. Bayesian statistics, 8(3):3–24.

16. Chan S., Chu J., Zhang Y., Nadarajah S. 2021. Count regression models for COVID-19. Physica A: Statistical Mechanics and its Applications,Vol.563, 125460, ISSN 0378-4371. https://doi.org/10.1016/j.physa.2020.125460.
17. Chiles, J. P., Delfiner, P. 2012. Geostatistics: Modeling Spatial Uncertainty. Second Edition. John Wiley & Sons, New York.
18. Cepeda, E.D., Nunez, A.V. 2013. Spatial Double Generalized beta Regression Models Extensions and Application to Study Quality of Education in Colombia. Journal of Educational and Behavioral Statistics, 38, 604–628.
19. Clark J. S., 2005. Why environmental scientists are becoming Bayesian. Ecology Letters, 8, 2–14. https://doi.org/10.1111/j.1461-0248.2004.00702.x.
20. Clark J. S., Gelfand A. E. 2006. Hierarchical Modelling for Environmental Sciences. New York: Oxford University Press.
21. Conde D, Fernández MA, Salvador B. 2005. A classification rule for ordered exponential populations. J Statist Plan Inference.;135(2):339–356.
22. Cook D.O., Kieschnick R., McCullough B.D. 2008. Regression analysis of proportions in finance with self selection. Journal of Empirical Finance Vol. 15, Issue 5, p. 860-867. https://doi.org/10.1016/j.jempfin.2008.02.001.
23. Cressie N. A. C. 1993. Statistics for spatial data. New York: John Wiley & Sons.
24. Diggle, P. J., Ribeiro, P. J. 2007. Model-based Geostatistics. Springer.
25. Diggle, P.J., Tawn, J., Moyeed, R. 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C. Appl. Stat. 47, 299–350.
26. Ding P. 2016. On the Conditional Distribution of the Multivariate t Distribution, The American Statistician, 70:3, 293-295. https://doi.org/10.1080/00031305.2016.1164756.
27. Dreižienė L. 2019. Classification risk of Gaussian spatial data using linear discriminant functions (Doctoral dissertation) Vilnius University.
28. Dreižienė L., Dučinskas. K. 2020. Comparison of spatial linear mixed models for ecological data based on the correct classification rates. Spatial Statistics 35, 100395.
29. Dreižienė, L., Dučinskas, K., Saltyte-Vaisiauske, L. 2018. Statistical classification of multivariate conditionally autoregressive Gaussian random field observations. Spat. Stat., 28, 216–225.
30. Duda R. O., Hart P. E., Stork D.G. 2001. Pattern Classification, 2nd edition. Wiley.
31. Dučinskas, K. 1995. Optimal training sample allocation and asymptotic expressions for error rates in discriminant analysis. Acta Applicandae Mathematicae, 38, p. 3-11.

32. Dučinskas, K. 1997. An asymptotic analysis of the regret risk in discriminant analysis under various training schemes. Lithuanian Mathematical Journal, 37(4):337-351.
33. Dučinskas, K. 1998. Analysis of the risk regret for classification of gamma populations. Informatica, 9(4):401-414.
34. Dučinskas K. 2009. Approximation of the expected error rate in classification of the Gaussian random field observations. Statistics and Probability Letters, vol. 79, p. 138-144.
35. Dučinskas K., Baltmiškytė E., Bučas M. 2012. Diskrečių skirstinių taikymas duomenų apie gamtos plotų padengimą Bajesinėje analizėje. Lietuvos matematikos rinkinys: Lietuvos matematikų draugijos darbai. ISSN 0132-2818 T. 53, ser. B, p. 51-55.
36. Dučinskas K., Dreižienė L. 2011. Supervised classification of the scalar Gaussian random field observations under a deterministic spatial sampling design. Austrian Journal of Statistics, vol. 40 (1-2), p. 25-36.
37. Dučinskas, K., Dreižienė, L. 2018. Risks of classification of the gaussian markov random field observations. J. Classification, 35, 422–436.
38. Dučinskas, K., Dreižienė, L. 2021a. Actual error rates in linear discrimination of spatial Gaussian data in terms of semivariograms. Commun. Stat. - Theory Methods, 98 (published online). https://doi.org/10.1080/03610926.2021.1968903.
39. Dučinskas K., Dreižienė L. 2021b. On the classification error rates in terms of semivariograms for Gaussian universal kriging models. Lithuanian Journal of Statistics Lietuvos statistikos darbai. Vol. 60, 1–7. https://doi.org/10.15388/LJS.2021.26442.
40. Dučinskas, K., Stabingienė, L. 2011. Expected Bayes error rate in supervised classification of spatial Gaussian data. Informatica, 22(3), 371–381.
41. Dučinskas K., Šaltytė-Benth J. 2003. Erdvinė skatistika. KU, Klaipėda.
42. Fang K. T., Kotz S., Ng K. W. 1990. Symmetric multivariate and related distributions. Chapman and Hall.
43. Fauvel M., Tarabalka Y., Benediktsson J.A., Chanussot J., Tilton J.C. (2013) Advances in Spectral-Spatial Classification of Hyperspectral Images. Proceedings of the IEEE, 101:652-675. https://doi.org/10.1109/JPROC.2012.2197589.
44. Ferrari, S.L.P., Cribari-Neto, F. 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31, 799–815.
45. Fitzmaurice G., Laird N., Ware J. 2004. Applied Longitudinal Analysis. Wiley and Sons, Hoboken, New Jersey, USA.
46. Gelman A., Carlin J. B., Stern H. S. 2003. Bayesian data analysis, 2nd edition Boca Raton: CRC Press.

47. Gelfand A. E., Schliep E. M., 2016. Spatial statistics and Gaussian processes: A beautiful marriage. Spatial Statistics, 18, 86–104. https://doi.org/10.1016/j.spasta.2016.03.006.
48. Guo X., Du Y., Zhao L. 2021. Deep Generative Models for Spatial Networks. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021.
49. Hardouin, C., Yao, J.F. 2008. Multi-parameter auto-models and their application. Biometrika, 95 (2), 335–349.
50. Jackson, Q., Landgrebe, D. A. 2002. Adaptive Bayesian Contextual classification based on Markov random fields. IEEE Transactions on Geoscience and Remote Sensing. 40 (11), 2454-2463.
51. Jana, N., Kumar, S., Misra, N. 2014. Classification rules for exponential populations under order restrictions on parameters. Springer Proceedings in Mathematics & Statistics, 91, 243–250.
52. Jana N., Kumar S. 2016. Classification into Two-Parameter Exponential Populations with a Common Guarantee Time, American Journal of Mathematical and Management Sciences, 35:1, 36-54, https://doi.org/10.1080/01966324.2015.1074132.
53. Jana N., Kumar S., Misra N. 2016. Classification rules for two parameter exponential populations under order restrictions on parameters. Journal of Statistical Computation and Simulation, Vol. 86, No. 8, 1559–1581, https://doi.org/10.1080/00949655.2015.1075540.
54. Jensen, J. L., Moller, J. 1991. Pseudolikelihood for exponential family models of spatial processes, Ann. Appl. Probab., 1, 445-461.
55. Kaiser, M.S., Cressie, N. 2000. The construction of multivariate distributions from Markov random fields. J. Multivariate Anal., 73, 199–220.
56. Kaiser, M.S., Cressie, N., Lee, J. 2002. Spatial mixture models based on exponential family conditional distributions. Statistica Sinica, 12, 449–474.
57. Kalhory, L., Mahammadzadeh, M. 2018. Bayesian inference for Spatial Beta Generalized Linear Mixed Models. Journal of Sciences, Islamic Republic of Iran, 292, 173–186.
58. Kauermann G., Ormerod J. T., Wand M. P. 2010. Parsimonious Classification via Generalized Linear Mixed Models. Journal of Classification, vol 27(1), p. 89-110.
59. Khedama R., Belhadj-Aissa A. 2004. Contextual classification of remotely sensed data using MAP approach and MRF. ISPRS Journal of Photogrammetry and Remote Sensing 35 (7), 11-16.

60. Kim H. M., Mallick B. K. 2003. A note on Bayesian spatial prediction using the elliptical distribution. Statistics and Probability Letters, vol. 64, p. 271–276.
61. Kinderman R., Snell J. L. 1980. Markov random fields and their applications. Amer. Math. Soc., vol. 1, pp. 1–142.
62. King R., Morgan B. J. T., Gimenez O., Brooks S. P. 2010. Bayesian Analysis for Population Ecology. Boca Raton: CRC Press.
63. Klein, R., Press, S.J., 1992. Adaptive Bayesian classification of spatial data. Journal of the American Statistical Association. 87, 844–851.
64. Kleinman K., Lazarus R., Platt R. 2004. A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism. American Journal of Epidemiology, vol. 159(3), p. 217-224.
65. Lagos-Álvarez, B.M., Fustos-Toribio, R., Figueroa-Zúñiga, J., Mateu, J. 2017. Geostatistical mixed beta regression: a Bayesian approach. Stoch Environ Res Risk Assess, 31, 571–584.
66. Lee J., Kaiser M.S., Cressie N., 2001. Multiway dependence in exponential family conditional distributions. Journal of Multivariate Analysis, 79, 171-190.
67. Le N. D., Zidek J. V. 2006. Statistical Analysis of Environmental Space-Time Processes. New York: Springer Science Business Media.
68. Li M., Zang Sh., Zhang B, Li Sh., Wu Ch. 2014. A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information, European Journal of Remote Sensing, 47:1, 389-411, https://doi.org/10.5721/EuJRS20144723.
69. Liu F., Eugenio E. 2016. A review and comparison of Bayesian and likelihood-based inferences in beta regression and zero-or-one-inflated beta regression. Statistical Methods in Medical Research. Vol. 27 issue 4 p. 1025-1044.
70. Mackowiak R., Ardizzone L., Kothe U., Rother C. 2021. Generative Classifiers as a Basis for Trustworthy Image Classification. CVPR (inproceedings).
71. Mardia K. V. 1984. Spatial discrimination and classification maps. Communications in Statistics - Theory and Methods, vol. 13(18), p. 2181–2197.
72. Mardia K. V., Kent J. T., Bibby J. M. 1979. Multivariate Analysis. London: Academic Press.
73. Mase S. 1995. Consistency of the maximum pseudo-likelihood estimator of continuous state space Gibbsian processes. The Annals of Applied Probability, 5(3):603–612.

74. McCullagh P, Nelder J. A. 1989. Generalized Linear Models London: Chapman and Hall.
75. McLachlan G. J. 2004. Discriminant analysis and statistical pattern recognition. New York: Wiley.
76. Mohebbi M., Wolfe R., Jolley D. 2011. A Poisson regression approach for modelling spatial autocorrelation between geographically referenced observations. Mohebbi et al. BMC Medical Research Methodology 2011, 11:133
77. Moller, J. 1999. Perfect Simulation of Conditionally Specified Models. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 61 (1), 251–264.
78. Moubayed A., Injadat M., Nassif A., Lutfiyya H., Shami A. 2018. E-Learning: Challenges and Research Opportunities Using Machine Learning and Data Analytics. IEEE Access, 6, 39117-39138. https://doi.org/10.1109/ACCESS.2018.2851790.
79. Murphy K. P. 2012. Machine Learning. A probabilistic Perspective. England: MIT Press.
80. Nadarajah S., Kotz S. 2007. Multitude of beta distributions with applications. Statistics, vol. 41 (2), 153-179.
81. Nadarajah S., Kotz S. 2008. Estimation methods for the multivariate t distribution. Acta Applicandae Mathematicae: An International Survey Journal on Applyng Mathematics and Mathematical Applications, vol. 102(1), p. 99-118.
82. Nishii, R., Eguchi, S. 2006. Image classification based on Markov random field model with Jeffreys divergence. Journal of Multivariate Analysis. 97 (9), 1997-2008.
83. Osei F.B., SteinA., Andreo,V. 2022. A zero-inflated mixture spatially varying coefficient modeling of cholera incidences. Spatial Statistics Vol. 48, 100635. https://doi.org/10.1016/j.spasta.2022.100635.
84. Ospina, R., Ferrari, S.L.P. 2010. Inflated beta distributions. Stat. Papers. 51: 111–126.
85. Ospina, R., and Ferrari, S.L.P. 2012. A general class of zero-or-one inated beta regression models. Comput. Stat. Data Anal. 56, 1609–1623.
86. Paradinas I., Pennino M. G., Marín González M., López-Quílez A., Bellido J., Conesa D. 2018. Modelling spatially sampled proportion processes. Revstat - Statistical Journal 16 (1) p. 71-86.
87. Peng X., Li G., Liu Z. 2016. Zero-Inflated Beta regression for Differential Abundance Analysis with Metagenomics Data. Jurnal of computational biology. vol. 23 (2), 102-110.

88. Pfeiffer D. U, Robinson T. P, Stevenson M., Stevens K. B., Rogers D. J, Clements A. C. A 2008. Spatial Variation in Risk. Spatial Analysis in Epidemiology. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198509882.001.0001.
89. Ridout M., Hinde J., Demétrio C.G. 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. Biometrics. 57 (1): 219-23. https://doi.org/10.1111/j.0006-341x.2001.00219.x. PMID: 11252601.
90. Roislien, J., Omre, H. 2006. T-distributed random fields: A parametric model for heavy-tailed well-log data. Math. Geology, 38(7), 821–849.
91. Sengupta A., Cressie N. 2013. Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. Spatial Statistics 4, 14–44.
92. Shah A., Wilson, A., and Ghahramani Z., 2014. Student-*t* processes as alternatives to Gaussian processes. In Artificial Intelligence and Statistics, pages 877–885.
93. Shekhar S., Schrater P.R., Vatsavai R.R., Wu W.L., Chawla S. 2002. Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. IEEE Transactions on Multimedia, 4, 174-188.
94. Solberg A., Taxt T., Jain A.K. 1996 A Markov random field model for classification of multisource satellite imagery. IEEE Trans. Geosci. Remote. Sens. 34 100-113.
95. Song X., Abu-Mostafa Y., Sill J., Kasdan H. 1997. Incorporating contextual information in white blood cell identification. Advances in Neural Information Processing Systems 10 - Proceedings of the 1997 Conference, NIPS 1998. 950-956.
96. Stabingienė L. 2012. Vaizdų analizė naudojant Bajeso diskriminantines dunkcija. (Doctoral dissertation) Vilnius University.
97. Stabingienė L., Dučinskas K. 2010. Error rates in spatial classification of Gaussian data with random labeling. Lietuvos matematikos rinkinys. LMD darbai, vol. 51, p. 426-430. ISSN 0132-2818.
98. Stabingienė L., Stabingis G., Dučinskas K. 2010. Comparison of linear discriminant functions in image classification. Lietuvos matematikos rinkinys. LMD darbai, vol. 51, p. 227-231. ISSN 0132-2818.
99. Stabingis G. 2019. Statistiniai sprendimai erdvinei informacijai skaitmeniniuose vaizduose. (Doctoral dissertation) Vilnius University.
100. Stabingis G., Dučinskas K., Stabingienė L. 2014. Comparison of spatial classification rules with different conditional distributions of class label. Nonlinear Analysis: Modelling and Control, 2014, Vol. 19, No. 1, 109–117.

101. Sutradhar B. C., Ali M. M. 1986. Estimation of the parameters of a regression model with a multivariate t error variable, Communications in Statistics-Theory and Methods, vol. 15, p. 429–450.
102. Sun Sh., Zhong P., Xiao H., Wang R., 2016. Spatial contextual classification of remote sensing images using a Gaussian process. Remote sensing letters 7(2): 131-140. https://doi.org/10.1080/2150704X.2015.1117152.
103. Switzer P. 1980. Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. Mathematical Geology, vol. 12(4), p. 367-376.
104. Šaltytė, J. 2001. The asymptotic expansion of the expected risk for the LDA of spatially correlated Gaussian observations. Klaipėda University. (Doctoral dissertation).
105. Šaltytė J., Dučinskas K. 2002. Comparison of ML and OLS estimators in discriminant analysis of spatially correlated observations. Informatica, vol. 13(2), p. 297-238.
106. Šaltytė-Benth J., Dučinskas K. 2005. Linear discriminant analysis of multivariate spatial-temporal regressions. Scandinavian Journal of Statistics, vol. 32, p. 281-294.
107. Theodoritis S., Kountroumbas K. 2009. Pattern recognition. Academic Press.
108. Thompson G. Z., Maitra R., Meeker W. Q. 2020. Classification With the Matrix-Variate-t Distribution. Journal of Computational and Graphical Statistics 29 (3) 668-674. https://doi.org/10.1080/10618600.2019.1696208.
109. Tu, W. 2002. Zero-inflated data. Encycl. Environ. 4, 2387-2391.
110. Wakaki H. 1998. Asymptotic Expansions for the Studentized Best Linear Discriminant Function Under Two Elliptical Populations, American Journal of Mathematical and Management Sciences, 18:1-2, 89-108.
111. Wang Z., Ritou M., da Cunha C., Furet B. 2020. Contextual classification for smart machining based on unsupervised machine learning by Gaussian mixture model. International Journal of Computer Integrated Manufacturing, vol. 33 (10-11), p. 1042- 1054. https://doi.org/10.1080/0951192X.2020.1775302.
112. Warton DI, Foster SD, De'ath G, Stoklosa J, Dunstan PK. 2015. Model-based thinking for community ecology. Plant Ecology 216(5):669–682. https://doi.org/10.1007/s11258-014-0366-3.
113. Zhang, H. 2002. On Estimation and Prediction for Spatial Generalized Linear Mixed Models. Biometrics, 58 (1), 129–136.

SUMMARY

ĮVADAS

Mašininis mokymasis yra dirbtinio intelekto sritis, esminė augančios duomenų mokslo srities sudedamoji dalis, kuri remiasi idėja, kad sistemos gali mokytis iš duomenų, nustatyti modelius ir priimti sprendimus. Taikant statistinius metodus, algoritmai mokomi klasifikuoti arba prognozuoti, atskleidžiant svarbias duomenų savybes.

Mašininio mokymosi algoritmai su visiškai pažymėtais duomenimis priskiriami prižiūrimojo mokymosi kategorijai. Prižiūrimojo klasifikavimo uždaviniui spręsti reikalingi mašininio mokymosi algoritmai, kurie išmoksta nustatyti naujų stebėjimų klasę pagal mokymo duomenis. Du pagrindiniai prižiūrimojo mokymosi modeliai: generatyviniai (angl. *generative*) ir diskriminatyviniai (angl. *discriminative*) (Bishop ir Lasserre, 2007). Šiame darbe pagrindinis dėmesys skiriamas prižiūrimiesiems generatyviniams modeliams. Matematiškai generatyviniai modeliai skirti bendro tikimybinio skirstinio (angl. *joint probability distribution*) $P(Z,Y)$ nustatymui; čia *Z* žymi požymio reikšmių vektorių, o *Y* žymi klasių žymių vektorių. Naudojant Bajeso taisyklę (angl. *Bayes Rule*, *BR*), požymio reikšmių *Z* klasės sąlyginį tikimybinį skirstinį (angl. *class-conditional probability distribution*) $P(Z|Y)$ ir klasės žymių *Y* tikimybinį skirstinį $P(Y)$, gaunamas sąlyginis tikimybinis skirstinys $P(Y|Z)$, kitaip – aposteriorinis klasių žymių skirstinys (angl. *posterior class distribution*) (Duda R.O. ir kt., 2001).

$P(Z|Y)$ išraiška priklauso nuo tikimybinio skirstinio pasirinkimo, kuris aprašo *Z* su žinomu *Y*. Pasirinkimas atliekamas atsižvelgiant į *Z* reikšmių savybes: tolydžios ar diskrečios reikšmės, begalinis arba ribotas įgyjamų reikšmių intervalas. Šiame darbe nagrinėjami atvejai, kai *Z* skirstinys priklauso eliptinių arba eksponentinių skirstinių šeimoms. Šios šeimos plačiau pristatomos disertacijos 1 skyriuje. $P(Z|Y)$ išraiška taip pat priklauso nuo to, ar *Z* reikšmės laikomos nepriklausomomis, ar turinčiomis statistinę priklausomybę, kurią galima apibrėžti naudojant kovariacines funkcijas, variogramas arba erdvinės priklausomybės parametrus apibrėžiančius ryšius tarp klasifikuojamo taško ir jo kaimynų. $P(Y)$ atspindi tyrėjo apriorines žinias apie klasių žymes *Y*. Įvairios $P(Z|Y)$ ir $P(Y)$ išraiškos leidžia išplėsti

prižiūrimųjų generatyvinių modelių taikymą įvairioms klasifikavimo problemoms spręsti.

Problema ir jos aktualumas

Paprastai sprendžiant klasifikavimo problemas naudojant prižiūrimuosius generatyvinius modelius, vadovaujamasi stebėjimų nepriklausomumo prielaida (Jana ir Kumar, 2016). Tačiau erdvinė informacija atlieka esminį vaidmenį analizuojant ir suprantant įvairias mokslo sritis, pavyzdžiui, ekologinius, biologinius procesus (gamtos mokslus), biomediciną, inžineriją ir socialinius mokslus. Erdviniame klasifikavime reikia atsižvelgti į informaciją apie duomenų (stebinių) padėtį erdvėje. Šiame darbe erdvinės kontekstinės informacijos sąvoka naudojama apibrėžti erdvinės informacijos įtraukimą į klasifikatoriaus struktūrą. Erdvinio konteksto sąvoka, paprastai yra naudojama vaizdų klasifikavimui, nurodant ryšį tarp klasifikuojamo pikselio ir jam gretimų pikselių, kurie įtraukiami į modelius. Detalus erdvinio kontekstinio klasifikavimo metodų, skirtų nuotolinio stebėjimo vaizdų klasifikavimui, palyginimas pateikiamas Li ir kt. (2014). Kontekstinio klasifikavimo modeliai, kuriuose naudojama erdvinė informacija kiekybiškai įvertinant erdvinius ryšius, gali būti naudojami vaizdų klasifikavimui ir objektų aptikimui (Sun ir kt., 2016; Stabingis, 2019). Šiame darbe kovariacinės funkcijos ir auto-modeliai, kuriuos pasiūlė Besag (1974), naudojami erdvinei informacijai apibrėžti, kuri susijusi su ryšiu tarp klasifikuojamo erdvės taško ir kaimyninių erdvės taškų. Erdvinės kontekstinės informacijos įtraukimas į prižiūrimojo generatyvinio klasifikavimo algoritmų konstrukciją gali būti atliekamas modeliuojant požymio reikšmių pasiskirstymą ir (arba) klasės žymių pasiskirstymą. Prižiūrimųjų generatyvinių modelių su erdvine kontekstine informacija naudojimas klasifikavimo uždaviniams spręsti šiame darbe vadinamas erdviniu kontekstiniu klasifikavimu.

Kita klasifikavimo uždavinių prielaida yra ta, kad duomenys turi atitikti Gauso skirstinį (Wang ir kt., 2020, Dreižienė ir Dučinskas, 2021a). Šis reikalavimas retai įgyvendinamas atliekant realią duomenų analizę. Dėl šios priežasties tyrimas turi atlikti tam tikras duomenų transformacijas. Pavyzdžiui, Box-Cox laipsninė, logit ir arcsinus transformacijos tiriamus duomenis transformuoja taip, kad jie kuo geriau atitiktų Gauso atsitiktinio lauko (angl. *Gaussian Random Field, GRF*) savybes. Transformuotus duomenų modelius yra sunkiau interpretuoti ir nebūtinai jie gerai atitinka GRF savybes.

Norint išspręsti klasifikavimo uždavinius stebiniams, kurie aprašomi ne GRF ir panaudojant erdvinę kontekstinę informaciją, yra reikalingi nauji klasifikavimo metodai, kurie praplėstų prižiūrimojo generatyvinio klasifikavimo algoritmų taikymą.

Tyrimo objektas

Prižiūrimojo generatyvinio klasifikavimo metodai (angl. *supervised generative classification approach*), Bajeso diskriminantinė funkcija, (angl. *Bayes Discriminant Function, BDF*), eliptinių ir eksponentinių šeimų sąlyginiai skirstiniai (ang. *elliptical and exponential families conditional distributions*).

Darbo tikslas ir uždaviniai

Tyrimo tikslas – sukonstruoti prižiūrimojo generatyvinio klasifikavimo algoritmus, paremtus BDF ir skirtus požymiams su erdvinės kontekstinės informacijos skirstiniais, priklausančiais eksponentinių ir eliptinių skirstinių šeimoms.

Siekiant numatyto tikslo buvo suformuluoti šie uždaviniai:

1. praplėsti prižiūrimųjų generatyvinių klasifikatorių, paremtų BDF, panaudojimą požymiams, aprašomiems GRF;
2. sukonstruoti prižiūrimojo generatyvinio klasifikavimo algoritmą, paremtą BDF, požymiams, aprašomiems T-skirstinio atsitiktiniu lauku (ang. *T-distributed Random Field, TRF*);
3. sukonstruoti prižiūrimojo generatyvinio klasifikavimo algoritmą, paremtą BDF, požymiams su erdvinės kontekstinės informacijos skirstiniais, priklausančiais eksponentinei skirstinių šeimai.

Tyrimo metodika

Literatūros apžvalga buvo atlikta siekiant įvertinti metodus, naudojamus požymio reikšmių modeliavimui su statistine erdvine kontekstine informacija, bei įvertinti generatyvinius modelius, naudojamus išspręsti prižiūrimojo klasifikavimo uždavinius duomenų modeliams su erdvine kontekstine informacija. Darbe aprašytas erdvinės kontekstinės informacijos įtraukimas į požymio reikšmių arba klasės žymių tikimybinį skirstinį. Auto-modelių savybės panaudotos eksponentinės šeimos modeliams sudaryti. Siūlomo generatyvinio klasifikavimo algoritmo sudarymui naudojamas BDF metodas, paremtas vienmačių sąlyginių tankių (tikimybių) funkcijų logaritmų santykiu.

Išvestos tikrosios klaidos išraiškos (angl. *Actual Error Rate, AER*) panaudotos tikslumui įvertinti.

Empiriniame tyrime pasiūlytieji klasifikavimo algoritmai buvo išanalizuoti naudojant simuliuotus ir realius duomenis, skirtus AER įvertinti. Nežinomi populiacijos parametrai buvo įvertinti naudojant Bajeso analizės (angl. *Bayes Analysis, BA*), maksimalaus tikėtinumo (angl. *Maximum Likelihood, ML*) ir pseudo maksimalaus tikėtinumo (angl. *Maximum Pseudo-Likelihood, MPL*) metodus. Pasiūlytų klasifikatorių palyginimas atliktas panaudojant įvairias klasių žymių funkcijas. Remiantis realių duomenų požymio reikšmių savybėmis, auto-beta ir perteklinių nulių (angl. *zero-inflated*) auto-beta modeliai buvo pasirinkti klasifikavimo uždaviniui spręsti.

Darbo mokslinis naujumas ir jo reikšmė

Šis darbas praplečia prižiūrimųjų generatyvinių metodų taikymą statistiniam erdvinių duomenų klasifikavimui. Pagrindinius šios disertacijos indėlius galima apibūdinti taip:

1. GRF stebinio klasifikavimo uždavinio išplėtimas, parametrų vertinimui naudojant Bajeso metodą;
2. klasifikavimo uždavinio sprendimas vienmačiam TRF stebiniui;
3. klasifikavimo uždavinio sprendimas vienmačiam atsitiktinio lauko stebiniui iš eksponentinės šeimos skirstinių, naudojant auto-modelius;
4. klasifikavimo uždavinio sprendimas vienmačiam stebiniui su perteklinių nulių beta (angl. *Zero-Inflated Beta, ZIB*) skirstiniu.

Ginamieji teiginiai

1. BA metodas parametrų vertinimui turi pranašumą prieš ML metodą požymių, kurių skirstinys aprašomas GRF modeliu, klasifikavimo uždaviniams.
2. ML metodas parametrų vertinimui turi pranašumą prieš mažiausių kvadratų (angl. Least Squares, LS) metodą vienmačio TRF stebinio klasifikavimui.
3. Erdvinis kontekstinis klasifikavimas, paremtas BDF, turi pranašumą prieš klasifikavimą, paremtą tiesine diskriminantine funkcija (angl. Linear Distriminant Function; LDF) požymiams su beta skirstiniu.
4. Erdvinės informacijos, paremtos kaimynystės schemomis tarp mokymosi aibės elementų, įtraukimas į naujo taško vertinimą per klasių žymių skirstinį pagerina klasifikavimo tikslumą.

Disertacijos struktūra ir apimtis

Disertaciją sudaro įvadas, 3 skyriai, išvados, literatūros sąrašas ir santrauka lietuvių kalba. Įvade pateikiama įžanga į tyrimą ir disertacijos apžvalga. Pirmas skyrius yra skirtas darbų analizei, susijusiai su prižiūrimojo generatyvinio klasifikavimo metodais bei erdvinių duomenų modeliais. Jame pateikiamas prižiūrimojo generatyvinio klasifikavimo metodas, pagrįstas BDF, ir AER išraiškos. Antrajame skyriuje pateikti pagrindiniai disertacijos rezultatai, susiję su siūlomu BDF paremtu prižiūrimojo generatyvinio klasifikavimo algoritmu. Šių funkcijų sudarymui naudojamos vienmačių sąlyginių tankių (tikimybių) išraiškos, aprašančios erdvinių duomenų modelius. Trečiajame skyriuje pateikiami skaitiniai eksperimentai ir pritaikymas. Bendrosios išvados pristatomos po trečiojo skyriaus. Disertacijoje įtraukti 113 literatūros šaltiniai; jie yra pateikti darbo pabaigoje. Disertaciją sudaro 150 puslapiai, 21 paveikslų ir 10 lentelių. Disertacija parašyta anglų kalba.

1. ERDVINIO KONTEKSTINIO KLASIFIKAVIMO, SĄLYGINIŲ ELIPTINIŲ IR EKSPONENTINIŲ ŠEIMŲ SKIRSTINIŲ LITERATŪROS APŽVALGA

Šiame skyriuje analizuojami darbai, susiję su erdvinių duomenų modeliavimu ir erdviniu kontekstiniu klasifikavimu, remiantis BR temomis. Siūlomų prižiūrimojo generatyvinio klasifikavimo algoritmų teorinis pagrindas pateikiamas remiantis šiais darbais: [A1], [A2], [A3].

Erdvinių duomenų modeliavimas – tai tyrimo sritis, kai sudarant statistinius erdvinių duomenų modelius atsižvelgiama į vietos informaciją, susijusią su tiriamu požymiu ir jų sąveiką. (Banerjee ir Gelfand, 2003). GRF modeliai, kurie aprašomi vidurkio ir kovariacinėmis funkcijomis, plačiai nagrinėjami erdvinio statistinio modeliavimo literatūroje (Cressie, 1993; Diggle ir Ribeiro, 2007; Chiles ir Delfiner, 2012; Gelfand and Schliep, 2016). Duomenims, turintiems daugiamatį simetrinį skirstinį, bet netenkinantiems GRF savybių, modeliuoti pasirinkti eliptinės šeimos skirstiniai, kurie patrauklūs savo tankio funkcijos išraiškomis (Kim ir Mallick, 2003, Boente ir kt., 2014, Bankestad, 2020). Vienas iš dažnai sutinkamų yra *t*-skirstinys, taikomas duomenims su sunkiomis „uodegomis“ (angl. *heavy tails*). Daugiamačiu atveju tai yra TRF, kuris turi naudingų analitinių savybių: tiesinės kombinacijos turi *t*-skirstinį, o marginaliniai ir sąlyginiai skirstiniai sudaromi išlaikant tą patį principą (Roislien ir More, 2006, Nadarajah ir Kotz, 2008, Shah ir kt., 2014, Ding, 2016).

Kita svarbi skirstinių šeima, taikoma statistiniam duomenų modeliavimui, yra eksponentinės šeimos skirstiniai. Šie skirstiniai taikomi tiek tolydiems, tiek diskretiems duomenims, turintiems simetrinį ar asimetrinį pasiskirstymą (Zhang, 2002; Fitzmaurice ir kt., 2004, Peng X. ir kt., 2016, Chan ir kt., 2021).

Pereinant nuo erdvinių duomenų modeliavimo prie klasifikavimo uždavinių, erdvinės kontekstinės informacijos sąvoka naudojama apibrėžti erdvinės informacijos įtraukimą į klasifikatoriaus struktūrą. Switzer (1980) vienas pirmųjų nagrinėjo erdvinių duomenų klasifikavimą. Jo darbą pratęsė Mardia (1984), palygindamas klasifikavimo klaidas, kai stebėjimai laikomi priklausomais ir nepriklausomais. GRF stebinių statistinio klasifikavimo uždaviniai yra plačiai išnagrinėti (pvz. Atkinson ir Lewis, 2000, Shekhar ir kt., 2002, McLachlan, 2004). Tačiau nė vienas iš paminėtų autorių neanalizavo AER išraiškų. Klasifikavimo riziką, susijusią su nekoreliuotais stebėjimais ir įvairiais mokymo imčių planais, tyrė Dučinskas (1997). Vėliau Šaltytė (2001), Šaltytė ir Dučinskas (2002) pasiūlė AER vidurkio aproksimacijos formulę skaliariniam geostatistiniam GRF stebiniui dviejų

klasių atveju. Vėliau rezultatus apibendrino daugiamačiam erdvės-laiko modeliui (Šaltytė-Benth, Dučinskas (2005). Išsamų empirinį skirtingų klasifikavimo procedūrų palyginimą galima rasti Atkinson ir Lewis (2000) bei Berret ir Calder (2016) darbuose. Tačiau visose paminėtose publikacijose buvo daroma prielaida, kad stebinys, kurį siekiama klasifikuoti (dar vadinamą fokaliniu stebiniu, (angl. *Focal Observation, FO*), ir mokymo imtis yra nepriklausomi, t. y. formuojant diskriminantines funkcijas naudojami marginaliniai FO tankiai. Šios nepriklausomumo prielaidos erdvinių duomenų klasifikavimo uždaviniuose pirmą kartą atsisakė K. Dučinskas (2009a, 2009b), pateikdamas klasifikavimo klaidos aproksimacijos formulę, kai nežinomos tikrosios vidurkio parametrų reikšmės bei naudojama kovariacinė funkcija su nežinomu vieninteliu mastelio parametru, kitus parametrus laikant žinomais. Dreižienė (2019) pasiūlė klasifikavimo riziką vertinti, kai nežinomi populiacijų parametrai, įtraukiant ir anizotropijos parametrus. Pasiūlytas erdvinių tiesinių mišrių modelių, paremtų hibridiniu AER vertinimu, palyginimo metodas. Nagrinėtas pasiūlyto metodo pranašumas prieš indikatorinio krigingo metodą (Dreižienė ir Dučinskas, 2020). Buvo sukurtas GRF stebinių klasifikavimo metodas, paremtas BDF naudojant semivariogramas (Dučinskas ir Dreižienė, 2021a). Taip pat nagrinėtas universalaus krigingo modelis, kai yra keletas populiacijų, apibrėžtų skirtingais GRF su antros eilės regresijos parametrais; šis modelis išreikštas naudojant semivariogramas ir dispersijas (Dučinskas ir Dreižienė, 2021b).

Conde ir kt. (2005) nagrinėjo eksponentinio skirstinio populiacijų klasifikavimą dviejų ir daugiau klasių atveju su nepriklausomais kintamaisiais. Kleinman (2004) pasiūlė metodą, paremtą apibendrintais tiesiniais mišriais modeliais, (angl. *Generalized Linear Mixed Models, GLMM*) erdvinių duomenų klasterizavimui naudojant logistinės regresijos modelį. Praktiškai taikant ligų protrūkių tyrime, erdvinė priklausomybė buvo pašalinta. Kauermann ir kt. (2010) pasiūlė naują klasifikavimo algoritmą, pagrįstą faktorių parinkimu ir GLMM pritaikyta logistine regresija. Šis algoritmas taip pat naudoja neerdvinius GLMM. Kumar ir Misra (2014) pasiūlė klasifikavimo taisykles eksponentiniu skirstiniu aprašomoms populiacijoms. Darbas praplėstas analizuojant eksponentinio skirstinio vidurkio ir dispersijos parametrus nepriklausomų populiacijų atveju (Jana ir Kumar, 2016; Jan ir kt., 2016).

Andrews ir kt. (2011) pasiūlė nepriklausomų stebėjimų klasifikavimo metodą, pagrįstą daugiamačių *t*-skirstinių mišiniais.

Batsidis ir Zografos (2011) išvedė eliptinio atsitiktinio lauko stebėjimų klaidingo klasifikavimo tikimybių pasiskirstymo funkcijos asimptotinę aproksimaciją. Tačiau jie netyrė vidutinės klaidos tikimybės (angl. *Expected*

Error Eate, EER) aproksimacijų ir įverčių. Thompson ir kt. (2020) pasiūlė prižiūrimojo generatyvinio klasifikavimo metodą nepriklausomų stebėjimų, turinčių matricinį *t*-skirstinį (angl. *matrix-variate t-distribution*).

Remiantis nagrinėtais literatūros šaltiniais galima daryti išvadą, kad BDF paremtas erdvinis kontekstinis klasifikavimas yra atliekamas naudojant Gauso diskriminantines funkcijas. Analizuojant susijusius darbus erdvinių duomenų analizės tema, galima daryti išvadą, kad GRF savybių netenkinantiems duomenims modeliuoti naudojami ir kiti tikimybių modeliai, tokie kaip *t*-skirstinys, arba eksponentinės šeimos skirstiniai. Nauji erdvinių duomenų klasifikavimo metodai leistų išplėsti prižiūrimojo generatyvinio klasifikavimo metodų, paremtų BDF, galimybes. Kitoje šio skyriaus dalyje pateikiama BR pagrįsta kontekstinio klasifikavimo taisyklė, kurios pagrindu konstruojamas prižiūrimojo generatyvinio klasifikavimo algoritmas, skirtas erdvinių ne Gauso duomenų klasifikavimui.

Duomenų modelis. Nagrinėjamas atsitiktinis laukas (angl. *Random Field, RF*) $\left\{Z(s): s \in D \subset \mathbb{R}^d\right\}$ aprašomas baigtiniamačiais skirstiniais (angl. *multivariate distributions*). Pagrindinis tikslas yra vienmačio stebinio $Z_0 = Z(s_0)$, $s_0 \in D$ klasifikavimas į vieną iš *m* populiacijų Ω_l, $l = 1,...,m$. Pagal Beret ir Calder (2016) erdvės taškas s_0, kuriame stebinys klasifikuojamas, yra vadinamas fokaliniu erdvės tašku. Iš to seka, kad anksčiau aprašytas stebinys Z_0 yra vadinamas FO. Tegu $P_l\left(Z_0 \middle| Z = z; \Psi\right)$ – sąlyginė tikimybė, FO Z_0, kuris aprašomas diskrečiu skirstiniu, o $p_l\left(Z_0 \middle| Z = z; \Psi\right)$ – sąlyginė tankio funkcija, FO Z_0, kuris aprašomas tolydžiu skirstiniu, kai $Z = z$ populiacijoje Ω_l $l = 1,...,m$. Pagrindinė prielaida, kuria remiamasi šiame darbe: duomenyse egzistuoja statistinė erdvinė priklausomybė ir sąlyginio tankio (tikimybės) funkcijos išraiškos apibrėžiamos įtraukiant kaimyninių taškų požymių reikšmes. Ši informacija vadinama kontekstine informacija. Mokymo imties erdvės taškų aibė (angl. *Set of Training Locations, STL*) t. y. $S_N = \left\{s_i \in D; i = 1,...,N\right\}$ apibrėžia mokymo imties erdvinę struktūrą (Shekhar ir kt., 2002).

Klasių žymių tikimybės. Klasių žymių tikimybės turi užkoduotą apriorinę tyrėjo informaciją apie klasifikavimo klases. Kai klasės subalansuotos, paprastai klasių žymių tikimybės laikomos lygiomis. Šios tikimybės gali būti vertinamos naudojant mokymo imčių dydžių funkcijas. Taip pat galima daryti

prielaidą, kad klasių žymių tikimybės priklauso nuo FO ir jo kaimynų. Klasių žymių tikimybės formulė populiacijai Ω_l, kuri įtraukia kaimyninių taškų informaciją naudojant atvirkštinio atstumo funkciją, galima apibrėžti taip:

$$\pi_i^l = \sum_{j \in NN_i^1} \left(1/d_{ij}\right) \Big/ \sum_{j \in NN_i} \left(1/d_{ij}\right), \tag{1.2}$$

čia d_{ij} yra Euklido atstumas tarp erdvės taškų s_i ir s_j, $i, j = \overline{1, N}$; $NN_i = \sum_{l=1}^{m} NN_i^l$, čia NN_i^l erdvės taško s_i artimiausių kaimynų (angl. *Nearest Neighbourhood, NN*) aibė populiacijoje $\Omega_l, l = 1, ..., m.$ NN galima apibrėžti įtraukiant fiksuotą kaimynų skaičių arba tam tikra erdvės sritimi su fiksuotu atstumu nuo FO. NN aibės, naudojamos klasių žymių tikimybių vertinimui, ir sąlyginio tankio funkcijų (tikimybių) išraiškose; jos gali būti tos pačios arba skirtingos.

Diskriminantinė funkcija. Taisyklė, minimizuojanti klaidingo klasifikavimo tikimybę, yra vadinama Bajeso klasifikavimo taisykle (McLachlan, 2004, Anderson, 2003). Tegu $W^B\left(Z_0, \Psi\right)$ žymi BDF dviejų klasių atveju, paremtą santykių logaritmu:

$$W^B\left(Z_0, \Psi\right) = \ln\left(\frac{p_1\left(Z_0 \middle| Z = z; \Psi\right)}{p_2\left(Z_0 \middle| Z = z; \Psi\right)}\right) + \gamma, \tag{1.8}$$

čia Ψ žymi modelio parametrų rinkinį; $\gamma = \ln\left(\pi_0^1 / \pi_0^2\right)$. Pagal (1.8) Z_0, kai duota $Z = z$ yra priskiriama populiacijai Ω_1, jei $W^B\left(Z_0, \Psi\right) \geq 0$. Klaidingo klasifikavimo tikimybė arba klaidos tikimybė, susijusi su šiomis diskriminantinėmis funkcijomis, užrašoma taip (Anderson, 2003):

$$P_{0z}^B\left(\Psi\right) = \sum_{l=1}^{2} \pi_0^l P_{lz}, \tag{1.9}$$

čia $P_{lz} = P_{lz}\left(\left(-1\right)^l W^B\left(Z_0, \Psi\right) \geq 0\right)$, P_{lz} tikimybinis matas, priklausantis nuo FO $Z_0 \in \Omega_l$ sąlyginių klaidingo klasifikavimo tikimybių, susijusių su $p_l\left(Z_0 \middle| Z = z; \Psi\right)$ ar $P_l\left(Z_0 \middle| Z = z; \Psi\right)$, $l = 1, 2$.

Bajeso taisyklė yra optimali taisyklė, minimizuojanti klasifikavimo tikimybę tais atvejais, kai populiacijos yra pilnai apibrėžtos. Praktiniuose uždaviniuose retai pasitaiko pilno populiacijų apibrėžtumo atvejų. Tuomet nežinomos parametrų reikšmės yra pakeičiamos jų įvertiniais $\hat{\Psi}$, gautais iš mokymo imties. Tokiu būdu, diskriminantinė funkcija vadinama įterpta (angl.

Plug-in) BDF (PBDF). Taigi, (1.8) formulėje nežinomas parametrų reikšmes pakeitus jų įvertiniais, gaunama $W_{lk}^{B}\left(Z_0,\hat{\Psi}\right)$. Naudojant PBDF $W_{lk}^{B}\left(Z_0,\hat{\Psi}\right)$, galima apibrėžti AER (Dučinskas, 2009):

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}\ ,$$

čia $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W^B\left(Z_0,\hat{\Psi}\right)\geq 0\right),\ l=1,2$.

EER gaunama apskaičiavus AER vidurkį, atsižvelgiant į mokymo imties pasiskirstymą:

$$EER=E_z\left(P_{0z}^{B}\left(\hat{\Psi}\right)\right). \qquad (1.10)$$

Analitinių EER išraiškų paieška yra sudėtinga matematinė užduotis, todėl šiame darbe empiriniai EER įverčiai apskaičiuojami naudojant AER išraiškas.

Pereinant nuo Gauso skirstinio prie kitų modelių, nuspręsta pasirinkti eliptinės šeimos *t*-skirstinį, kuris turi naudingų analitinių savybių: tiesinės kombinacijos turi tą patį pasiskirstymą, o marginaliniai ir sąlyginiai skirstiniai išlaikant tą patį principą.

Pereinant nuo simetriškų skirstinių, siekiant išplėsti taikymo sritį, pasirinkti eksponentinės šeimos skirstiniai, kurie apibūdina asimetrinius duomenis su sunkiomis uodegomis. Pasirinkti dažnai naudojami du diskretieji skirstiniai: Puasono ir binominis. Taip pat pasirinkti du tolydieji skirstiniai: gama ir beta. Erdvinių duomenų, aprašomų eksponentinės šeimos skirstiniu, modeliavimui pasirinkti Besag (1974) pasiūlyti auto-modeliai, kurie sudaromi remiantis dviem prielaidomis. Pirma, priklausomybė tarp požymio reikšmių erdvės taškuose yra porinė. Antra, požymio reikšmių sąlyginis tikimybinis pasiskirstymas priklauso eksponentinei skirstinių šeimai. Kitame skyriuje pateikiamos BDF ir AER išraiškos, kurias sudarant naudojamos pasirinktų skirstinių vienmačių sąlyginių tankių išraiškos.

2. PRIŽIŪRIMOJO GENERATYVINIO KLASIFIKAVIMO ALGORITMO KŪRIMAS POŽYMIAMS SU ERDVINIU NE GAUSO SKIRSTINIU, DISKRIMINANTINIŲ FUNKCIJŲ IR AER IŠRAIŠKOS

Šiame darbe pagrindinis dėmesys skiriamas prižiūrimo generatyvinio klasifikavimo algoritmui, kai klasifikavimas atliekamas naudojant BDF. Šioje dalyje pateikiamos BDF išraiškos, sudaromos naudojant sąlyginių tankio (tikimybės) funkcijų santykio logaritmą, BDF bendras pavidalas pateikiamas (1.8). Taisyklių kūrimo metodika suformuluota FO Z_0 su pasiskirstymu iš eliptinės arba eksponentinės šeimos, naudojant vienmates sąlyginio tankio (tikimybių) funkcijas. Paprastumo dėlei buvo išnagrinėti atskiri populiacijų atvejai, kurie gali būti naudojami kaip baziniai siūlomų klasifikavimo funkcijų panaudojimo pavyzdžiai. Šie pavyzdžiai taip pat gali būti apibendrinami ir panaudojami sprendžiant klasifikavimo uždavinius su įvairioms parametrinėms struktūroms. Klasifikavimo tikslumui įvertinti išvestos AER išraiškos. Šiame skyriuje pateikti rezultatai publikuoti straipsniuose [A2], [A3].

***t*-skirstinio modelis.** Tegu FO Z_0 populiacijoje Ω_l, $l=\overline{1,m}$ turi sąlyginį *t*-skirstinį su sąlyginiu vidurkiu μ_{0z}^l ir sąlyginiu mastelio parametru ω_{0z}. Dviejų klasių $m=2$ ir lygių klasių žymių tikimybių atveju, BDF apibrėžta (1.8) išraiška šiuo atveju:

$$W^B\left(Z_0,\Psi\right)=\left(Z_0-\frac{1}{2}\left(\mu_{0z}^1+\mu_{0z}^2\right)-\alpha'\left(z-X\beta\right)\right)\left(\mu_{0z}^1-\mu_{0z}^2\right), \qquad (2.1)$$

čia $\alpha=R^{-1}r_0$; r_0 yra vektorius, apibrėžiantis erdvinę koreliaciją tarp Z_0 ir *Z*; *R* yra *Z* erdvinės koreliacijos matrica; *z* yra mokymo imties požymio reikšmių realizacijos, $Z=z$; *X* yra *Z* plano matrica; $\beta'=\left(\beta_1',\beta_2'\right)$ yra $1\times 2q$ regresijos parametrų vektorius; q yra regresorių skaičius; *N* yra mokymo aibės dydis; $\Psi=\left\{\beta',\sigma^2,\nu\right\}$ yra modelio parametrų rinkinys; σ^2 yra dispersijos parametras; ν žymi laisvės laipsnius. Toliau laikoma, kad regresijos parametrai yra nežinomi.

FO vidurkis $\mu_0^l=x_0'\beta_l$, $l=1,2$, x_0' yra neatsitiktinių regresorių vektorius taške s_0, β_l yra $q\times 1$ regresijos parametrų vektorius klasei Ω_l, $l=1,2$. Tegu $\Delta\mu_0=\mu_0^1-\mu_0^2>0$ ir sąlyginis Mahalanobio atstumas $\Delta_0=\Delta\mu_0\big/\left(\sigma\sqrt{\rho_0}\right)$; čia

ρ_0 žymi sąlyginės ir marginalinės dispersijos santykį. Tada PBDF išraiška gaunama iš BDF (2.1), vidurkio parametrus β pakeičiant jų įverčiais $\hat{\beta}$:

$$W^B\left(Z_0,\hat{\beta}\right)=\left(Z_0-\alpha'\left(z-X\hat{\beta}\right)-\frac{1}{2}x_0'H\hat{\beta}\right)\left(x_0'G\hat{\beta}\right), \tag{2.2}$$

su $H=\left(I_q,I_q\right)$ ir $G=\left(I_q,-I_q\right)$; čia I_q yra q eilės vienetinė matrica.

Lema 3. Klaidos tikimybė, remiantis BDF $W^B\left(Z_0,\Psi\right)$ (2.2), apibrėžiama taip:

$$P_{0z}^B(\beta)=S_{\nu+N}\left(-(\Delta_0/2)\sqrt{\left(\nu/((\nu-2)\zeta_z(\beta))\right)}\right), \tag{2.3}$$

Lema 4. AER, naudojant klaidos tikimybės išraišką (2.3) ir PBDF (2.2), galima išreikšti taip:

$$P_{0z}^B\left(\hat{\beta}\right)=\sum_{l=1}^{2}S_{\nu+N}\left(\hat{Q}_l\right)/2, \tag{2.4}$$

čia $\hat{Q}_l=(-1)^l\left(\left(\hat{a}_l+\hat{b}\hat{\beta}\right)\mathrm{sgn}(x_0'G\hat{\beta})\Big/\sigma\sqrt{\rho_0\zeta\left(z\right)}\right)\sqrt{(\nu+N)/(\nu+N-2)}$,

kai $\hat{a}_l=x_0'\hat{\beta}_l-\alpha'X\hat{\beta}$, $\hat{b}=\alpha'X-x_0'H/2$; $S_N(\bullet)$ žymi pasiskirstymo funkciją (angl. *Cumulative Distribution Function; CDF*) vienmačiam *t*-skirstiniui $T_1(0,1,N)$. Lemos 3 ir Lemos 4 įrodymai pateikiami disertacijos 2.1 poskyryje.

Auto-Puasono modelis. Tegu FO Z_0 populiacijoje Ω_l, $l=1,2$ turi sąlyginį Puasono skirstinį su sąlyginiu vidurkiu μ_0^l. BDF $W^B\left(Z_0,\Psi\right)$, apibrėžtos (1.8), išraiška:

$$W^B\left(Z_0,\Psi\right)=\left(\theta_0^1-\theta_0^2\right)Z_0-\left(\mu_{0z}^1-\mu_{0z}^2\right)+\gamma, \tag{2.5}$$

čia $\Psi=\left\{\theta_0^l,\mu_0^l;\,l=1,2\right\}$ yra modelio parametrų rinkinys; $\gamma=\ln\left(\pi_0^1/\pi_0^2\right)$.

Lema 5. AER, susijusi su BDF (2.5), apibrėžiama taip:

$$P_{0z}^B\left(\hat{\Psi}\right)=\pi_0^1\sum_{B_1}\exp\left\{Z_0\ln\hat{\mu}_{0z}^1-\hat{\mu}_{0z}^1-\ln\left(Z_0!\right)\right\}$$
$$+\pi_0^2\left(\sum_{B_2}\exp\left\{Z_0\ln\hat{\mu}_{0z}^2-\hat{\mu}_{0z}^2-\ln\left(Z_0!\right)\right\}\right)$$

Čia $B_1 = \left\{ Z_0 : Z_0 \in \mathbb{N} \cup \{0\}, W^B\left(Z_0, \Psi\right) < 0 \right\}$

ir $B_2 = \left\{ Z_0 : Z_0 \in \mathbb{N} \cup \{0\}, W^B\left(Z_0, \Psi\right) \geq 0 \right\}$ yra sumavimo sritys. Lemos 5 įrodymas pateikiamas disertacijoje 2.2 poskyryje.

Auto-Binominis modelis. Tegu Z_0 populiacijoje Ω_l, $l = 1,2$ turi sąlyginį binominį skirstinį su fiksuotų bandymų skaičiumi n_0 ir sėkmės tikimybe kiekvieno bandymo metu p_0^l. BDF $W^B\left(Z_0, \Psi\right)$, apibrėžtos (1.8), išraiška:

$$W^B\left(Z_0, \Psi\right) = Z_0\left(\theta_0^1 - \theta_0^2\right) + n_0\left(\ln\left(1 - p_{0z}^1\right) - \ln\left(1 - p_{0z}^2\right)\right) + \gamma, \qquad (2.6)$$

čia $\Psi = \left\{ \theta_0^l, p_{0z}^l; l = 1,2 \right\}$ yra modelio parametrų rinkinys; $\gamma = \ln\left(\pi_0^1 / \pi_0^2\right)$.

Lema 6. AER, susijusi su BDF (2.6), apibrėžiama taip:

$$P_{0t}^B\left(\hat{\Psi}\right) = \pi_0^1 \sum_{B_1} \exp\left\{ \begin{array}{l} \ln\left[\dfrac{n_0!}{Z_0!\left(n_0 - Z_0\right)!}\right] + Z_0 \ln \hat{p}_{0z}^1 + \\ + \left(n_0 - Z_0\right)\ln\left(1 - \hat{p}_{0z}^1\right) \end{array} \right\} +$$

$$+ \pi_0^2 \left(\sum_{B_2} \exp\left\{ \begin{array}{l} \ln\left[\dfrac{n_0!}{Z_0!\left(n_0 - Z_0\right)!}\right] + Z_0 \ln \hat{p}_{0z}^2 + \\ + \left(n_0 - Z_0\right)\ln\left(1 - \hat{p}_{0z}^2\right) \end{array} \right\} \right),$$

čia $B_1 = \left\{ Z_0 : Z_0 \in \mathbb{N} \cup \{0\}, W^B\left(Z_0, \Psi\right) < 0 \right\}$

ir $B_2 = \left\{ Z_0 : Z_0 \in \mathbb{N} \cup \{0\}, W^B\left(Z_0, \Psi\right) \geq 0 \right\}$ yra sumavimo sritys, $\hat{\Psi} = \left\{ \hat{\theta}_0^l, \hat{\theta}_{0j} \right\}$ yra įvertintų modelio parametrų rinkinys. Lemos 6 įrodymas pateikiamas disertacijoje 2.2 poskyryje.

Auto-Gama modelis. Tegu FO Z_0 populiacijoje Ω_l, $l = \overline{1,m}$ turi sąlyginį gama skirstinį su mastelio parametru α_{0z}^l, skirtingu kiekvienai klasei *l*, ir formos parametru γ_{0z}. BDF, apibrėžtos (1.8), išraiška:

$$W^B\left(Z_0, \Psi\right) = \left(\frac{\alpha_{0z}^1 - \alpha_{0z}^2}{\alpha_{0z}^1 \alpha_{0z}^2}\right) Z_0 + \gamma_{0z} \ln\left(\frac{\alpha_{0z}^2}{\alpha_{0z}^1}\right) + \gamma, \qquad (2.7)$$

čia $\Psi = \left\{ \alpha_{0z}^l, \gamma_{0z}, l = 1,2 \right\}$ yra modelio parametrų rinkinys; $\gamma = \ln\left(\pi_0^1 / \pi_0^2\right)$.

Lema 7. AER, susijusi su BDF (2.7), apibrėžiama taip:

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}$$

čia $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int_{B_l}H\left(\left(-1\right)^l W\left(u,\hat{\Psi}\right)\right)p_l\left(u\right)du,$ kai $l=1,2,$

$B_l=\left\{u:u\in\left(0;\infty\right)\right\}$ yra integravimo sritis, $H\left(\bullet\right)$ yra Heaviside žingsninė funkcija (angl. *Heaviside step function*). Lemos 7 įrodymas pateikiamas disertacijoje 2.2 poskyryje.

Auto-beta modelis. Tegu FO Z_0 populiacijoje Ω_l, $l=1,2$ turi sąlyginį beta skirstinį su vidurkiu μ_{0z}^l ir tikslumo parametru ϕ_{0z}^l apibrėžtais naudojant natūraliųjų parametrų išraiškas. BDF $W^B\left(Z_0,\Psi\right)$, apibrėžtos (1.8), išraiška:

$$W^B\left(Z_0,\Psi\right)=\left(A_{01}^1-A_{01}^2\right)\ln\left(Z_0\right)+\left(A_{02}^1-A_{02}^2\right)\ln\left(1-Z_0\right)+\gamma_0\left(\Psi\right),\qquad(2.8)$$

čia $\gamma_0\left(\Psi\right)=\ln\left\{\pi_0^1 Be\left(a_{02},b_{02}\right)/\,\pi_0^2 Be\left(a_{01},b_{01}\right)\right\}$, $Be\left(a_{0l},b_{0l}\right)$ yra Euler Beta funkcija su $a_{0l}=A_{01}^l+1$, $b_{0l}=A_{02}^l+1$; A_{0h}^l yra natūraliųjų parametrų išraiškos, skirtingos kiekvienai populiacijai Ω_l, $l=1,2$, su parametrų rinkiniu Ψ; h yra beta skirstinio pakankamų statistikų (angl. *sufficient statistics*) skaičius.

Lema 8. AER, susijusi su BDF (2.8), apibrėžiama taip:

$$P_{0z}^{B}\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz}\,,\qquad(2.9)$$

čia $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int H\left(\left(-1\right)^l W\left(u,\hat{\Psi}\right)\right)p_l\left(u\right)du,$ kai $l=1,2$;

$B_l=\left\{u:u\in\left(0;1\right)\right\}$ yra integravimo sritis; p_l yra beta skirstinio sąlyginio tankio funkcija. Lemos 8 įrodymas pateikiamas disertacijoje 2.2 poskyryje.

Perteklinių nulių auto-beta modelis. Tegu FO Z_0 populiacijoje Ω_l, $l=1,2$ turi sąlyginį ZIB skirstinį su mišinio parametru c_{0z}^l, vidurkiu μ_{0z}^l ir tikslumo parametru ϕ_{0z}^l iš beta skirstinio sąlyginės tankio funkcijos. BDF $W^B\left(Z_0,\Psi\right)$, apibrėžtos (1.8), išraiška:

$$W^B\left(Z_0,\Psi\right)=\ln\left(\frac{\pi_0^1 p_1\left(Z_0\left|Z_j;\Psi\right.\right)}{\pi_0^2 p_2\left(Z_0\left|Z_j;\Psi\right.\right)}\right)=\ln\left(\frac{\pi_0^1}{\pi_0^2}\right)+\ln\left(\frac{c_{0z}^1}{c_{0z}^2}\right)I\left(Z_0=0\right)+$$
$$+\left(1-I\left(Z_0=0\right)\right)\ln\left(\frac{\left(1-c_{0z}^1\right)}{\left(1-c_{0z}^2\right)}\right)+ \tag{2.10}$$
$$+\left(1-I\left(Z_0=0\right)\right)\left(A_{01}^1-A_{01}^2\right)\ln\left(Z_0\right)+\left(A_{02}^1-A_{02}^2\right)\ln\left(1-Z_0\right),$$

čia $I\left(\bullet\right)$ – indikatorinė funkcija.

Lema 9. AER, naudojantis įterptąją Bajeso klasifikavimo taisyklę, susijusią su BDF (2.10), kai $m=2$, apibrėžiamas taip:

$$P_{0z}^B\left(\hat{\Psi}\right)=\sum_{l=1}^{2}\pi_0^l\hat{P}_{lz} \tag{2.11}$$

čia $\hat{P}_{lz}=P_{lz}\left(\left(-1\right)^l W\left(Z_0,\hat{\Psi}\right)\geq 0\right)=\int\limits_{B_l}H\left(\left(-1\right)^l W\left(u,\hat{\Psi}\right)\right)p_l\left(u\right)du,$

$B_l=\left\{u:u\in\left[0;1\right)\right\}$ yra integravimo sritis; p_l yra ZIB skirstinio sąlyginio tankio funkcija, $l=1,2$.

Aukščiau pateikiamos BDF išraiškos, leidžiančios išspręsti porinio klasifikavimo uždavinius erdviniams duomenims, modeliuojamiems ne Gauso skirstiniu. Tai įgyvendinama naudojant siūlomą prižiūrimojo generatyvinio klasifikavimo algoritmą, pagrįstą siūlomomis BDF išraiškomis. Siūlomas algoritmas pateikiamas disertacijos 2.3 poskyryje. Taip pat išvedamos AER išraiškos, leidžiančios įvertinti klasifikavimo tikslumą. Tolesnėje dalyje pateikiami rezultatai, gauti pritaikius pasiūlytą algoritmą atliekant empirinį tyrimą.

3. EMPIRINIS TYRIMAS IR DISKUSIJOS

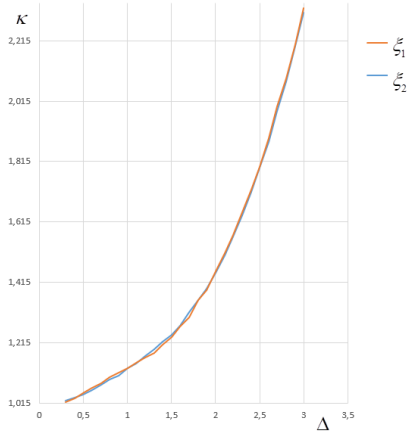
Disertacijos trečiajame skyriuje pateikti skaitiniai eksperimentai ir taikymai. Šioje dalyje trumpai pateikiami eksperimentų rezultatai.

AER įverčiai naudojami siekiant įvertinti diskriminantinės funkcijos tikslumą klasifikuojant būsimus stebėjimus. AER įverčiai priklauso nuo mokymo imties stebėjimų reikšmių ir jų erdvės vietų, kuriose buvo surinkti stebėjimai. EER, apibrėžtos (1.10), įvertinimai gaunami apskaičiuojant AER vidurkį, atsižvelgiant į mokymo imties pasiskirstymą. Empirinis EER įvertinimas, naudojamas disertacijoje, apibrėžiamas taip:

$$\overline{EER}^{(\delta)} = \sum_{i=1}^{M} P_{0z(i)}^{\delta}\left(\hat{\Psi}\right) \Big/ M \tag{3.1}$$

čia $P_{0z(i)}^{\delta}\left(\hat{\Psi}\right)$ yra AER įverčiai; δ žymi skirtingus skaičiavimo metodus; M yra simuliacijų skaičius. Lyginant $\overline{EER}^{(\delta)}$ reikšmes atliekamas klasifikavimo metodų palyginimas.

Pirmame empirinio tyrimo etape atliktas GRF stebinio klasifikavimas dviejų klasių atveju, kai nežinomų modelio parametrų vertinimui naudojami BA ir ML metodai. Tyrimas atliktas simuliuotiems duomenims, o metodų palyginimas atliktas analizuojant indekso $\kappa_{\xi_k} = \overline{EER}^{ML} \Big/ \overline{EER}^{BA}$ reikšmes.

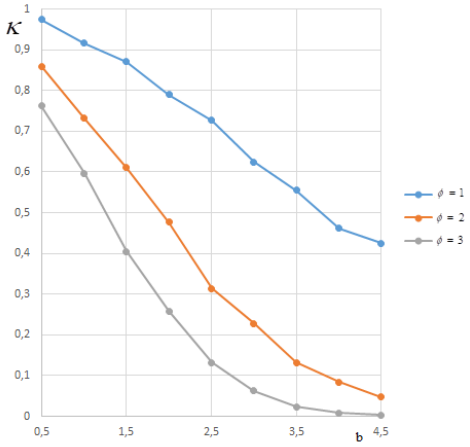


3.3 pav. Indekso κ reikšmės skirtingoms TLC, $M = 10^4$.

Pateikiami rezultatai, kai simuliacijų skaičius $M = 10^4$ ir ξ_k, $k = 1,2$ žymi simetrinę ir asimetrinę mokymo žymių konfigūraciją (angl. *Training Labels Configuration, TLC*). 3.3. pav. matomos dvi vienodo elgesio augimo kreivės, kurios atspindi BA metodo pranašumą prieš ML metodą. Šis pranašumas

reikšmingesnis labiau besiskiriančioms populiacijoms (didesnės Δ reikšmės). Šios išvados galioja tiek simetrinei, tiek asimetrinei TLC.

Kitame tyrimo etape atliktas GRF stebinio kelių klasių klasifikavimo uždavinio sprendimas. Klasifikavimo tikslumas vertinimas naudojant $\overline{EER}$ (3.1) reikšmes, nagrinėjant situacijas su STL struktūromis: grupuota STL (angl. *Grouped STL*, *STLG*) ir mišria (angl. *Mixed STL*, *STLM*). Nustatyta, kad $\overline{EER}$ mažėja didėjant klasių atskiriamumui. Taip pat nustatyta, kad grupuotai klasių žymių struktūrai $\overline{EER}$ mažėja didėjant erdvinės koreliacijos pločiui ϕ (angl. *range*). $\overline{EER}$ reikšmių santykio grafikas pateikiamas 3.7 pav. Skirtingos kreivės atspindi skirtingas kovariacines funkcijas, kurios skiriasi pločio parametru ϕ.

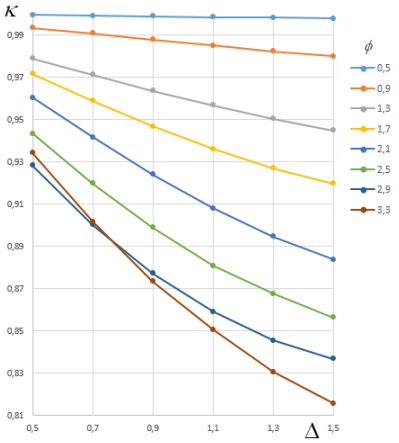


3.7 pav. $\overline{EER}$ reikšmių santykis skirtingoms STL struktūroms.

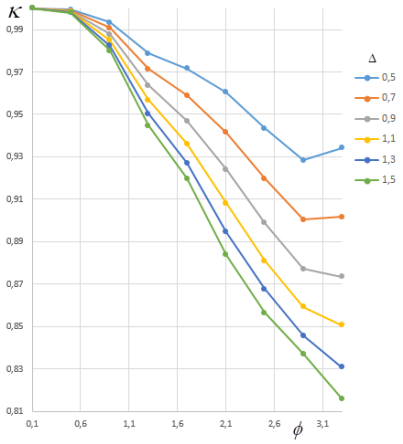
Tyrimas parodo STLG naudojančio metodo pranašumą prieš metodą su STLM. Šis pranašumas yra reikšmingesnis labiau besiskiriančiomis populiacijoms (didesnės *b* reikšmės) nei artimoms populiacijoms. Rezultatai suteikia argumentų, leidžiančių tikėtis, kad populiacijos STLG struktūra gali būti veiksmingai panaudota erdvinių Gauso duomenų klasifikavime naudojant BDF paremtą metodą.

Kitame tyrimo etape PBDF (2.2) funkcija pritaikyta stebėjimo iš TRF klasifikavimo uždaviniui, kai parametrų vertinimui naudojamas ML ir LS metodai. Tiriama, kaip skirtingi statistiniai parametrai veikia klaidingo klasifikavimo tikimybę. Taip pat tirta priklausomybė nuo klasių žymių struktūros. Nustatytas ML pranašumas prieš LS metodą; visais nagrinėtais

atvejais $\overline{EER}^{LS} \geq \overline{EER}^{ML}$. 3.9 pav. ir 3.10 pav. pateikiamos indekso $\kappa = \overline{EER}^{ML} / \overline{EER}^{LS}$ kreivės.



3.9 pav. κ reikšmės, ϕ ir S_{24}.

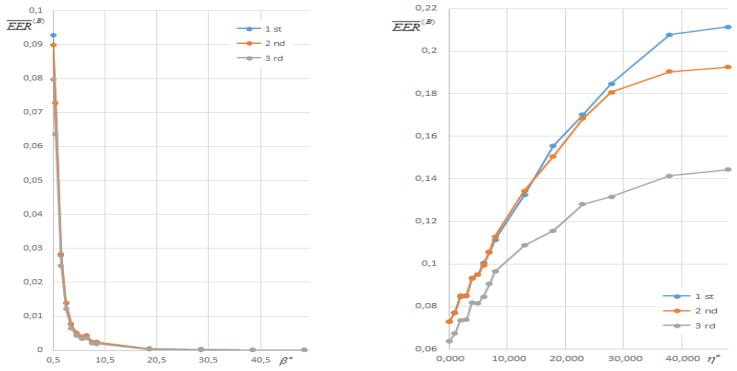


3.10 pav. κ reikšmės, Δ ir S_{24}.

Tyrimas parodo ML vertinimu pagrįsto PBDF pranašumą, palyginti su tuo, kuris pagrįstas LS vertinimu. Šis pranašumas yra reikšmingesnis tais atvejais, kai erdvinė priklausomybė tarp stebėjimų yra didesnė (t. y. didesnės ϕ reikšmės). Ši išvada galioja skirtingiems atstumams tarp populiacijų Δ. Rezultatai suteikia svarių argumentų, kad dažnai sudėtingi erdvinių vidutinių parametrų ML įverčiai turėtų būti naudojami labiau koreliuojantiems erdviniams duomenims, modeliuojamiems TRF, o paprastesnis LS vertinimas galėtų pakeisti šiuos silpnai koreliuojančių erdvinių duomenų įverčius be reikšmingo PBDF našumo praradimo.

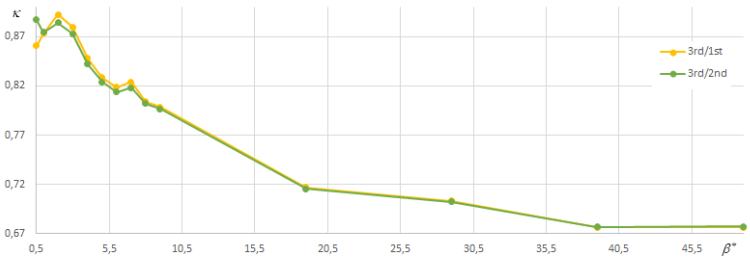
Kitame tyrimo etape atliktas BDF (2.8) funkcijos taikymas stebėjimo iš beta skirstinio atsitiktinio lauko simuliuotų duomenų klasifikavimo uždaviniui. Tiriama $\overline{EER}$ priklausomybė nuo klasių žymių tikimybių vertinimo esant skirtingoms parametrinėms struktūroms. Pritaikyti trys klasių žymių vertinimo metodai: 1 metodas, kai klasių tikimybės laikomos lygiomis, 2 metodas, kai klasių tikimybės skaičiuojamos pagal (1.2), FO kaimynais laikant visus mokymo aibės taškus, ir 3 metodas, kai skaičiuojama pagal (1.2), įtraukiant kaimynus, esančius iki ketvirtos eilės. Analizuotos dvi skirtingos parametrinės struktūros. A struktūra naudojama, kai vertinamos klaidos tikimybės didėjant skirtumui β^* tarp regresijos parametrų, apibrėžiančių skirtingas klases. B struktūra naudojama, kai vertinamos klaidos tikimybės didėjant skirtumui η^* tarp erdvinės priklausomybės parametrų, aprašančių

priklausomybę nuo kaimyninių taškų šiaurės–pietų kryptimi ir vakarų–rytų kryptimi. Šie parametrai tarp klasių nesiskiria. $\overline{EER}^{B}$ reikšmės, kai simuliacijų skaičius $M = 100$, pateikiamos 3.13 pav.



3.13 pav. $\overline{EER}^{B}$ reikšmės su skirtingais klasių žymių tikimybių vertinimo metodais: A struktūra (kairėje) ir B struktūra (dešinėje).

3.13 pav. kairėje pusėje, matoma, kad didėjant β^*, $\overline{EER}^{B}$ reikšmės mažėja. Visų kreivių tendencija rodo mažėjančią klasifikavimo klaidos tikimybę, tačiau sunku pastebėti klasės žymių tikimybių skaičiavimo metodo įtaką gautiems įverčiams. Dėl šios priežasties 3.14 pav. pateikiamas $\overline{EER}^{B}$ reikšmių santykis, kai κ žymi 3-ojo metodo santykį su 1-uoju ir 2-uoju. Šiame grafike matoma, kad didėjant β^*, santykio reikšmės mažėja. Galima daryti išvadą, kad naudojant 3 metodą, klasifikavimas atliekamas tiksliau. Pagal 3.13 pav. dešinėje esantį grafiką matyti, jog didėjant η^*, $\overline{EER}^{B}$ reikšmės taip pat didėja. Tai reiškia, kad klasifikavimo tikslumas mažėja. Išsiskiria situacija, kai klasių tikimybės skaičiuojamos pagal 3 metodą. $\overline{EER}^{B}$ reikšmės didėja ne taip sparčiai, kaip kitais nagrinėtais atvejais.



3.14 pav. κ reikšmių kreivės su skirtingais klasių žymių skaičiavimo metodais naudojant A struktūrą.

Klasifikavimas buvo taip pat atliktas taikant modifikuotą LDF. Wilcoxon (angl. *Wilcoxon signed rank*) testas naudojamas statistiškai įvertinti skirtumą tarp klasifikavimo metodų su BDF ir LDF. Gauti rezultatai pateikiami 3.5 lentelėje.

3.5 lentelė. Wilcoxon testas.

	Klasių žymių tikimybės		
Struktūra	Lygios tikimybės	Atvirkš. atstumo, NN su visais mokymo aibės steb.	Atvirkštinio atstumo, NN su kaimyniniais steb. iki 4 eilės
A	$V = 136$, (3,052e-05)	$V = 120$, (0,0007265)	$V = 126$, (0,0 02942)
B	$V = 153$ (0,0003204)	$V = 103{,}5$ (0,07023)	$V = 47{,}5$, (0,1769)

Čia V yra Wilcoxon testo statistika, o skliaustuose pateikiamos p- reikšmės (angl. *p-values*). Pagal gautus rezultatus galima teigti, kad analizuojant situacijas apibrėžiama struktūra A su reikšmingumo lygmeniu (0,002942). LDF ir BDF $\overline{EER}$ reikšmės statistiškai skiriasi prie skirtingų klasės žymių tikimybių skaičiavimo metodų. Struktūroje B LDF ir BDF $\overline{EER}$ reikšmės statistiškai skiriasi taikant vienodas klasės žymių tikimybes ir kai klasės žymių tikimybės įvertinamos atvirkštiniu atstumu, naudojant visus mokymo aibės stebėjimus (0,070023).

Remiantis gautais rezultatais, galima daryti išvadą, kad išankstinės informacijos, susijusios su FO padėtimi erdvėje ir mokymo aibės taškų padėtimi erdvėje įtraukimas į modelius turi įtakos klasifikavimo klaidai. Ši įtaka žymiai padidėja, kai klasės žymių tikimybės skaičiuojamos naudojant pasirinktą NN aibę iš mokymo aibės.

Kitoje empirinio tyrimo dalyje atliktas generatyvinių algoritmų, pagrįstų disertacijoje pasiūlytomis PBDF, taikymas dugno tipo nustatymui vertinant šakotojo banguolio dumblių (angl. *black carrageen*, lot. *Furcellaria lumbricalis*) padengimą Baltijos jūros pietryčių priekrantės zonoje. Taip pat atliktas AER vertinimas.

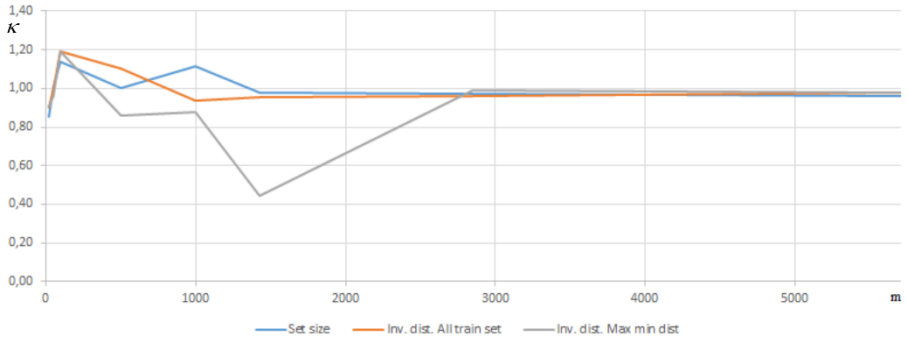
Dumblių padengimo duomenys buvo surinkti 641 $1\ \mathrm{m}^2$ plotuose, fiksuojant jų padengimą procentais. Buvo pasirinkti beta skirstinio ir ZIB skirstinio modeliai. Beta skirstinių šeima siūlo daug tankių ribotame intervale [a, b]. Dėl savo universalumo šie skirstiniai tampa svarbia skirstinių klase analizuojant tokio tipo duomenų tyrimus, kaip santykių, proporcijų ar koncentracijos. Dėl šios priežasties darbe tirti auto-beta modeliai tampa erdvinių duomenų analizei reikšminga modelių klase. Nagrinėjamuose duomenyse yra daugiau nei 70 % nulinių reikšmių. Tokiu atveju yra natūralu

duomenis modeliuoti naudojant dviejų skirstinių mišinį: beta skirstinį ir išsigimusį skirstinį, kurio vertė 0 (angl. *degenerate distribution in a value 0*). Dėl šios priežasties pasirinktas ZIB skirstinio modelis.

Buvo išskirtos dvi dugno paviršiaus klasės: $l=1$, kai dugno paviršių sudaro ne rieduliai, ir $l=2$, kai dugno paviršių sudaro rieduliai. Erdvinė kontekstinė informacija į modelį įtraukiama trimis skirtingais būdais: aprašant determinuoto trendo modelį įtraukiamos klasifikuojamo erdvės taško koordinatės, aprašant erdvinę priklausomybę apibrėžiamos NN sritys bei vertinant klasių žymių tikimybes.

Tikslas yra remiantis turima mokymo imtimi sudaryti tiksliausią modelį, kurį būtų galima taikyti testavimo imties požymio reikšmių klasifikavimui dugno paviršiaus nustatymui. Ieškant geriausio modelio disertacijoje buvo varijuojama NN srities parinkimu pasirenkant sąlyginio tankio funkcijos išraiškas. Klasių žymių tikimybės buvo skaičiuojamos trimis metodais. 1 metodas pagal stebėjimų skaičių kiekvienoje populiacijoje, 2 metodas pagal (1.2) formulę, kai FO NN sritis apibrėžiama mažiausiu galimu atstumu taip, kad joje būtų bent po 1 mokymosi aibės tašką iš kiekvienos populiacijos. 3 metodas pagal (1.2) kai FO NN sritis apima visus mokymosi aibės taškus.

Klasifikavimo metodų su BDF (2.8) ir LDF $\overline{EER}$ reikšmių santykio $\kappa=\overline{EER}^{B}/\overline{EER}^{L}$ palyginimas pateiktas 3.20 pav. Kai atstumas tarp kaimynų yra nuo 20 iki 50 m, tikimybės koeficientas artėja prie 1. Kai atstumas yra 100 m, santykis yra didesnis nei 1. Šiuo atveju LDF pagrįstas klasifikavimo metodas turi pranašumą. Kai atstumas didėja, κ indeksas mažėja ir BDF pagrįstas klasifikavimo metodas įgyja pranašumą. Taip pat lyginant skirtingus klasių žymių tikimybių vertinimo metodus, kai atstumas didėja, BDF pranašumas matomas, kai klasės žymės tikimybės skaičiuojamos pagal 2 metodą.



3.20 pav. κ reikšmių kitimas atsižvelgiant į NN atstumą.

Remiantis gautais rezultatais galima daryti išvadą, kad ankstesnės informacijos, susijusios su FO vietomis ir mokymo aibės stebėjimų vietomis, įtraukimas į modelius turi įtakos klaidos tikimybių įverčiams, o ši įtaka žymiai padidėja, kai klasės žymės tikimybės įvertinamos naudojant (1.2). 3.10 lentelėje pateikiamos $\overline{EER}$ reikšmės pritaikius perteklinių nulių auto-beta modelį, kai AER įverčiai skaičiuojami pagal BDF (2.10) ženklą testinei aibei.

3.10 lentelė $\overline{EER}$ reikšmės perteklinių nulių auto-beta modeliui.

Klasės žymių tikimybė	Artimiausių kaimynų (NN) sritis			
	1 km	3 km	7 km	all sample
Imties dydis	**0,2435**	0,2642	0,2487	**0,2435**
Atvirk. atst.su maks. min	**0,2021**	0,2383	0,2228	0,2176
Atvirk. atst. visa mok. aibe	**0,2228**	0,2487	0,2435	0,2383

Nustatyta, kad mažiausi klaidų įverčiai gaunami į modelį įtraukiant 1 km atstumu esančių kaimyninių erdvės taškų informaciją. Taip pat pastebėta, kad vertinant klasių žymių tikimybes yra tikslinga įtraukti tik artimiausių kaimynų duomenis. Remiantis gautais rezultatais galima daryti išvadą, kad kontekstinės informacijos įtraukimas į modelius turi įtakos klaidos tikimybių įverčiams. Klasifikavimas yra tikslesnis, kai klasių žymių tikimybės skaičiuojamos įtraukiant kaimyninių taškų erdvinę informaciją.

BENDROSIOS IŠVADOS

Šis darbas praplečia prižiūrimojo generatyvinio klasifikavimo metodų taikymą požymiams su erdvinės kontekstinės informacijos skirstiniais, aprašomais GRF ir ne GRF modeliais. Sukurtas BDF paremtas algoritmas, leidžiantis panaudoti prižiūrimojo generatyvinio klasifikavimo modelius požymiams, kurių skirstinys priklauso eksponentinei (Puasono, binominis, gama, beta, ZIB) ir eliptinei (*t*-skirstinio) šeimoms. Toliau pateikiamos atlikto tyrimo metu gautos išvados.

1. Siekiant praplėsti prižiūrimų generatyvinių modelių panaudojimą GRF stebinių klasifikavimui ir ištirti jų efektyvumą, atliktas EER įverčių palyginimas atskleidė BA populiacijų parametrų vertinimo metodo pranašumą prieš ML metodą, kai klasifikuojami GRF stebiniai 2 klasių atveju. Šis pranašumas yra reikšmingesnis stipriai atskirtoms populiacijoms nei artimoms populiacijoms, tiek su simetrine, tiek su asimetrine TLC. Esant asimetrinei TLC, santykio $\overline{EER}^{ML} / \overline{EER}^{BA}$ reikšmės monotoniškai didėja nuo 1,0221 iki 1,3938, kai atstumas tarp populiacijų auga nuo 0,5 iki 1,9. Rezultatai suteikia svarių argumentų, leidžiančių tikėtis, kad erdvinių populiacijos parametrų BA įverčiai gali būti veiksmingai naudojami PBDF atliekant GRF stebinių klasifikavimą.
2. Atliktas EER įverčių palyginimas, siekiant praplėsti prižiūrimųjų generatyvinių modelių taikymą GRF stebinių klasifikavimui 3 klasių atveju ir ištirti jų efektyvumą, parodė STLG pranašumą prieš STLM. Šis pranašumas yra reikšmingesnis stipriai atskirtoms populiacijoms nei artimoms populiacijoms tirtose situacijose su kovariacinėmis funkcijomis, besiskiriančiomis pločio parametru ϕ, kuris įgyja reikšmes 1, 2 arba 3. Kai $\phi = 3$, santykio $\overline{EER}^{STLG} / \overline{EER}^{STLM}$ reikšmės mažėja nuo 0,7618 iki 0,0029, kai atstumas tarp populiacijų didėja. Rezultatai suteikia svarių argumentų, leidžiančių tikėtis, kad erdvinės populiacijos STLG struktūra gali būti veiksmingai panaudota GRF klasifikavime, naudojant PBDF.
3. Palyginus sukonstruoto modelio realizavimo algoritmo TRF stebinių klasifikavimui 2 klasių atveju, kai naudojami prižiūrimieji generatyviniai modeliai, siekiant praplėsti jų panaudojimą, pritaikymo metu gautus EER įverčius, nustatytas ML parametrų vertinimo metodo pranašumas prieš LS metodą. Pranašumas didėja

didėjant atstumui tarp populiacijų, taip pat didėjant kovariacinės funkcijos pločio parametrui ϕ. Kai atstumas tarp populiacijų $\Delta = 1,5$, santykio $\overline{EER}^{ML} / \overline{EER}^{LS}$ reikšmės monotoniškai mažėja nuo 1 iki 0,8159, kai kovariacinės funkcijos pločio parametras ϕ auga nuo 0,1 iki 3,3. Rezultatai suteikia svarių argumentų, kad dažnai sudėtingi erdvinių duomenų vidurkio parametrų ML įverčiai turėtų būti naudojami labiau koreliuojantiems erdviniams duomenims, modeliuojamiems TRF, o paprastesnis LS metodas galėtų pakeisti šiuos įverčius silpnai koreliuojančių erdvinių duomenų atveju be reikšmingo PBDF našumo praradimo.

4. Modelio realizavimo algoritmo, sukonstruoto siekiant praplėsti prižiūrimųjų generatyvinių modelių panaudojimą požymių, kurių skirstinys erdvėje aprašomas auto-Puasono, auto-binominiu, auto-gama ar auto-beta modeliu, klasifikavimui 2 klasių atveju ir ištirti jų efektyvumą, pritaikymo metu gautų EER įverčių palyginimas atskleidė modelio realizavimo algoritmo, paremto BDF, pranašumą prieš algoritmą, paremtą modifikuota LDF, kai požymio skirstinys aprašomas auto-beta modeliu. Šis skirtumas statistiškai reikšmingas, kai klasės yra labiau besiskiriančios. Santykio $\overline{EER}^{BDF} / \overline{EER}^{LDF}$ reikšmės mažėja nuo 1,0018 iki 0,0262, kai skirtumas tarp regresijos parametrų skirtingoms klasėms didėja nuo 0,5 iki 49. Gauti rezultatai suteikia argumentų, leidžiančių tikėtis, kad siūlomas BDF paremtas algoritmas gali būti veiksmingai naudojamas klasifikuoti požymius, kurių skirstinys erdvėje aprašomas auto-beta modeliu.
5. Siekiant įvertinti erdvinės informacijos įtraukimo per klasių žymių tikimybių skaičiavimą efektą atliktas EER įverčių palyginimas atskleidė metodo, kai klasės žymės tikimybės skaičiuojamos naudojant atvirkštinio atstumo funkciją su apibrėžta NN sritimi (3 metodas), pranašumą prieš metodą, kai tikimybės laikomos lygiomis (1 metodas). Santykio $\overline{EER}^{3rd} / \overline{EER}^{1st}$ reikšmės mažėja nuo 0,8608 iki 0,6766. Tai taip pat patvirtino ir šakotojo banguolio dumblių (angl. black carrageen) dugno padengimo duomenų tyrimas. Sudarius perteklinių nulių auto-beta modelį, EER įverčiai gauti mažiausi, kai į FO skirstinio modelį įtraukiami 1 km atstumu esantys kaimyniniai taškai. Kai klasės žymės tikimybės skaičiuojamos naudojant atvirkštinio atstumo funkciją su apibrėžta NN sritimi, EER įvertis yra 0,2021, o kai klasės žymės tikimybės skaičiuojamos naudojant

mokymo imties dydį, jis lygus 0,2435. Rezultatai suteikia argumentų, leidžiančių teigti, kad išankstinės informacijos, susijusios su FO ir mokymo aibės vietomis, įtraukimas į modelius sumažina klasifikavimo klaidą.

SUMMARY

The dissertation is devoted to supervised generative classification algorithms for the contextual classification of data with statistical spatial dependence. Original discriminant functions based on the plug-in Bayes classification rule and Actual Error Rate (AER) expressions of these classifiers are proposed. This doctoral thesis consists of the introduction, three chapters, conclusions, bibliography, and summary in lithuanian language. The introduction presents the research and an overview of the dissertation. The first chapter is designated for related work analysis. The Contextual classification rule based on the Bayes rule, general forms of Bayes Discriminant Function (BDF) and AER, spatial data models and their descriptions by conditional density (mass) functions are submitted. Chapter 2 presents the main results of the dissertation related to the supervised generative classification algorithm based on BDF. The BDFs are constructed using the log ratio of univariate conditional density (probability) functions by incorporating the spatial context information into the data models. The Actual Error Rate (AER) expressions of these classifiers are proposed. The last chapter introduces the numerical experiments and applications.

LIST OF PUBLICATIONS

1. Zikarienė E., Dučinskas K. 2021. Application of spatial auto-beta models in statistical classification. Lietuvos matematikos rinkinys. Proceedings of The Lithuanian Mathematical Society, 62(A), p. 36- 43. https://doi.org/10.15388/LMR.2021.25219.
2. Zikarienė E., Dučinskas K. 2019 Implementation of generalized additive models for spatial beta regression. Proceedings of the XII International Conference. Computer data analysis and modelling: stochastics and data science. p. 341-343.
3. Dučinskas K., Dreižienė L., Zikarienė E. 2015. Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. Statistics and probability letters. p. 107-114. Available online 2014-12-16.
4. Dučinskas K., Zikarienė E. 2015. Actual error rates in classification of the T-distributed random field observation based on plug in linear discriminant function. Informatica. vol. 26, no. 4, p. 557-568. https://doi.org/10.15388/Informatica.2015.64.
5. Dučinskas K., Zikarienė E., Dreižienė L. 2014. Comparison of Performances of Plug-in Spatial Classification Rules Based on Bayesian and ML Estimators. Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods. p. 161-166.

TRUMPOS ŽINIOS APIE AUTORIŲ

Eglė Zikarienė gimė 1986 m. liepos 5 d. 2009 m. Klaipėdos universitete įgijo matematikos bakalauro kvalifikacinį laipsnį. 2011 m. Klaipėdos universitete įgijo statistikos magistro kvalifikacinį laipsnį. 2013-2021 m. ji buvo Vilniaus universiteto doktorantė. Nuo 2013 m. yra Lietuvos Matematikų Draugijos narė. Nuo 2011 m. dirba Klaipėdos universitete dėstytoja.

Eglė Zikarienė

Spatial Contextual Classification Based on Conditional Distributions Belonging to Elliptical and Exponential Families

Doctoral Dissertation
Natural Sciences
Informatics (N 009)
Thesis Editor: Liutauras Bartašius

Eglė Zikarienė

Erdvinis kontekstinis klasifikavimas naudojant sąlyginius eliptinių ir eksponentinių šeimų skirstinius

Daktaro disertacija
Gamtos mokslai
Informatika (N 009)
Santraukos redaktorius: Liutauras Bartašius

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p.: info@leidykla.vu.lt, www.leidykla.vu.lt
Tiražas 20 egz.