Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Mathematics

DOCTORAL THESIS

Statistical analysis of the spatiotemporal processes

Study program: P2612 - Electrical Engineering and Information Technology
Specialization: 3901V021 - Mathematical Engineering
Supervisor: doc. RNDr. Kateřina Helisová, Ph.D.
Supervisor-specialist: Ing. Mgr. Jaroslav Hlinka, Ph.D.

Prague, 2020

Anna Pidnebesna

Acknowledgements

I would like to thank my supervisors, doc. RNDr. Kateřina Helisová, Ph.D. and Ing. Mgr. Jaroslav Hlinka, Ph.D., for lots of discussions, advice, and helpful comments, and also for their patience, kindness, and friendship. Without your help, this thesis would not have been possible. I warmly thank COBRA group for support and friendship, I have enjoyed the environment I was working in. Particularly I am grateful to Pavel Šanda for helping with language and style editing. I thank David Tomeček for guidance through the world of fMRI data processing. Many thanks to my friends, Vilma Stepkova, Iegor Rafalovskyi, Diana Omelianchyk, and many others, for their help, encouragement, and sense of humor. Last but not least, I would like to thank my family for their inspiration and support, and for their everlasting love and patience.

I declare that this thesis was written by myself and using the quoted references. I agree with using this thesis for study purposes.

In Prague, 15th January 2020 Anna Pidnebesna

Abstrakt

Tato disertační práce se zabývá časoprostorovými procesy a jejich aplikacemi na dva reálné problémy. Prvním z nich je proces korespondence mezi úřady a soukromými osobami, respektive mezi úřady navzájem, v České republice, s cílem získat co nejvíce informací užitečných pro logistiku těchto úřadů. V práci se zaměřujeme jak na modelování tohoto procesu, tak na hledání závislostí mezi různými typy korespondence. Na data přitom nahlížíme jako na realizaci časoprostorového bodového procesu, přičemž ukazujeme, že známé modely jako např. Coxův proces nebo Poissonovo rozdělení nelze v této situaci použít, tudíž je potřeba využít méně standardních přístupů. Analýza pouze časové části procesu se pak provádí metodou change point, lineárním regresním modelem a vektorovým autoregresním modelem.

Druhý studovaný jev je z oblasti neurozobrazování. Neuronální aktivace různých oblastí mozku jsou popsány jako časoprostorový bodový proces. Přímé měření neuronálního signálu je však velmi obtížné, široce používanou neinvazivní technikou pro nepřímé pozorování neuronální aktivity je tak funkční magnetická rezonance (fMRI). Singál BOLD (blood-oxygen-level-dependent) měřený v fMRI je modelován jako lineární konvoluce neuronálního signálu a funkce hemodynamické odpovědi. V naší práci představujeme přístup indexace složek směsi (MCI) jakožto novou metodu pro odhad neuronálního signálu. Dále je představen softwarový nástroj BRAD vyvinutý pro odhad neuronálního signálu z neurozobrazovacích dat. Nakonec ukazujeme užitečnost aplikace MCI pro odhad efektivní konektivity v oblasti mozku.

Klíčová slova: časoprostorový proces, vícerozměrné časové řady, odhad neuronální aktivity, dekonvoluční metody, funkce intenzity, K-funkce, párová korelační funkce, bod změny, Dantzig Selector, LASSO, upravené Akaikeova informační kritéria, Bayesovská informační kritéria, efektivní konektivita, funkční zobrazování magnetickou rezonancí, hemodynamická odpověď'.

Abstract

In this thesis, the theory of spatiotemporal processes is applied to the analysis of two real-world phenomena. One of them is the process of submissions to municipalities in the Czech Republic. The aim of the analysis is to find out as much as possible of potentially useful information for municipalities functioning. We focus on the modelling of the process and finding helpful consistent patterns in the data. A wide range of approaches is used in this task. We start from modelling of the process as a spatiotemporal point pattern and show that generally used approaches, such as modelling using the Cox process, or modelling using the Poisson distribution, cannot be applied for this task. The analysis of the temporal part of the process is carried out through the change point techniques, linear regression model and vector autoregressive model.

The second studied phenomena arise in the neuroimaging field. Neuronal activations of different brain regions are described as a spatiotemporal point process. However, the direct measurement of the neuronal signal is extremely difficult. Instead, the functional magnetic resonance imaging (fMRI) is a widely used non-invasive technique for indirect observation of neuronal activity. The blood-oxygen-level-dependent (BOLD) signal measured in fMRI is modelled as a linear convolution of the neuronal signal and the hemodynamic response function. We present the Mixture Components Inference (MCI) approach as a new method for neuronal signal estimation. Also, the software tool BRAD, developed for estimation of the neuronal signal from neuroimaging data, is presented in this thesis. Moreover, we show the usefulness of application the MCI to the estimation of effective connectivity between brain regions.

Keywords: spatiotemporal point process, multivariate time series, neuronal activity estimation, deconvolution methods, intensity function, K-function, pair-correlation function, change point, Dantzig Selector, LASSO, adjusted Akaike Information Criterion, Bayesian Information Criterion, effective connectivity, functional magnetic resonance imaging, hemodynamic response.

Contents

List of abbreviations

- AIC - Akaike information criterion
- AICc - adjusted Akaike information criterion
- BC - Bayes classifier
- BIC - Bayes information criterion
- BOLD - blood oxygenation level dependent
- DS - Dantzig Selector
- EEG - electroencephalogram
- FN - false negative
- FP - false positive
- FPR - false positive rate
- GMM - Gaussian mixture model
- HRF - hemodynamic response function
- JI - Jaccard index
- LARS - least angle regression
- LASSO - least absolute shrinkage and selection operator
- LR - linear regression
- MCI - mixture component indexing
- MISE - mean integrated squared error
- MLE - maximum likelihood estimate
- MVC - model with varying concentrations
- OLS - ordinary least squares
- PD-pursuit - Primal-Dual pursuit

- PMLE - maximum pseudo-likelihood estimate
- PSD - power spectral density
- ROC - receiver operating characteristic
- ROI - region of interest
- SNCP - shot-noise Cox process
- SNS - sensitivity
- SORIS - second-order intensity re-weighted stationary
- SPC - specificity
- TN - true negative
- TP - true positive
- TPR - true positive rate
- VAR model - vector autoregressive model

Introduction

In probability theory, a stochastic process is a mathematical object, defined as a collection of random variables. These processes are widely used as models of systems that vary in a random manner. Thus, stochastic processes have applications in such disciplines in modern science as physics, biology, economy, chemistry and many others.

A stochastic process is a general notion and there is a lot of types of random processes that are different, for instance, by state space, or index set, and so on. In this work, we concentrate on a specific type of random processes, namely spatiotemporal point processes. Spatiotemporal point process is a random collection of points where each point represents the time and location of an event. In other words, it is a process of the form $\{(x_i,t_i), i=1,\ldots,n\}$ where each x_i denotes the location and t_i the corresponding time of occurrence of an event.

In this thesis, two real-world processes are studied. Both of them can be considered as spatiotemporal point processes, however, the origin of data and their measurements are very different. Therefore, the tasks concerning the datasets are distinct, which leads to the diversity of methods of data analysis. The goal of this thesis is to study the data according to particular tasks using tailored statistical and data analysis methods.

The first one of the studied processes is the socio-economical process of submissions to municipalities in the Czech Republic. One of their main attributions is the processing of different submissions like forms, documents, emails, letters, submissions sent through data boxes, etc. The information about sent and obtained correspondence of the set of municipalities is studied. The geographical coordinates of the sender, as well as the date of submission, are known. The dataset for a particular municipality typically covers daily observations during a few years. Every submission can be considered as a point that has spatial and temporal coordinates, thus, these data form a spatiotemporal point process.

There are various approaches to analysing performances of municipalities, see for instance [Hulianytskyi and Omelianchyk, 2012], [Parkinson and Roseland, 2002], [Giroux and McLelland, 2003], [López and Peters, 2010]. The data studied in this thesis were first analysed in [Lechnerová and Lechner, 2010], where the economic importance of the data for public administration is explained in detail. The authors analyse the evolution of the municipality communications in time, finding trends and seasonal components. Then in [Lechnerová and Lechner, 2013], spatial distribution and interactions of municipalities are studied. However, blank spots are still present in understanding the behaviour of the underlying process. Therefore, we got the datasets from the authors of these two articles for further analysis. The main aim of our work with this data is to find out as much as possible of potentially useful information for municipalities functioning. We focus on the modelling of the process and finding helpful consistent patterns in the data.

The second studied process comes from neuroscience. Some of the most important cells in the human brain are neurones - electrically excitable cells that communicate with each other and form an extremely complex system. It is excessively difficult to measure, analyse and model the whole brain activity on the neuronal level. Thus, in many studies, the neurones are studied at the level of regions of interest (ROI) in the brain, which are larger structures of the neurones are organised into (see, for example, [Poldrack, 2007], [Etzel et al., 2009]). We can talk about the electrical activity of a single neuron, that can be described by firing times. With some simplification, it is reasonable to discuss also the ROI activation times. Since every ROI has its separate space position in the brain, the neuronal activations of the brain regions can be considered as a spatiotemporal point process.

Information about the neuronal activity can be useful in various studies of brain structure and functioning. Specifically, a common way to characterise the brain is to estimate the connectivity of the neuronal system. Thus, the knowledge of the activations of brain regions can be used for the brain effective connectivity analysis.

However, direct measurement of the neuronal signal corresponding to a particular brain part is extremely complicated. Measuring the precise signal would require an invasive procedure. For example, intracranial electroencephalography (EEG) can be used for that. The non-invasive scalp EEG suffers from a poor spatial resolution and requires solving the inverse problem for reconstruction signals from deeper brain regions. Nevertheless, there are methods of indirect measurements, such as functional magnetic resonance imaging (fMRI), that is a useful instrument for noninvasive imaging of brain activity. The most common approach to fMRI measurement is using the Blood Oxygenation Level Dependent (BOLD) contrast. The BOLD-signal can be approximately modeled as a linear convolution of the neuronal signal with the hemodynamic response (see, for example, [Heeger and Ress, 2002]).

Despite that fMRI-BOLD signal is not directly the neuronal signal, it is widely used for the estimation of brain connectivity. Very often, the correlation of the measured signal is used for that (functional connectivity, see, for example, [Hlinka et al., 2011]). However, it has a set of well-known disadvantages. The first one is based on the fact that the BOLD signal does not measure the neuronal activity directly and there are no proofs in the literature that functional connectivity is an appropriate way to characterise the connections in neuronal system. Other problems of this method are connected with the correlation itself. For example, such a matrix has false positive links and undirected connections.

The most famous approach to brain activity and effective connectivity estimation that is aimed to solve the described problems, is Dynamic Causal Model (DCM). The newest versions of this model are described in [Frässle et al., 2017], [Yao et al., 2018]. A range of alternative approaches to effective connectivity estimation from the BOLD signal is studied. For example, the partial correlation or the precision matrix can be used instead of the correlation, and the Granger causality based methods can be useful. A detailed review can be found in [Smith et al., 2011].

Regarding the estimation of the neuronal signal itself, the paradigm-free mapping (PFM) approach was presented in [Gaudes et al., 2013]. However, the method proposed by authors is supposed to find the neuronal activation corresponding to a single stimuli during the resting state, which is a relatively strict assumption for real-world tasks.

In this thesis, we focus on an alternative approach to the neuronal activity estimation problem and discuss the effective connectivity estimates. We represent the BOLD signal as a convolution of the neuronal signal with a fixed hemodynamic response and assume sparsity of the neuronal activity.

The thesis is organised as follows. Chapter 1 includes basic theoretical notions and results used in this work. In Chapter 2, the submissions process analysis is described. First, the description of the data is given in Section 2.1. Then, the continuous and discrete domain modelling results are discussed in Section 2.2 and Section 2.3 respectively. The submission subgroups analysis is presented in Section 2.4. Chapter 3 is devoted to the neuroimaging data analysis. In Section 3.1, the problem context is discussed; Section 3.2 contains the description of the data-related questions. In Section 3.3, the newly developed method of neuronal activity estimation is presented. Section 3.4 includes the effective connectivity estimation results. In Section 3.5, the description of the developed software is given. The thesis is finishing with the Discussion and Conclusions sections, and Appendix section where the technical details are presented.

Chapter 1

Theoretical background

1.1 Space point processes

Definition 1.1.1. Let $(\Omega, \mathscr{F}, P)$ be a probability space. ***Point process*** (see for instance [Stoyan et al., 1995]) X is a measurable mapping from $(\Omega, \mathscr{F})$ to $(N, \mathscr{N})$, where N is a system of locally finite subsets of $\mathbf{R}^d$ with the σ-algebra $\mathscr{N} = \sigma(\{\mathbf{x} \in N : \sharp(\mathbf{x} \cap A) = m\} : A \in \mathscr{B}, m \in \mathbb{N}_0)$, where $\mathscr{B}$ denotes bounded Borel sets.

Definition 1.1.2. Let Y be a point process on $\Pi \subseteq \mathbb{R}^d$. Given some space M, if a random "mark" $m_\xi \in M$ is assigned to each point $\xi \in Y$, then $Z = \{(\xi, m_\xi) : \xi \in Y\}$ is called a ***marked point process*** with points in Π and mark space M.

Definition 1.1.3. A point process X on $\mathbb{R}^d$ is called ***stationary point process***, if its distribution is invariant under translations. It means the distribution of $X + s = \{x + s : x \in X\}$ is the same as that of X for any $s \in \mathbb{R}^d$.

Definition 1.1.4. Let us denote the number of points of the process X in the set A by $X(A)$. A locally finite diffusion measure μ on $\mathscr{B}$ satisfying $\mu(A) = \mathbb{E}X(A)$ for all $A \in \mathscr{B}$ is called an ***intensity measure***.

Definition 1.1.5. If there exists a function $\lambda(x)$ for $x \in \mathbb{R}^d$ such that $\mu(A) = \int_A \lambda(x)\,\mathrm{d}x$, then $\lambda(x)$ is called an ***intensity function***. If $\lambda(x) = \lambda$ is constant, then the constant λ is called ***intensity***.

Definition 1.1.6. A point process X on $\mathbb{R}^d$ is called a ***homogeneous point process***, if its intensity function is constant $\lambda(x) = \lambda$. Otherwise, X is said to be ***inhomogeneous***.

Here, I denotes the indicator function.

Definition 1.1.7. Let X be a point process on $\mathbb{R}^d$. Then

$$\alpha_k(A) = \mathbb{E}\Big(\sum_{u_1,\dots,u_k \in X}^{\neq} I\{(u_1, \dots, u_k) \in A\}\Big), A \in \mathscr{B}$$

is called the ***k-th order factorial moment measure.*** If $\alpha_k(.)$ has a density with respect to the Lebesgue measure on $(\mathbb{R}^d)^k$, then this density is called the ***k-th order intensity function of*** X and denoted by $\lambda^{(k)}$.

Remark: The 1st order intensity function $\lambda^{(1)}$ of X is the intensity function $\lambda(x)$ from Definition 1.1.5.

Theorem 1.1.1. **(Campbell's theorem.)** Suppose that X has the k-th-order intensity function $\lambda^{(k)}(u_1,\dots,u_k)$. Then for an arbitrary integrable Borel-measurable functions $h:(\mathbb{R}^d)^k \to [0,\infty)$, it holds

$$\mathbb{E} \sum_{x_1,\dots,x_k \in X}^{\neq} h(x_1,\dots,x_k) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} h(u_1,\dots,u_k) \lambda^{(k)}(u_1,\dots,u_k) \,\mathrm{d}u_1 \dots \mathrm{d}u_k.$$

Definition 1.1.8. Let X be a point process on $\mathbb{R}^d$. If there exist the first and the second order intensity functions $\lambda(u)$ and $\lambda^{(2)}(u,v)$, respectively, then the ***pair-correlation function*** (also called ***g-function***) is defined by

$$g(u,v) = \frac{\lambda^{(2)}(u,v)}{\lambda(u)\lambda(v)}, \quad u,v \in \mathbb{R}^d.$$

If the point process X is stationary with the intensity $\lambda(x) = \lambda > 0$, the pair-correlation function could be written as a function of single argument, i.e.

$$g(u,v) = \frac{\lambda^{(2)}(u,v)}{\lambda(u)\lambda(v)} = \frac{\lambda^{(2)}(0,v-u)}{\lambda^2} = g(u-v), \quad u,v \in \mathbb{R}^d.$$

Definition 1.1.9. Let X be a point process on $\mathbb{R}^d$ with the well-defined pair-correlation function $g(x,y), \quad x,y \in \mathbb{R}^d$. The process X is called the ***second-order intensity re-weighted stationary (SORIS)***, if there exists a function $g' : \mathbb{R}^d \to \mathbb{R}$ such that $g(x,y) = g'(y-x), \quad x,y \in \mathbb{R}^d$.

Definition 1.1.10. Let X be a point process on $\mathbb{R}^d$ and suppose that X has the intensity function $\lambda(x)$. Suppose that the measure

$$\mathscr{K}(B) = \frac{1}{|A|} \sum_{x,y \in X}^{\neq} \frac{I\{x \in A, y-x \in B\}}{\lambda(x)\lambda(y)}, \quad B \subseteq \mathbb{R}^d$$

does not depend on the choice of $A \subseteq \mathbb{R}^d$ with $0 < |A| < \infty$. Then X is said to be the ***second order intensity re-weighted stationary*** and K is called the ***second order reduced moment measure***.

The stationarity of X implies the second order intensity re-weighted stationarity. If the pair correlation function exists and it is invariant under translations, then we have the second order intensity re-weighted stationarity and

$$\mathscr{K}(B) = \int_B g(x) \,\mathrm{d}x, \quad B \subseteq \mathbb{R}^d.$$

Definition 1.1.11. Let X be a stationary point process on $\mathbb{R}^d$ with the intensity $\lambda > 0$. Let $A \in \mathscr{B}_0^d$ be an arbitrary set with positive Lebesgue measure $|A|$. The ***reduced second order moment function (K-function)*** is defined by

$$K(r) = \mathbb{E} \sum_{x,y \in X}^{\neq} \frac{I\{x \in A, \| x-y \| \leq r\}}{\lambda^2 |A|}, \quad r \geq 0.$$

For stationary process X with the existing $\lambda^{(2)}$, it holds that

$$K(r) = \int_{B(0,r)} g(x)\,\mathrm{d}x, \quad r \geq 0.$$

Thus $K(r) = \mathscr{K}(B(0,r))$, where $B(0,r)$ is the ball with the radius r and with the center in the origin.

Definition 1.1.12. Consider a stationary point process X_0 on $\mathbb{R}^d$ with the intensity $\lambda > 0$ and let $f : \mathbb{R}^d \to [0,1]$ be a function on $\mathbb{R}^d$. Point process X is a ***thinned version of*** X_0 with the ***retention function*** $f(x)$ ***(inhomogeneity function)***, if it is defined by

$$X = \{x \in X_0 : U(x) < f(x)\},$$

where $U(x)$ are independent, identically distributed random variables, $U(x) \sim Uniform[0,1]$, $U(x)$ are independent from X_0.

Definition 1.1.13. ***Poisson point process*** is a point process Φ, satisfying:

- for any finite collection $\{A_n\}$ of disjoint sets in $\mathbb{R}^d$, the numbers of points in these sets, $\Phi(A_n)$, are independent random variables,
- for each $A \subset \mathbb{R}^d$ such that $\mu(A) < \infty$, $\Phi(A)$ has Poisson distribution with parameter $\mu(A)$, i.e. $P[\Phi(A) = k] = \frac{\mu(A)^k}{k!} e^{-\mu(A)}, \quad k = 0,1,2,\ldots$.

Definition 1.1.14. Let $P = \{p_1, p_2, \ldots, p_n\}$, $p_i \in \mathbb{R}^2$ $\forall i = 1,\ldots,n$, $2 \leq n < \infty$, such that $p_i \neq p_j$, $i \neq j$, $\forall i,j = 1,\ldots,n$, be a set of generator points. Denote the location of the point p_i as (x_{i1}, x_{i2}), and the corresponding vector as $\vec{x}$. **Voronoi region** of p_i is the convex set

$$V_i = V(p_i) = \{\vec{x} \,|\, \|\vec{x} - \vec{x}_i\| \leq \|\vec{x} - \vec{x}_j\| \,\forall j : i \neq j\},$$

where $\|.\|$ is the usual Euclidean distance.

Voronoi tessellation of P is the union of Voronoi regions $V = \cup_{i=1}^{n} V_i$.

1.2 Spatiotemporal point processes

Definition 1.2.1. ***Spatiotemporal point process*** X is a random countable subset of $\mathbb{R}^2 \times \mathbb{R}$, where the point $(u,t) \in X$ corresponds to the point $u \in \mathbb{R}^2$ in space, occurring at the time $t \in \mathbb{R}$.

The process X is assumed to have no multiple points.

We assume the process X has the first and the second-order intensity functions λ and $\lambda^{(2)}$ such that

$$\int h_1(u,t)\lambda(u,t)\,\mathrm{d}u = \mathbb{E} \sum_{(u,t)\in X} h_1(u,t),$$

$$\int\int h_2((u,t),(v,s))\lambda^{(2)}((u,t),(v,s))\,\mathrm{d}u = \mathbb{E} \sum_{(u,t),(v,s)\in X} h_2((u,t),(v,s)),$$

for arbitrary Borel functions $h_1 : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^+$, $h_2 : (\mathbb{R}^2 \times \mathbb{R}) \times (\mathbb{R}^2 \times \mathbb{R}) \to \mathbb{R}^+$.

Definition 1.2.2. ***The spatial and the temporal projection processes*** of the process X are defined as follows:

$$X_{sp} = \{u : (u,t) \in X(W \times [0,T])\} \text{ and } X_{tm} = \{t : (u,t) \in X(W \times [0,T])\}, \text{ respectively,}$$

where $W \times [0,T]$ is an observation window.

Remark: Here, $W \subset \mathbb{R}^2$, $|W| > 0$ and $[0,T]$ is a bounded time interval, $T > 0$. It is assumed that in the point process X, any pairs (u,t) and (v,s) are distinct, i.e. $P(u \neq v) = 1$, and $P(t \neq s) = 1$, thus we may disregard multiple points in the observed projections X_{tm} and X_{sp}.

The existence of the first and the second-order intensity λ and $\lambda^{(2)}$ implies the existence of the first and second-order intensity functions for the projection processes. They could be obtained as

$$\lambda_{tm}(t) = \int_W \lambda(u,t)\,\mathrm{d}u, \quad \lambda_{tm}^{(2)}(t,s) = \int_W \int_W \lambda^{(2)}((u,t),(v,s))\,\mathrm{d}u\,\mathrm{d}v,$$

$$\lambda_{sp}(u) = \int_0^T \lambda(u,t)\,\mathrm{d}t, \quad \lambda_{sp}^{(2)}(u,v) = \int_0^T \int_0^T \lambda^{(2)}((u,t),(v,s))\,\mathrm{d}t\,\mathrm{d}s.$$

In Chapters 1 and 2, we assume the separability of the first-order spatiotemporal intensity, i.e.

$$\lambda(u,t) = \lambda_1(u)\lambda_2(t), \quad (u,t) \in \mathbb{R}^2 \times \mathbb{R}^+,$$

where λ_1 and λ_2 are non-negative functions. Then, the first-order intensity could be represented as

$$\lambda(u,s) = \frac{\lambda_{sp}(u)\lambda_{tm}(s)}{\int_{W\times[0,T]} \lambda(u,s)\,d(u,s)}.$$

1.3 Shot-noise Cox process

Definition 1.3.1. Suppose that $\Lambda = \{\Lambda(x) | x \in W\}$ is a nonnegative random field such that $x \to \Lambda(x)$ is a locally integrable function almost surely. If conditionally given Λ, X is a Poisson process on W with the intensity function Λ, then X is said to be a ***Cox process driven by*** Λ.

Definition 1.3.2. Let X be a Cox process on $\mathbb{R}^d$ driven by

$$\Lambda(x) = \sum_{(c,\gamma)\in\Phi} \gamma k(c,x),$$

where $k(.,.)$ is a kernel function (non-negative integrable function such that $\int_{\mathbb{R}^2} k(u)\,\mathrm{d}u = 1$), and Φ is a Poisson process on $\mathbb{R}^2 \times \mathbb{R}^+$ with a locally integrable intensity function. Then X is called a ***shot-noise Cox process (SNCP)***.

Definition 1.3.3. A ***spatiotemporal shot-noise Cox process*** X is a Cox process with the driving field Λ, given by

$$\Lambda(u,t) = \sum_{(r,v,s)\in\Phi} rk((u,t),(v,s)), \quad (u,t) \in \mathbb{R}^2 \times \mathbb{R}^+,$$

where Φ is a Poisson process on $\mathbb{R}^+ \times \mathbb{R}^2 \times \mathbb{R}^+$ with an intensity measure U and k is a kernel function.

Remark: The shot-noise Cox point process X is a stationary spatiotemporal process, if the measure U is of the form $U(\mathrm{d}(r,v,s)) = \gamma V(\mathrm{d}r)\,\mathrm{d}v\,\mathrm{d}s$ ($\gamma > 0$, and $V(\mathrm{d}r)$ is an arbitrary measure on $\mathbb{R}^+$, $\int_{\mathbb{R}^+} \min(1,r)V(\mathrm{d}r) < \infty$), and the kernel function $k((u,t),(v,s)) = k(v-u, s-t)$.

In case of $V(\mathrm{d}r) = \delta_1(\mathrm{d}r)$ being a Dirac measure concentrated in 1, X is a Poisson cluster process with centers coming from the stationary Poisson process on $\mathbb{R}^2 \times \mathbb{R}^+$ with the intensity γ. Then the clusters are independent conditionally on the position of cluster centers with Poisson distributed number of points. The described procedure allows us to obtain a class of Neyman-Scott processes.

1.4 Non-parametric estimates

For a homogeneous point process X on $\mathbb{R}^d$, a non-parametric estimate of the intensity λ is obtained as $\hat{\lambda} = \frac{X(W)}{|W|}$. Such estimate is unbiased. For an inhomogeneous point process, the non-parametric estimate could be obtained as the kernel smoothing with the kernel function h_b with the bandwidth $b > 0$ and the edge correction factor $\omega_{b,W}(x)$ as

$$\hat{\lambda}(u) = \sum_{x \in X \cap W} \frac{h_b(u-x)}{\omega_{b,W}(x)}, \quad u \in W.$$

Here, the edge correction factor is given by $\omega_{b,W}(x) = \int_W h_b(u-x)\,\mathrm{d}u$, so that $\int_W \hat{\lambda}(u)\,\mathrm{d}u = X(W)$. Note, that $\omega_{b,W}(x) < 1$ for all x that are laying close to the observation window boundary.

For a spatiotemporal point process X, the estimate of the first-order intensity could be obtained by

$$\hat{\lambda}(u,t) = \frac{\hat{\lambda}_{sp}(u)\hat{\lambda}_{tm}(t)}{X(W \times [0,T])}.$$

A non-parametric estimate of the K-function for the homogeneous point process and SORIS point process could be obtained as

$$\hat{K}(r) = \frac{1}{|W|} \sum_{x,y \in X \cap W}^{\neq} \frac{I\{\| x-y \| \leq r\}}{\omega(x,y)\hat{\lambda}(x)\hat{\lambda}(y)}, \quad r \geq 0,$$

where $\hat{\lambda}(.)$ is the estimate of the intensity function and $\omega(x,y)$ is an edge correction factor. In this work, we used two types of edge corrections for the estimates of K-function:

- **Border edge correction**

  $$\hat{K}(r) = \sum_{x,y \in X \cap W}^{\neq} \frac{I\{x \in W_{\ominus r}, \| x-y \| \leq r\}}{|W_{\ominus r}|\hat{\lambda}(x)\hat{\lambda}(y)}, \quad r \geq 0,$$

  where $W_{\ominus r} = W \ominus B(0,r) = \{w \in W : B(w,r) \subseteq W\}$, and $B(w,r)$ is the ball with the center in w and with the radius r, $B(w,r) = \{x \in \mathbb{R}^2 : |x-w| < r\}$.

- **Translation edge correction**

$$\hat{K}(r) = \sum_{x,y \in X \cap W}^{\neq} \frac{I\{\| x-y \| \leq r\}}{|W \cap W + x - y| \hat{\lambda}(x)\hat{\lambda}(y)}, \quad r \geq 0,$$

  where $W + x = \{w + x | w \in W\}$.

A non-parametric estimate of the pair-correlation function for the homogeneous or SORIS point process X can be computed as

$$\hat{g}(r) = \frac{1}{|W|} \sum_{x,y \in X \cap W}^{\neq} \frac{h_b(\| x-y \| - r)}{\sigma_d r^{d-1} \omega(x,y) \hat{\lambda}(x)\hat{\lambda}(y)}, \quad r \geq 0.$$

An unbiased non-parametric estimate of the K-function for a spatiotemporal point process can be calculated as

$$\hat{K}(r,\tau) = \frac{1}{|W||T|} \sum_{(u,t),(v,s) \in X}^{\neq} \frac{I\{\| u-v \| \leq r, |t-s| \leq \tau\}}{\hat{\lambda}(u,t)\hat{\lambda}(v,s)\omega((u,t),(v,s))}.$$

Here, $\omega((u,t),(v,s))$ is an edge correction factor, border or translation edge correction, as defined above. Further, the unbiased non-parametric estimates for the K-functions of the process projections are

$$\hat{K}_{sp}(r) = \frac{1}{|W|} \sum_{u,v \in X_{sp}}^{\neq} \frac{I\{\| u-v \| \leq r\}}{\hat{\lambda}_{sp}(u)\hat{\lambda}_{sp}(v)\omega(u,v)}, \quad \hat{K}_{tm}(\tau) = \frac{1}{|T|} \sum_{t,s \in X_{tm}}^{\neq} \frac{I\{|t-s| \leq \tau\}}{\hat{\lambda}_{tm}(t)\hat{\lambda}_{tm}(s)\omega(t,s)}.$$

1.5 Autoregressive and vector autoregressive model

Definition 1.5.1. Suppose that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and $\mathbb{T} \subset \mathbb{R}^K$, $K \geq 1$. A ***vector stochastic process*** is a collection of K-dimensional random vectors $\{X_t(\omega), t \in \mathbb{T}\}$ that depend on the real parameter t, where $\omega \in \Omega$ is an elementary random event. For each fixed $\omega \in \Omega$, $\{X_t, t \in \mathbb{T}\}$ is a real function called the ***realisation*** of the stochastic process.

Definition 1.5.2. A ***time series*** is a realisation $\{X_t, t \in \mathbb{T}\}$ of 1-dimentional stochastic process where index parameter t is interpreted as time. A ***multivariate time series*** is a realisation of vector stochastic process.

Definition 1.5.3. Let $\{Y_t \in \mathbb{R} : t = 1, \ldots, T\}$ be a time series. The ***p-th order autoregressive model*** $AR(p)$ is defined by

$$Y_t = b_0 + \sum_{k=1}^{p} b_k Y_{t-k} + e_t,$$

where $\{b_k \in \mathbb{R}, k = 0, \ldots, p\}$ are coefficients of AR-model, and $e_t \sim N(0, \sigma_u^2)$ is a white noise.

Definition 1.5.4. Let $\{Y_t \in \mathbb{R}^n : t = 1, \dots, T\}$ denote a multivariate time series. The ***p-th order vector autoregressive model*** $VAR_n(p)$ is defined by

$$Y_t = B_0 + \sum_{k=1}^{p} B_k Y_{t-k} + e_t,$$

where $Y_t, u_t \in \mathbb{R}^n$, for $t = 1, \dots, T$, each B_k represents a coefficient matrix of dimension $n \times n$, $e_t \sim N(0, \Sigma_u)$ is a white noise and p denotes the maximal lag length.

1.6 Sparse regression

In this section, we focus on sparse regression methods. Let us define the linear regression problem

$$y = Hs + e, \tag{1.1}$$

where y is the dependent variable ($n \times 1$ vector), H is a $n \times m$ regressors matrix, $n \geq m$, e is a $n \times 1$ vector of an additive Gaussian noise and s is a sparse $m \times 1$ vector of linear regression coefficients to be estimated.

Definition 1.6.1. ***The Ordinary Least Squares (OLS)*** estimate of an unknown vector s in (1.1) is the maximum likelihood estimate:

$$\hat{s} = (H^T H)^{-1} H^T y.$$

Definition 1.6.2. ***LASSO (least absolute shrinkage and selection operator)*** is a sparse linear regression estimator that was proposed in [Tibshirani, 1996]. It is a regression analysis method that performs variable selection by estimating the parameter vector s through the optimization problem:

$$\hat{s} = \arg\min_s \left[\lambda \|s\|_1 + \|y - Hs\|_2^2 \right], \tag{1.2}$$

where $\lambda \geq 0$ is a tuning parameter and $\|.\|_p$ is a norm in L_p-space, i.e. $\|x\|_p = \left(\sum_{i=1}^N |x_i|^p \right)^{1/p}$. Note that this is an instance of convex optimization and also of quadratic programming.

Definition 1.6.3. ***Dantzig Selector*** is a sparse linear regression estimator that was proposed in [Candes and Tao, 2007]. It is formulated as follows:

$$\hat{s} = \arg\min_s \|s\|_1, \quad \|H^T(y - Hs)\|_\infty < \lambda, \tag{1.3}$$

where $\lambda \geq 0$ is a tuning parameter, $\|x\|_\infty = \sup_n |x_n|$.

Note, that the optimization problem (1.2) can be reformulated in a form similar to Dantzig Selector:

$$\hat{s} = \arg\min_s \|s\|_1, \quad \|y - Hs\|_2^2 < \lambda_L, \tag{1.4}$$

where $\lambda_L \geq 0$ is a tuning parameter. It was shown that solution s^L of the problem (1.4) with a given parameter value λ_L is always a feasible solution for the problem (1.3) with the same parameter, i.e. $\|H^T(y - Hs^L)\|_\infty < \lambda_L$, but possibly not the optimal one. Similarity of LASSO and Dantzig Selector estimators were discussed, for instance, in [James et al., 2009], [Bickel et al., 2009], [Asif and Romberg, 2009].

1.7 Selection criteria

In the model selection context, we assume that for given data, we obtain a set of candidate models. The selection criteria inform us about the relative quality of the candidate models, generally by finding a compromise between accuracy of the data fit and model parsimony. The model with the lowest value of the selection criteria among all of the tested models would be considered the best model.

The ***Akaike Information Criteria (AIC)*** is defined as

$$AIC = -2\ln L + 2k,$$

where L is the maximised value of the likelihood function of the model, and k is the number of estimated model parameters.

The ***Bayes Infromation Criteria (BIC)*** is defined as

$$BIC = -2\ln L + k\ln N,$$

note that it also takes into account the sample size N.

In practice, the ***adjusted Akaike Information Criterion (AICc)*** is recommended to be used instead of AIC (see [Hurvich and Tsai, 1989], [Burnham and Anderson, 2004]) when k is large relative to sample size N. It is defined as

$$AIC_c = AIC + \frac{2k(k+1)}{N-k-1}.$$

Clearly, that AICc converges to AIC for $N \to \infty$.

Usage of these criteria for LASSO problems is discussed in [Zou et al., 2007]. For LASSO and Dantzig Selector, AIC and BIC have the form

$$\lambda^{AIC} = \min_{\lambda} \left[N\ln(\|y - H\hat{s}(\lambda)\|_2^2) + 2k(\lambda) \right],$$

$$\lambda^{BIC} = \min_{\lambda} \left[N\ln(\|y - H\hat{s}(\lambda)\|_2^2) + k(\lambda)\ln N \right],$$

where $\hat{s}(\lambda)$ is the solution of (1.2) and (1.3), that corresponds to λ, y is the measured signal, H is the matrix of regressors, $k(\lambda)$ is the number of non-zero parameters in $\hat{s}(\lambda)$ (see [Zou et al., 2007]). Comparison of using the AIC and BIC to Dantzig Selector in deconvolution tasks in neuroimaging was discussed in [Gaudes et al., 2013].

1.8 Deconvolution

Definition 1.8.1. ***Convolution*** $\star$ of two functions $h(.), s(.) : \mathbb{R} \to \mathbb{R}$ is defined as

$$y(t) = (h \star s)(t) = \int_{-\infty}^{\infty} h(\tau)s(\tau - t)\,\mathrm{d}\tau = \int_{-\infty}^{\infty} s(\tau)h(\tau - t)\,\mathrm{d}\tau, \quad t \in \mathbb{R}.$$

For real valued functions $h(.)$ and $s(.)$ defined on $\mathbb{Z}$, the convolution is given by

$$y(n) = (h \star s)(n) = \sum_{i=-\infty}^{\infty} h(i)s(n-i) = \sum_{i=-\infty}^{\infty} s(i)h(n-i), \quad n \in \mathbb{Z}.$$

Deconvolution is an inverse problem that is concerned with the restoration and/or reconstruction of information from known data. In terms of Definition 1.8.1, deconvolution is an algorithm-based technique for estimation of s, when y and h are known. Note, that in real-life tasks, y is usually known up to white noise e: $y = h \star s + e$. Then, deconvolution also can be used as a denoising procedure.

Definition 1.8.2. ***Toeplitz Matrix*** is an $n \times n$ matrix $H_n = \{h_{k,j}, k, j = 0, 1, \ldots, n-1\}$, where $h_{k,j} = h_{k-j}$, i.e. a matrix of the form

$$H_n = \begin{bmatrix} h_0 & h_{-1} & h_{-2} & \ldots & h_{-(n-1)} \\ h_1 & h_0 & h_{-1} & \ldots & h_{-(n-2)} \\ \multicolumn{5}{c}{\dotfill} \\ h_{n-1} & h_{n-2} & h_{n-3} & \ldots & h_0 \end{bmatrix}.$$

Note, that in discrete case for finite vectors h and s, the convolution equation $y = h \star s$, can be rewritten in the matrix form as $y = Hs$, where H is the Toeplitz matrix, corresponding to the kernel $h = (h_0, h_1, \ldots, h_{n-1})$, $h_{-i} = 0$, $i = 1, \ldots, n$. The most known deconvolution approach is the Ordinary Least Squares (OLS) estimate, which is the maximum likelihood estimate of the unknown vector s:

$$\hat{s} = (H^T H)^{-1} H^T y.$$

Chapter 2

Submissions analysis

2.1 Data description

The municipalities in the Czech Republic provide both local government and delegated powers of government. The scope of delegated powers is different for different municipalities. They are divided into three categories depending on this scope: municipalities performing basic scope of delegated powers (the first type), municipalities with authorised municipal office (the second type; performing in addition for example environment and landscape protection and including registry office and building authority), and municipalities with extended powers (the third type; including in addition for example trade licensing office). The numbers of the municipalities according to their extent of delegated powers are shown in Table 2.1.

The data stemming from the electronic records management systems kept by the municipalities are examined. The original data include the information about the date of sending or receiving the submission, addressee of a sender, agenda and way of communication. The information is anonymised by the provider so that we have no specific addresses but only postcodes (ZIP codes) at our disposal. Therefore we identify the spatial position of the communicating subject by position of an appropriate post office. The coordinate system of the unified trigonometric cadastral network (S-JTSK) converted into

Extent of delegated powers	**Number**
Total number of Municipalities in the Czech Republic	6253
Municipality of type I.	5660
Municipality of type II.	388
Municipality of type III.	205
Municipality with Registry Office	1230
Municipality with Building Authority	618

Table 2.1: The numbers of municipalities according to their extent of delegated powers, state for 1'st of January 2014, Source: Czech Statistical Office. Small Lexicon of Municipalities of the Czech Republic – 2014. Available from WWW: https://www.czso.cz/csu/czso/small-lexicon-of-municipalities-of-the-czech-republic-2014-2hzarbj7wx.

Description	Value
Amount of unique time points	2 229
The most frequent time point (amount)	38
The most frequent time point (%)	0.02%
Amount of unique space points	323
The most frequent space point (amount)	7 412
The most frequent space point (%)	37%
Total number of spatiotemporal points	20 268
Amount of unique spatiotemporal points	10 805
The most frequent spatiotemporal point (amount)	24

Table 2.2: Characterization of dataset Municipality 1.

kilometres with the shifted origin is used in order to optimise numerical calculations. Also, we consider the submissions to be divided into five groups according to the way of their processing as follows:

- type 1: electronic guaranteed (containing mainly data boxes),
- type 2: electronic non-guaranteed (containing mainly emails),
- type 3: paper forms sent by classical mail,
- type 4: personal submissions,
- type 5: others.

Within the preprocessing of the data, we omit weekends and public holidays and we check postcodes with respect to sources provided by Czech Post. Further, we exclude communication with subjects outside the Czech Republic. The amount of discarded data (including unidentified senders or addressees) is below 1.5%.

The analysis is based on a collection of 36 datasets, more precisely on data from 18 randomly chosen municipalities of different types, each containing both the sent and the received submissions. For the presentation of the results, the received submissions of one particular municipality are chosen as an example (called ”Municipality 1” in the sequel). The dataset contains 20 268 spatiotemporal points, that belong to the time interval from the 21st of January 2002 till the 30th of December 2011. It contains 2 229 working days during this interval. The maximal number of submissions per day is 38. The dataset also contains information about 323 different municipality positions (see also Table 2.2). The municipality, providing the data, (in sequel, the main office) is located at (-23.09, 6.15). According to the coordinates, the municipality is situated in Nová Ves nad Lužnicí. Note that the behaviour of the remaining time series is very similar.

2.2 Continuous domain modelling

As seen from Figure 2.1, we may expect point clusters in the process, especially in the spatial variable. Our first approach is to use an inhomogeneous spatiotemporal Neyman-Scott process (see for instance [Moller and Waagepetersen, 2004]) for modelling the cluster point process that would fit the data. The inhomogeneous spatiotemporal Neyman-Scott process it is the Poisson process driven by a random intensity function

$$\Lambda(u,t) = \sum_{(v,s)\in\Phi} \gamma f(u,t) k_1(v-u) k_2(s-t), \quad (u,t) \in \mathbb{R}^2 \times \mathbb{R}, \tag{2.1}$$

where Φ is a stationary Poisson process with an intensity $\kappa > 0$, $k_1 : \mathbb{R}^2 \to \mathbb{R}$ and $k_2 : \mathbb{R} \to \mathbb{R}$ are probability density functions, whose shape should be chosen or estimated from the data, and whose width corresponds to the typical cluster size. Also, $\gamma > 0$ is an arbitrary constant and $f : \mathbb{R}^2 \times \mathbb{R} \to [0,1]$ is an inhomogeneity function whose larger values mean higher intensity in the corresponding spatial and temporal coordinates, and vice versa. Just note that the process X is an example of a Cox process (see Section 1.3).

2.2.1 Intensity estimates

In our work, we use the estimates of the first-order intensity $\hat{\lambda}(u,t)$ for the calculating the estimate of the K-function and the pair-correlation function. Their form could help us to identify clusters in the process and also to find the typical size of the cluster. There are several methods for estimating spatial and temporal intensity (see for instance [Chiu et al., 2013], [Daley and Vere-Jones, 2003] or [Moller and Waagepetersen, 2004]). The ones we use in this section are described below.

In this chapter, the estimate of spatiotemporal intensity function is considered in the product form of estimates of spatial and temporal intensity functions (see [Prokešová and Dvořák, 2014]), i.e.

$$\hat{\lambda}(u,t) = \frac{\hat{\lambda}_{sp}(u)\hat{\lambda}_{tm}(t)}{X(W \times [0,T])}, \tag{2.2}$$

where W denotes the spatial observation window (i.e. the area of the Czech Republic), $[0,T]$ is the observed time interval, and $\hat{\lambda}_{sp}$ and $\hat{\lambda}_{tm}$ denote spatial and temporal first-order intensities, respectively.

In this section, we study a few different first-order intensity estimates $\hat{\lambda}_{sp}$ and $\hat{\lambda}_{tm}$. Our goal is to estimate the K-function using these estimates. We believe that the suitable estimate has the next property: the value of integral $\int_W \hat{\lambda}_{sp}(u)\,\mathrm{d}u$ is a satisfactory approximation of the total number of points $|X(W \times [0,T])|$. This property is used to compare different intensity estimates.

Edge corrections

Border edge correction factor is required because the observed data are given for bounded observation window, while the pattern itself is theoretically assumed to extend out of the observation window. So we need the correction to compensate the lack of information. In this work, we apply two different types of edge correction for the K-function and the pair

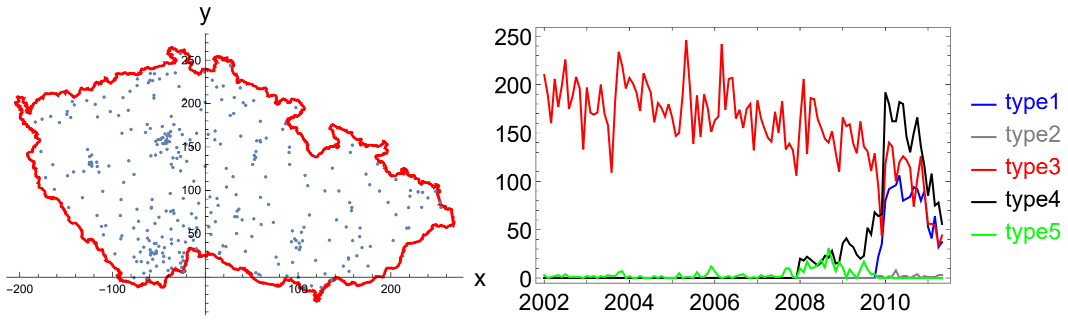


Figure 2.1: Municipality 1, received submissions. Spatial projection of all submissions (left) and temporal evolution of the numbers of particular submission types per month (right) in the time period January 2002 - December 2011.

correlation function: border and translation edge corrections. This section provides some technical details of corresponding calculations.

Border edge correction is based on the set $W_{\ominus r} = W \ominus b(0,r) = \{w \in W : b(w,r) \subseteq W\}$. However, we work with geographical boundary of the Czech Republic, so construction of W includes a problem of approximation of the 'cut' observation window $W_{\ominus r}$ for each value of r, which is a time-consuming procedure. Also, it could be time consuming to check whether a point $u \in W_{\ominus r}$ if $W_{\ominus r}$ is not a convex set. Therefore, in this work, we define W the smallest convex polygon which contains all of the process points (see Figure 2.2).

The border is constructed using the following procedure:

- From the set of space points, choose the one with the smallest y-coordinate (u_0). This point is the first vertex of the polygon.
- Choose a vertical ray, starting from u_0 and going upwards. Start rotating it clockwise, until it hits a point of the process. Denote this point by u_1.
- Continue rotating the ray from the new point u_1, until we touch the next point, denote it by u_2.
- Continue in this procedure until we have a convex polygon.

The advantages of this border are:

- It is easy to compute the convex polygon area, using the Shoelace Theorem (see for instance [Rhoad et al., 1991])

  $$A = \sum_{i=1}^{n} \frac{(x_{i+1} + x_i)(y_{i+1} - y_i)}{2},$$

  where $(x_{n+1}, y_{n+1}) = (x_1, y_1)$. Here, the (x_i, y_i), $i = 1, \ldots, n$ are points of the constructed boarder.
- It is easy to construct $W_{\ominus r}$.

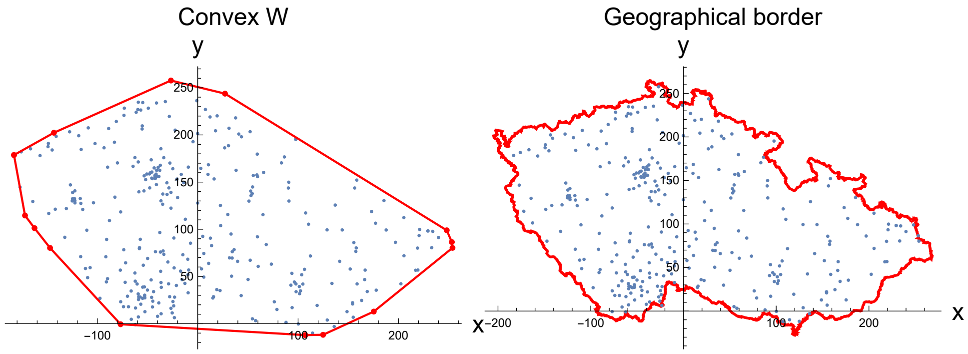


Figure 2.2: The borders of the observation window W, used for the border edge correction (left), and the geographical border, used for the translation edge correction (right).

- It is easy to check, whether an arbitrary point (x_0, y_0) lies inside the polygon $W_{\ominus r}$. If the equation of the $n-$th edge is $A_n x + B_n y + C_n = 0$, then values $A_n x_0 + B_n y_0 + C_n$ have the same sign for all n iff (x_0, y_0) lies inside $W_{\ominus r}$ and different signs for the outside points.

Translation edge correction uses an intersection area of the original observation window and the translated one $|W \cap (W+x)|$, where $W + x = \{w + x | w \in W\}$. For approximation of the area, we use the matrix M consisting of 0 and 1. Each element of this matrix $M(i, j)$ corresponds to the rectangle in geographical coordinates $(x_i, x_i + \Delta x) \times (y_i, y_i + \Delta y)$ with a known area $\delta = \Delta x \Delta y$, and the element $M(i, j) = 1$ if the corresponding part $(x_i, x_i + \Delta x) \times (y_i, y_i + \Delta y)$ at least partly lies in the Czech Republic. Otherwise, $M(i, j) = 0$. Hence, the normalized total sum of the elements of M is approximately equal to the geographical area of the Czech Republic:

$$|W| = \delta \sum_{i,j} M(i, j).$$

In this case, since M corresponds to W, the sifted window $(W + x)$ can be approximated by the matrix M_x, whose elements corresponds to the territory of Czech Republic translated on x. Then, the area of $|W \cap (W+x)|$ can be estimated as the sum of products of corresponding elements of original and shifted matrices:

$$|W \cap (W+x)| = \delta \sum_{i,j} M(i, j) M_x(i, j).$$

Spatial intensity

Kernel smoothing. The procedure of kernel smoothing is one of the methods of first-order intensity estimation.

$$\hat{\lambda}_{sp}(u) = \sum_{x \in X_{sp}} \frac{h_b^{sp}(u-x)}{\omega_{b,W}(x)}, \quad u \in W,$$

where X_{sp} is the spatial projection of the process X, $X_{sp} = \{u \,|\, (u,t) \in X\}$. Here, the edge correction factor is given by $\omega_{b,W}(x) = \int_W h_b^{sp}(u-x) \, \mathrm{d}u$. However, the obtained estimate

$\hat{\lambda}_{sp}$ depends on the chosen kernel bandwidth b, which selection is an important task because its influence on the result is relatively high.

Parametric model. The parametric model of the spatial intensity is based on the assumption that two parts play important role in the submission process intensity. The first part represents a population density in the Czech Republic, the second one assumes that intensity of mailing depends on the distance from the main office. Given this, we suggest to apply the following model of the parametric intensity:

$$\lambda(u;\beta) = \beta_0 \lambda_0(u) \max[1, \lambda_1(u;\beta)], \quad \beta = (\beta_0, \beta_1, \beta_2) > 0, \quad u \in W,$$

where $\beta = (\beta_0, \beta_1, \beta_2)$ is a parameter vector, $\lambda_0(u)$ corresponds to the population per km^2 in some region close to u, $\lambda_1(u;\beta)$ corresponds to the average number of submissions per month between the municipality, corresponding to the u coordinate, and the main municipality. One of the possible models for $\lambda_1(u;\beta)$ is a parabolic function

$$\lambda_1(u;\beta) = \beta_1 - \frac{||u-u_0||^2}{\beta_2^2}, \ \beta_1, \beta_2 > 0,$$

where u_0 represents coordinates of the main office. To estimate the parameters β, we use a maximum pseudo-likelihood estimate (PMLE), see for instance [Schoenberg, 2005], defined as

$$\hat{\beta} = \arg\max_{\beta} L(\beta), \quad L(\beta) = \exp\{-\int_W \lambda(u;\beta)\,\mathrm{d}u\} \prod_{x_i \in X_{sp}} \lambda(x_i;\beta).$$

Remarks:

- It is reasonable to maximise $\ln L(\beta)$ instead of $L(\beta)$ since

$$\ln L(\beta) = -\int_W \lambda(u;\beta)\,\mathrm{d}u + \sum_{x_i \in X_{sp}} \ln \lambda(x_i;\beta) =$$

$$= -\int_W (\beta_0 \lambda_0(u) \max[1, \lambda_1(u;\beta)])\,\mathrm{d}u + \sum_{x_i \in X_{sp}} \ln(\beta_0 \lambda_0(x_i) \max[1, \lambda_1(x_i;\beta)]) =$$

$$= -\beta_0 \int_W \lambda_0(u)[1 + (\lambda_1(u;\beta) - 1)^+]\,\mathrm{d}u +$$

$$+ \sum_{x_i \in X_{sp}} \{\ln \beta_0 + \ln \lambda_0(x_i) + \ln(\max[1, \lambda_1(x_i;\beta)])\} =$$

$$= -\beta_0 \int_W \lambda_0(u)\,\mathrm{d}u - \beta_0 \int_W \lambda_0(u)(\lambda_1(u;\beta) - 1)^+\,\mathrm{d}u +$$

$$+|X_{sp}| \ln \beta_0 + \sum_{i \in I} \ln \frac{N_i}{|W_i|} + \sum_{x_i \in X_{sp}} \ln(\max[1, \lambda_1(x_i;\beta)]),$$

where $N_0 = \int_W \lambda_0(u)\,\mathrm{d}u$, $\lambda_0(x_i;\beta) = \frac{N_i}{|W_i|}$, W_i is partition of W: $\forall x_i \in X_{sp} : \exists! x_i \in W_i$, $\sum_{i \in I} W_i = W$, and N_i is the number of people, living at W_i, i.e. $\sum_{i \in I} N_i = N_0$.

Note that $\sum_{i\in I}\ln\frac{N_i}{|W_i|}$ does not depend on β, so it does not have an influence on the maximizing.
The Riemann approximation can be used for computation of the integral

$$\int_W \lambda_0(u)(\lambda_1(u;\beta)-1)^+\,\mathrm{d}u = \sum_{i\in I}\frac{N_i}{|W_i|}(\lambda_1(x_i;\beta)-1)^+|W_i| = \sum_{i\in I}N_i(\lambda_1(x_i;\beta)-1)^+.$$

Finally,

$$\hat{\beta} = \arg\max_{\beta}[-\beta_0 N_0 + |X_{sp}|\ln\beta_0 - \beta_0\sum_{i\in I}N_i(\lambda_1(x_i;\beta)-1)^+ +$$
$$+\sum_{x_i\in X_{sp}}\ln(\max[1,\lambda_1(x_i;\beta)])].$$

- N_i is known from the data, and N_0 can be estimated as the population of the Czech Republic. For the calculations, we use $N_0 = 10\,625\,449$, that corresponds to beginning of 2018, and was taken from the website of Czech Statistical Office (https://www.czso.cz/csu/czso/domov).
- In the calculations of $\lambda_0(x_i) = \frac{N_i}{|W_i|}$ instead of the area of the "i-th-post-code territory" $|W_i|$, we use the area of a corresponding Voronoi-cell $|C_i|$ (see Definition 1.1.14). So, $\lambda_0(x_i) = \frac{N_i}{|C_i|}$. For the Voronoi tessellation, we use all post codes in the Czech Republic.

Model based on Voronoi tessellation. Another way to estimate the spatial intensity is to use the Voronoi tessellation as follows. Note, that spatial projection of the process X can be defined also as $X_{sp} = \{(x_i,y_i)\in W, i\in I_{sp}\}$, where I_{sp} is the index set of spatial points with at least one point of the pattern. In the data, there are multiple points in many spatial coordinates (x_i,y_i), therefore to each coordinates (x_i,y_i), we assign the number $n_i^{(sp)}$ of points lying in these coordinates. Then the estimate of the spatial intensity is given by

$$\hat{\lambda}_{sp}(u) = \sum_{i\in I_{sp}}\frac{n_i^{(sp)}}{|C_i|}\mathbb{I}\{u\in C_i\}, \tag{2.3}$$

where $\{C_i, i\in I_{sp}\}$ is the system of cells of the Voronoi tessellation built over the points X_{sp} on the observation window W.

Note that we got a maximum pseudo-likelihood estimate, that provides us with an interesting property that integral of the intensity over the whole observation window W is equal to the number of points, i.e.

$$I = \int_W \hat{\lambda}_{sp}(u)\,\mathrm{d}u = \sum_{i\in I}\frac{n_i^{(sp)}}{|C_i|}|C_i| = \sum_{i\in I}n_i = |X_{sp}|.$$

Temporal intensity

Kernel smoothing. As in the case of space first-order intensity, the first estimate of the temporal intensity is also done with a kernel smoothing procedure

$$\hat{\lambda}_{tm}(t) = \sum_{x\in X_{tm}}\frac{h_b^{tm}(t-x)}{\omega_{b,W}(x)},\ t\in[0,T].$$

Here, the edge correction factor is given by $\omega_{b,W}(x) = \int_W h_b^{tm}(t-x)\,\mathrm{d}t$.

Model based on the average periods

For estimation of temporal intensity, methods for estimating trends of time series (see [Green, 2011]), are employed since the data are recorded day by day. Analogously to the approach above, we denote

$$X_{tm} = \{t \,|\, (u,t) \in X\} = \{t_i \in [0,T], i \in I_{tm}\},$$

the temporal projections of the process X, where I_{tm} is the index set of temporal points with at least one point of the pattern, and assign the number of points $n_i^{(tm)}$ to each time t_i. Then we construct the periodogram, see [Green, 2011], which is a tool using for finding the importance of frequencies (periods) in the data. From this as well as from the nature of the data, it is obvious that the most important time period in the data is one month and the second important period is one week. Further, when looking at the temporal evolution of the data in detail, we can observe different behavior in different working days caused probably by the fact that the office hours of the selected municipality are Mondays and Wednesdays. Therefore we divided the days into two groups:

$$A = \{1,3\} \text{ that corresponds to } \{\textit{Monday, Wednesday}\},$$
$$B = \{2,4,5\} \text{ that corresponds to } \{\textit{Tuesday, Thursday, Friday}\}.$$

Let $t \in \{0,1,\ldots,T\}$ and consider functions $m(t)$ and $d(t)$ describing to which month and (working) day t corresponds, namely

$$m(t) = 1 \text{ if } t \text{ is a day in January}, \ldots, m(t) = 12 \text{ if } t \text{ is a day in December,} \tag{2.4}$$

and analogously

$$d(t) = 1 \text{ if } t \text{ corresponds to Monday}, \ldots, d(t) = 5 \text{ if } t \text{ corresponds to Friday.} \tag{2.5}$$

Then denote $M_k = \{t \in \{0,\ldots,T\} : m(t) = k\}$ for $k = 1,\ldots,12$ the set of days belonging to the month k, and $D_A = \{t \in \{0,\ldots,T\} : d(t) \in A\}$ and $D_B = \{t \in [0,T] : d(t) \in B\}$ are the sets of days belonging to the day group A and B, respectively. Estimate of the temporal intensity is then

$$\hat{\lambda}_{tm}(t) = y_m(t) + y_d(t), \;\; t \in [0,T], \tag{2.6}$$

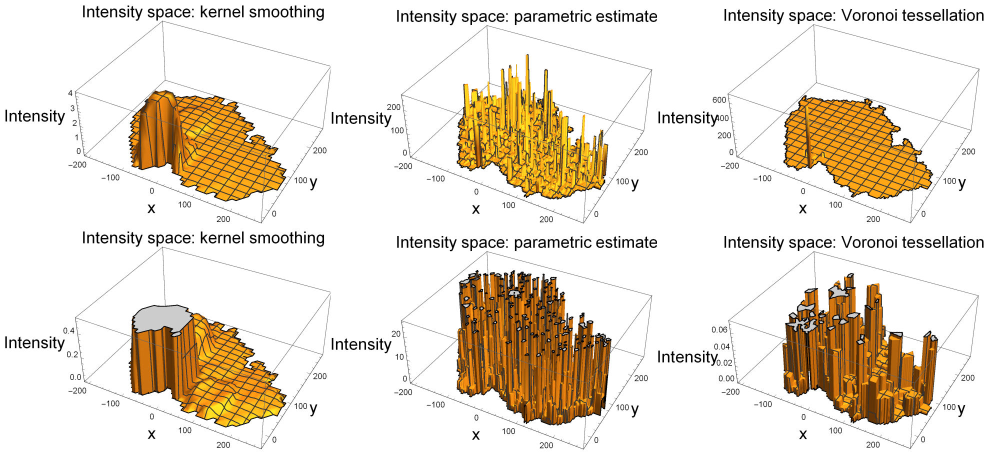
where the functions $y_m(t)$ and $y_d(t)$ are defined as

$$y_m(t) = \sum_{k=1}^{12} \frac{\mathbb{I}\{t \in M_k\}}{\sharp(\{0,\ldots,T\} \cap M_k)} \sum_{s=0}^{T} n_s^{(tm)} \mathbb{I}\{s \in M_k\},$$

where $n_s^{(tm)}$ is the number of points which time coordinate is equal to s. Here y_m denotes the average number of submitions in a single day of the month corresponding to t, and

$$y_d(t) = \frac{\sum_{s=0}^{T}(n_s^{(tm)} - y_m(s))\mathbb{I}\{s \in D_A\}}{\sharp(\{0,\ldots,T\} \cap D_A)}\mathbb{I}\{t \in D_A\} + \frac{\sum_{s=0}^{T}(n_s^{(tm)} - y_m(s))\mathbb{I}\{s \in D_B\}}{\sharp(\{0,\ldots,T\} \cap D_B)}\mathbb{I}\{t \in D_B\},$$

which can be interpreted as the mean deviation of the day t from the monthly average with respect to the type of the day A or B, respectively.



Intensity space: kernel smoothing

Intensity space: parametric estimate

Intensity space: Voronoi tessellation

Intensity space: kernel smoothing

Intensity space: parametric estimate

Intensity space: Voronoi tessellation

Figure 2.3: Comparison of different types of space intensity estimates. From left to right: kernel smoothing estimate, parametric estimate, and Voronoi tessellation estimate. The rows represent two different scales: the upper one shows full estimates with a dominating peak in the main office, the lower figures provide more details on the rest of the territory.

Numerical results

In this section, numerical characteristics of different intensity estimates are provided. Firstly the kernel smoothing estimate of space intensity is tested. The uniform kernel of width 80km is used for the estimation. The Riemann approximation of the integral of $\hat{\lambda}_{sp}(u)$ over all $u \in W$ is equal to 23 167, which is quite close to the actual point number. This occurs due to the ability of the translation edge correction to compensate for the lack of information about the process outside of the observation window. The estimate is suitable, however, while working with the kernel smoothing estimates, we noticed that the results are very sensitive to the width of the kernel, making us try another approach of intensity estimation.

Next model is the parametric one. The estimated number of points, calculated using this model, is extremely high (861 088), which makes such approach useless. The third intensity estimate is based on Voronoi tessellation. This approach has two main advantages. The first one is that Riemann integral of it over all points in observational window W is equal exactly to the real number of points, so the estimate is appropriate. The second advantage is an absence of any parameter, that should be chosen.

For the time projection, two different estimates of first-order intensity were used. The first one is the kernel smoothing estimate, that is calculated using the Epanechnikov kernel $k(u) = 0.75(1-u^2)\mathbb{I}\{|u| \leq 1\}$ of the width 50. Conclusions are very similar to the spatial case. The intensity estimate looks suitable enough due to the value of Riemann approximation of the integral $\int_{[0,T]} \hat{\lambda}_{tm}(t)\,\mathrm{d}t$, however it is dependent on the kernel width.

An alternative temporal intensity estimate is produced with the method of the average periods (grouped by working day and month). As well as the Voronoi tessellation based estimate, this one does not need any parameter selection. The comparison of the different types of intensity estimates is shown in Figure 2.3. The comparison of two estimates of the temporal intensity is shown in Figure 2.4.

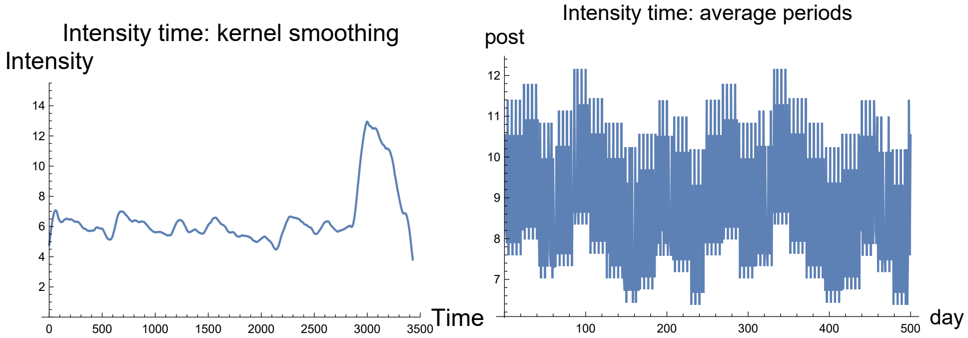


Figure 2.4: Comparison of the different types of time intensity estimates. The left picture corresponds to the kernel smoothing estimate, the right one represents the method of the average periods.

2.2.2 K-functions and pair-correlation functions

In this section, estimates of the K-function and the pair-correlation function are studied. For the smoothing kernel approach in space, these characteristics are obtained using both translation and border edge corrections. Corresponding boundaries of the observation window W are depicted in Figure 2.2. The comparison of the estimates of K-function and pair-correlation function, obtained for data, with the classical Poisson process, are shown in Figures 2.5 and 2.6. The K-function for the Poisson process is $K_{sp}(r) = \pi r^2$ for the 2D-case (space) and $K_{tm}(t) = 2t$ for 1D-case (time). The pair-correlation function for the Poisson process is also known and equal to 1 for both space and time.

The comparison of the functions estimated from the data with the Poisson process, can give more information about the real process. Namely, if the estimated K-function lies above the K-function for Poisson process, it means that data form a clusters, while the process whose K-function lies below the Poisson process' K-function is regular. The similar conclusions can be done for the pair-correlation function. Also, it is important to note that the typical sizes of clusters r corresponds to the rapid increase of $\hat{K}(r)$.

In our pictures, the estimates of the K-function and pair-correlation function for the data are shown in red color, and black color corresponds to the Poisson process. In the presented example, the uniform kernel with the width 80 is used for the space projection, and Epanechnikov kernel $k(u) = 0.75(1-u^2)\mathbb{I}\{|u| \leq 1\}$ with the width 50 is used for time projection.

As one can see in the pictures, the weakness of the border edge correction lies in a non-monotonous K-function. This occurs due to the main office being located quite close to the border, causing situation when, for some radius r, a large number of points appears out of the restricted observation window $W_{\ominus r}$. Hence, the estimate will be based only on the small number of points and becomes unreliable. The translation edge correction does not have this issue. Therefore, we chose it in the sequel.

For the both types of edge correction, the clustering in the spatial projection process X_{sp} can be seen. Both the estimates of the K-function lie above the curve of the Poisson process which means that the clusters are present in the data. The analogous pictures for the K-function and pair correlation function of the temporal projection process X_{tm} are provided in Figure 2.7.

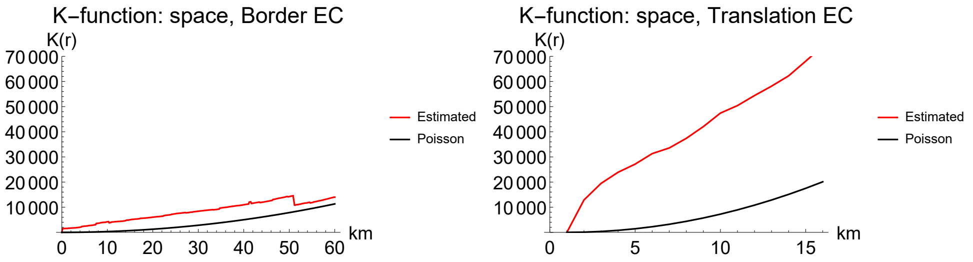


Figure 2.5: Comparison of the estimates for the K-function by different types of the edge correction. The left figure corresponds to the border edge correction, the right one to the translation edge correction. The black parabolic line correspond to the K-function of Poisson process.

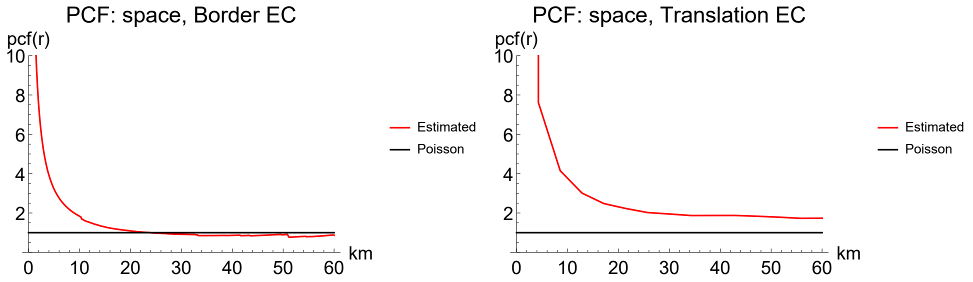


Figure 2.6: Comparison of the estimates of the pair-correlation function of space projection by different types of the edge correction. The left figure corresponds to the border edge correction, the right one to the translation edge correction.

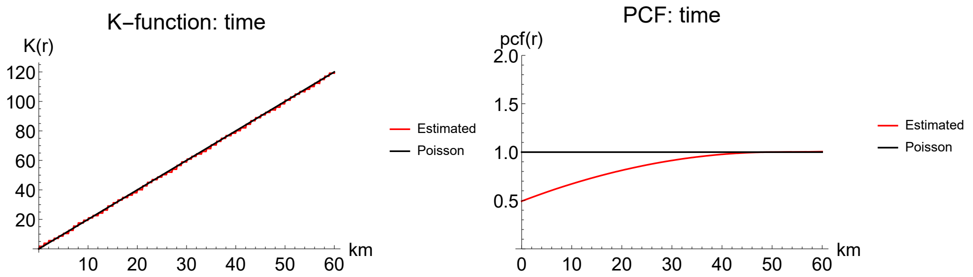


Figure 2.7: Comparison of the K-function (left) and the pcf (right) for the classical Poisson process (black) and temporal projection process (red). Intensity estimate is based on the kernel smoothing.

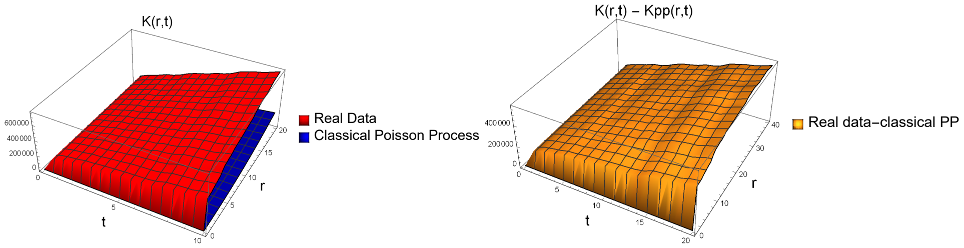


Figure 2.8: Left: spatiotemporal K-function estimated from the data (red) compared to the theoretical K-function for the Poisson point process with corresponding intensity (blue). Right: differences between these two functions. Spatial units are kilometers, temporal units are days.

For the temporal projection process, the K-function and the pair correlation function are very close to the corresponding characteristics of the classical Poisson process. K-function and pair correlation function, based on average periods estimate are similar to the estimates, based on the kernel smoothing: close to the characteristics of the classical Poisson process. However, analysing only projections and its characteristics can be insufficient, and a spatiotemporal K-function is needed in order to make a conclusion about the clustering in space or in spatiotemporal.

Finally, the estimates obtained with the Voronoi tessellation and method of the average periods are chosen as the best ones for the spatial and temporal intensities respectively. These estimates are used to get the empirical estimates of the spatiotemporal K-function.

As seen from Figure 2.8, the spatiotemporal K-function estimated from the data lies above the theoretical K-function for the Poisson process which indicates the presence of clusters in the data. However, the most important jump is in the point (0,0) which corresponds to the distance 0 km in space and 0 days in time. Thus the conclusion is that the pattern has clusters as expected, but their typical scale is less than 1 km in the space and less than 1 day in time. Since the temporal part of the data is recorded in days and spatial part is given by the coordinates of ZIP codes of the municipalities, typically located in the distance of many kilometers from each other, we cannot model the clusters themselves. The clusters are formed by events with the same spatial and temporal coordinates due to the rather rough discretization. Hence it is not possible to infer the precise scale of the clusters from the data.

2.3 Discrete domain modelling

In this section, we focus on discrete modelling. We consider the data to be a realisation of the point process placed on a discrete lattice where the points of the lattice (called knots in the sequel) are given by the cartesian product of the spatial positions (coming from a finite set) and dates of the submissions. First, we study the dependence between the spatial and temporal parts of the process, and then, we focus on modelling of the number of points in the knots of the lattice.

2.3.1 Testing independence of spatial and temporal coordinates

Information about the independence of the spatial and temporal parts of the process is of great benefit since, in the case of independence, we can analyse spatial and temporal projections independently on each other. Otherwise, we must consider them together.

Consider the hypothesis H_0 : "Spatial position of points is independent of their temporal coordinate." In order to test it, we deal with the process as couples of spatial and temporal coordinates, i.e.

$$X = \{(\xi, t), \xi \in S, t \in \{0, \dots, T\}\},$$

where S is the set of spatial positions of ZIP codes. The data form a realisation of such process represented as a collection of N points $\xi_i = (x_i, y_i)$ with the associated (random) times t_i, $i = 1, \dots, N$, so

$$X = \{(\xi_1, t_1), \dots, (\xi_N, t_N)\}.$$

For testing the null-hypothesis, we use so-called mark independence test which works as follows. Suppose that the process is characterised by some properties (for instance the number of lattice knots with more than one point etc.), called characteristics or statistics in the sequel. We simulate the a given number of permutations of times with respect to the spatial coordinates. The main idea is that in the case of independence, both the permuted patterns and the data have similar characteristics. Using mathematical terminology, denote $\{\pi(1), \dots, \pi(N)\}$ a uniform random permutation of the numbers $\{1, \dots, N\}$ and

$$\tilde{X} = \{(\xi_1, t_{\pi(1)}), \dots, (\xi_N, t_{\pi(N)})\}$$

the corresponding pattern with permuted times. Further, consider K independent uniform permutations $\pi_1, \dots, \pi_K$, which produce the patterns $\tilde{X}_1, \dots, \tilde{X}_K$. Using these patterns, we can construct the empirical 95%-confidence intervals for arbitrary statistics under the null hypothesis. Thus, if the value of the corresponding statistic calculated from the original dataset lies outside this interval, we reject the null-hypothesis and vice versa.

For our purposes, we chose the following statistics:

- T_1 – the number of knots of the lattice with at least one point,
- T_2 – the number of knots of the lattice with more than one point,
- T_3 – the average number of points in a knot in the set of knots with at least one point,
- T_4 – the average number of points in a knot in the set of knots with more than one point.

Numerical results

In order to test the independence of time and spatial coordinates, we simulate $K = 1000$ permutations and construct the 95% confidence interval for the given statistics $T_1, \dots, T_4$. The results are shown in Table 2.3. Independence of space and time coordinates was rejected, thus we conclude that there are interactions between space and time coordinates.

Statistics	$\mathbf{q_{0.025}}$	$\mathbf{q_{0.975}}$	**Data**	**Conclusion**
T_1	11 347	11 455	10 805	*Reject*
T_2	4 027	4 148	4 040	*Not reject*
T_3	1.77	1.79	1.88	*Reject*
T_4	3.14	3.19	3.34	*Reject*

Table 2.3: Testing independence between time and spatial coordinates. The second and the third column introduces the quantiles for the corresponding statistic calculated from 1000 simulated permutations, the fourth column is the value calculated from the data.

Statistics	$\mathbf{q_{0.025}}$	$\mathbf{q_{0.975}}$	**Data**	**Conclusion**
T_1	19 457	19 706	10 805	*Reject*
T_2	8 056	8 244	4 040	*Reject*
T_3	2.41	4.26	1.88	*Reject*
T_4	4.43	4.50	3.34	*Reject*
T_5	47 310	48 336	20 264	*Reject*

Table 2.4: Testing the hypothesis that the number of points in the knots of the lattice has Poisson distribution with the parameter $\hat{\lambda}(\xi,t)$ given by (2.2). The second and the third column introduces the quantiles for the corresponding statistic calculated from 100 simulations, the fourth column is the value calculated from the data.

2.3.2 Modeling the number of points by Poisson distribution

As seen from the numerical results above, we cannot work with space and time separately, so we still consider the process of points on the lattice. Now, we are interested in the distribution of the number of points in the knots of the lattice. The natural beginning of this study is testing the hypothesis that they have the Poisson distribution with the parameter given by the intensity of the corresponding knot. Thus, we consider the process represented as

$$X = \{(\xi_1,t_1,\eta_1),\ldots,(\xi_N,t_N,\eta_N)\},$$

where η_i are independent random variables having the Poisson distribution with the parameter $\hat{\lambda}(\xi_i,t_i))$, and $\hat{\lambda}(\xi,t)$ is the intensity estimate (2.2) described in Section 2.2.

In order to check the model, we apply again the empirical test based on a construction of 95% envelopes for the statistics $T_1,\ldots,T_4$ mentioned above. Since in this case, the intensity plays an important role, we moreover observe the statistic T_5 describing the total number of points.

Numerical results

In Table 2.4, there are shown the results for testing Poisson distribution of the number of points in the knots of the lattice. The confidence intervals were constructed from only 100 simulations because the calculations are more time-consuming than in the case of testing independence. It is seen that this null-hypothesis was rejected in all cases.

Statistics	$q_{0.025}$	$q_{0.975}$	Data	Conclusion
T_1	10 603	11 067	10 805	*Not reject*
T_2	3 943	4 137	4 040	*Not reject*
T_3	1.85	1.90	1.88	*Not reject*
T_4	3.30	3.39	3.34	*Not reject*
T_5	19 959	20 674	20 264	*Not reject*

Table 2.5: Testing the hypothesis that the number of points in the knots of the lattice has distribution estimated empirically for 24 combinations of the month and the type of working day (12 months and 2 types of working days) through the statistics $T_1,\dots,T_5$. The second and the third column introduces the quantiles for the corresponding statistic calculated from 39 simulations, the fourth column is the value calculated from the data.

2.3.3 Modeling the number of points by empirical distribution

In this section, we focus on the empirical distribution of number of points. We work with the collection of independent random variables $\eta'(\xi,m(t),\tilde{d}(t))$ where ξ is the spatial position, $m(t)$ is defined above by (2.4), and $\tilde{d}(t)=A$ if $t\in A$ and $\tilde{d}(t)=B$ if $t\in B$ is the type of working day. Thus for each spatial position ξ, 24 empirical distributions (corresponding to 12 months and 2 different groups of day) are to be estimated.

We use two approaches for simulation-based goodness-of-fit testing. The first one is again the construction of confidence intervals for the statistics $T_1,\dots,T_5$. The second one is construction of the confidence intervals for the following quantities:

- the mean number of points per day in a given month (for instance in January), expressed as a difference from the overall mean number of points per day (regardless of the month),
- the mean number of points per day for a given day of week (for instance for Mondays), expressed as a difference from the respective monthly averages (for instance Mondays in January contribute by the difference from the January average).

Note that the second set of quantities corresponds to the computation of $y_d(t)$ above where we consider five groups (Monday, ..., Friday) instead of two (A, B).

Numerical results

The results based on empirical distributions are represented in Table 2.5. Since the simulation is very time-consuming in this case, only 39 simulations were used. Recall that the model is based on the empirical distributions of 24 combinations of the month and the type of working day for each space point. As seen from the obtained results, it fits the data well.

In order to check the model from another point of view we also construct the confidence regions for the mean number of points per day in a given month and for the mean number of points per day for a given day of the week, as defined above, see Table 2.7 and Table 2.6. The results indicate that the fitted model describes the temporal dynamics of the data rather well.

Month	$q_{0.025}$	$q_{0.975}$	**Data**	**Conclusion**
January	-0.16	0.79	0.31	*Not reject*
February	0.17	1.33	0.69	*Not reject*
March	-0.71	0.17	-0.26	*Not reject*
April	-0.64	0.49	0.04	*Not reject*
May	0.28	1.66	1.06	*Not reject*
June	-0.39	0.76	0.35	*Not reject*
July	-0.62	0.32	-0.26	*Not reject*
August	-1.53	0.06	-0.85	*Not reject*
September	-1.11	-0.04	-0.53	*Not reject*
October	-0.47	1.13	0.29	*Not reject*
November	-0.56	0.31	-0.10	*Not reject*
December	-1.37	-0.45	-0.91	*Not reject*

Table 2.6: Testing the hypothesis that the number of points in the knots of the lattice has distribution estimated empirically for 24 combinations of the month and the type of working day (12 months and 2 types of working days). The mean number of points per day in a given month, expressed as a difference from the overall mean number of points per day, is considered here. The second and the third column gives the quantiles for the corresponding statistic calculated from 39 simulations, the fourth column shows the value calculated from the data.

Day	$q_{0.025}$	$q_{0.975}$	**Data**	**Conclusion**
Monday	1.58	2.12	2.28	*Reject*
Tuesday	-1.56	-1.00	-1.53	*Not reject*
Wednesday	1.39	2.19	1.41	*Not reject*
Thursday	-1.53	-0.95	-0.93	*Reject*
Friday	-1.45	-0.95	-1.24	*Not reject*

Table 2.7: Testing the hypothesis that the number of points in the knots of the lattice has distribution estimated empirically for 24 combinations of the month and the type of working day (12 months and 2 types of working days). The mean number of points per day for a given day of week, expressed as a difference from the respective monthly averages, is considered here. The second and the third column gives the quantiles for the corresponding statistic calculated from 39 simulations, the fourth column shows the value calculated from the data.

2.3.4 Modeling the number of repetitions by other distributions

In the analysis described above, we take into account both the spatial and temporal coordinate of the process and try to fit the distribution of the number of submissions sent from each municipality. Since it was not successful, in this section we concentrate on temporal projection only.

First, the Poisson distribution is tested. As demonstrated in the Table 2.8, for most of the municipalities, the null hypothesis about the Poisson distribution of the number of submissions is rejected. The behaviour of the number of submissions can be compared to the behaviour of the number of insurance claims. So, we use the experience of the insurance companies, where the gamma distribution is used to model the number of insurance claims, mainly because it is a fairly flexible positive-skewed distribution with convenient mathematical properties. Therefore, we try to approximate the distribution of the number of submissions per month by gamma distribution $\Gamma(\alpha,\beta)$. Since the gamma distribution with the specific parameters can be close to the normal distribution, the further hypothesis about the Gaussian distribution is tested.

The parameters for Poisson and Gaussian distributions are estimated as the average number of submissions per month and their standard deviation. The situation with the gamma distribution is different. Let us discuss the estimation of the gamma distribution parameters. The density of gamma distribution $\Gamma(\alpha,\beta)$ is

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad x > 0,$$

where $\alpha > 0$ and $\beta > 0$ are the shape and the scale parameters, respectively. The estimation of the parameters requires usage of numerical methods. More information about parameters estimation can be found for instance in [Bowman and Shenton, 1988].

A straight forward way is to use the moments method, but its efficiency is low. Another method is the maximum likelihood estimate (MLE). The equations for MLE estimates of α and β are

$$n^{-1}\sum_{i=1}^{n}\log X_i = \log\hat{\beta}_{MLE} + \psi(\hat{\alpha}_{MLE}), \quad \bar{X} = \hat{\alpha}_{MLE}\hat{\beta}_{MLE},$$

or equivalently

$$R_n = \log\hat{\alpha}_{MLE} - \psi(\hat{\alpha}_{MLE}), \quad \hat{\beta}_{MLE} = \bar{X}/\hat{\alpha}_{MLE},$$

where $\bar{X}$ is the average of the data, $R_n = \log(\bar{X}/\tilde{X})$, $\tilde{X}$ is the geometric mean of the sample data, and $\psi(\alpha) = \frac{\mathrm{d}}{\mathrm{d}\alpha}\log\Gamma(\alpha)$. However, it is known that MLE of the shape parameter is significantly biased, especially for small samples (see for instance [Choi and Wette, 1969], [Johnson et al., 1994], [Anderson and Ray, 1975]). Therefore, we use a modification of MLE.

There are exist various modifications of MLE (see for instance [Stacy, 1973], [Pradhan and Kundu, 2011], [Yanagimoto, 1988]). We use the approach described in [Zhang, 2013], which reduces the bias and improves the efficiency of the estimate. The parameters α and β are estimated as

$$\hat{\alpha}^{*}_{MLE} = \frac{n}{n+4.6}\hat{\alpha}_{MLE} + \frac{0.54}{n+4.6}, \quad \hat{\beta}^{*}_{MLE} = \hat{\beta}_{MLE},$$

Poisson distribution	In; p-value	Out; p-value
municipality1	0.0000	0.0000
municipality2	0.0000	0.0000
municipality3	0.1013	0.1151
municipality4	0.0000	0.0000
municipality5	0.0000	0.0000
municipality6	0.0000	0.0000
municipality7	0.4394	0.0223
municipality8	0.0000	0.0000
municipality9	0.0000	0.0147
municipality10	0.0000	0.0000
municipality11	0.0000	0.0000
municipality12	0.2867	0.0000
municipality13	0.0000	0.0000
municipality14	0.0001	0.0000
municipality15	0.0000	0.0000
municipality16	0.0000	0.0000
municipality17	0.0001	0.0000
municipality18	0.0000	0.0000

Table 2.8: Number of submissions per month obtained by ("In") and sent from ("Out") analysed municipalities fitted by Poisson distribution and tested by Pearson χ^2 test.

while the values 4.6 and 0.54 are obtained as the regression coefficients from the equation

$$\mathrm{E}(\hat{\alpha}_{MLE}) - \alpha \approx \frac{a\alpha}{n} + \frac{b}{n}$$

for minimising the bias of the estimate $\hat{\alpha}$.

Numerical results

We study the distribution of the number of submissions per one month without taking into account the space coordinates. Even when there are trends in evolutions of different types (see Fig. 2.1), we may suppose that their sum has approximately constant trend, so the distribution is independent of time.

For each dataset, we estimate the corresponding parameters and using the Pearson Chi-square test, we test the null-hypothesis about the Poisson, gamma and normal distributions. The results are shown in Table 2.8 for Poisson distribution, in Table 2.9 for the gamma distribution and in Table 2.10 for the normal distribution. Taking into account the number of tested hypotheses, the threshold for the p-value using the Bonferroni correction (see for instance [Armstrong, 2014], [Bonferroni, 1936]) is 0.0005. Thus, as can be seen from the results, the number of submissions per month can be fitted by normal or gamma distribution for most of the municipalities.

Gamma distribution	In; p-value	In; $\hat{\alpha}$	In; $\hat{\beta}$	Out; p-value	Out; $\hat{\alpha}$	Out; $\hat{\beta}$
municipality1	0.0000	12.67	14.70	0.0000	655.56	0.50
municipality2	0.0904	11.08	27.45	0.0134	2.84	167.23
municipality3	0.0934	4.76	2.98	0.3726	2.12	2.81
municipality4	0.0010	15.44	17.79	0.5249	8.32	33.01
municipality5	0.4398	11.02	5.15	0.1120	3.88	7.32
municipality6	0.0000	10.25	409.86	0.0000	2.21	1,700.24
municipality7	0.0133	14.78	1.70	0.1769	4.35	3.32
municipality8	0.0000	1.23	20.60	0.9161	1.04	13.74
municipality9	0.0012	1.70	25.19	0.4981	1.77	7.37
municipality10	0.1661	8.19	3.69	0.4062	2.45	8.07
municipality11	0.0000	2,106.39	0.50	0.0000	1,320.52	0.50
municipality12	0.0000	66.20	1.37	0.4159	0.47	42.17
municipality13	0.0206	4.27	5.43	0.0000	16.85	0.50
municipality14	0.0000	4.75	53.11	0.0000	17.61	19.86
municipality15	0.0296	14.65	106.18	0.0072	22.83	83.00
municipality16	0.0293	20.41	7.80	0.0454	4.14	75.27
municipality17	0.0855	12.14	6.65	0.3438	11.75	7.49
municipality18	0.1985	27.81	11.85	0.0107	1.36	87.26

Table 2.9: Numbers of submissions per month obtained by (”In”) and sent from (”Out”) the analysed municipalities fitted by gamma distribution and tested by Pearson χ^2 test.

Gaussian distribution	In; p-value	Out; p-value
municipality1	0.0000	0.2955
municipality2	0.0255	0.0000
municipality3	0.0001	0.0000
municipality4	0.6763	0.0030
municipality5	0.0024	0.0000
municipality6	0.0321	0.2073
municipality7	0.0000	0.0000
municipality8	0.8947	0.7384
municipality9	0.9703	0.1588
municipality10	0.4520	0.0000
municipality11	0.0038	0.0000
municipality12	0.7475	0.2296
municipality13	0.0000	0.0135
municipality14	0.9530	0.0010
municipality15	0.2056	0.2124
municipality16	0.3857	0.9944
municipality17	0.1487	0.1512
municipality18	0.3364	0.2964

Table 2.10: Numbers of submissions per month obtained by (”In”) and sent from (”Out”) the analysed municipalities fitted by Gaussian distribution and tested by Pearson χ^2 test.

2.4 Submissions subgroups analysis

In this part, we worked with the temporal projection of the data as multivariate time series. Each point in the dataset has its submission type, which describes the medium of communication used for submission. There are five subgroups/types of submissions: electronic guaranteed and non-guaranteed, postal, personal and other. We denoted subgroups of submissions per each medium by type1-type5 or $t_1 - t_5$ respectively. Thus, our process X_{tm} now can be denoted as $\{t_1(i), t_2(i), t_3(i), t_4(i), t_5(i),\, i = 1, \ldots, n\}$.

The main idea is the following. Since in the last years, there were changes in laws, especially the obligation to keep data boxes from the beginning of 2010, first, we detect whether there are significant changes in behaviour of the time series $\{t_1(i), t_2(i), t_3(i), t_4(i), t_5(i), i = 1, \ldots, n\}$. Then we focus on the series after the last change and using linear regression methods, we find a model of their mutual dependences. Finally, we fit the model for the number of submissions per month.

2.4.1 Change points

There is a wide class of change point detection methods and approaches (see for instance [Antoch et al., 2000], [Hawkins and Deng, 2010], [Mahmoud et al., 2007]). In our study, we test whether the parameters of linear regression describing the particular submissions types are changed during the observed period. We use a flexible change point detection technique. The method assumes that the number of submissions in time has partly linear trends. Thus, we test the null-hypothesis that there are no changes in the linear regression parameters against the hypothesis that there exists one change. The advantages of such an approach is simplicity of the method and easy calculations while the assumption of the linear trend is not too binding, because we care of the change points only and not of the trend itself. In general, the null hypothesis is formulated as

$$H_0 : Y_i = a + bx_i + e_i, \quad i = 1, \ldots, n,$$

while the alternative hypothesis is

$$\begin{aligned} H_A : \exists m \in \{2, \ldots, n-2\} : Y_i &= a + bx_i + e_i, \quad i = 1, \ldots, m, \\ Y_i &= a^0 + b^0 x_i + e_i, \quad i = m+1, \ldots, n, \end{aligned}$$

where $a \neq a^0$ or $b \neq b^0$ are the coefficients of linear regression, $Y = \{Y_1, \ldots, Y_n\}$ is the dependent variable, $x = \{x_1, \ldots, x_n\}$ is the explanatory variable and the noise $e = \{e_1, \ldots, e_n\}$ is formed by i.i.d. random variables with the mean $\mathrm{E}e_i = 0$ and the variance $\operatorname{var} e_i = \sigma^2 > 0$. In our case, the dependent variable $Y = t_j,\ j = 1, \ldots, 5$, is used for the numbers of particular submission types and the explanatory variable x is time [in months], i.e. $x_i = i$, $i = 1, \ldots, n$.

Denote

$$X_k = \begin{pmatrix} 1 & x_1 \\ \ldots & \ldots \\ 1 & x_k \end{pmatrix}, \quad X_k^o = \begin{pmatrix} 1 & x_{k+1} \\ \ldots & \ldots \\ 1 & x_n \end{pmatrix}$$

and

$$\chi_k^2 = \frac{1}{\sigma^2}(\hat{a}_k - \hat{a}_k^0, \hat{b}_k - \hat{b}_k^0)\big((X_k^T X_k)^{-1} + (X_k^{0^T} X_k^0)^{-1}\big)^{-1}(\hat{a}_k - \hat{a}_k^0, \hat{b}_k - \hat{b}_k^0)^T,$$

where $\hat{a}_k, \hat{a}_k^0, \hat{b}_k$ and $\hat{b}_k^0$ are estimates obtained by least squares under the alternative with $m = k$, and A^T denotes transposition of matrix A. Then under the null-hypothesis, for the maximum-type test statistics, it holds (see [Antoch et al., 2000]) that

$$P\left(\max_{2\le k\le n-2}\{\chi_k^2\} > \left(\frac{x+b_n}{a_n}\right)^2\right) \approx 1-\exp\{-e^{-x}\}, \quad x \in \mathbb{R}^1,$$

where

$$a_n = \sqrt{2\log\log n} \quad \text{and} \quad b_n = 2\log\log n + \frac{1}{2}\log\log\log n.$$

In the case that we reject the null-hypothesis, we are interested in estimation of the change point m and the parameters of linear regression. The estimate of the change point m is defined as

$$\hat{m}_{regr} = \arg\min\{\sum_{i=1}^{k}(Y_i - \hat{a}_k - \hat{b}_k x_i)^2 + \sum_{i=k+1}^{n}(Y_i - \hat{a}_k^0 - \hat{b}_k^0 x_i)^2; \quad k = 1,\dots,n\},$$

where $\hat{a}_k$, $\hat{b}_k$ and $\hat{a}_k^0$, $\hat{b}_k^0$ are least squares estimates of a, b and a^0, b^0 based on $Y_1,\dots,Y_n$.

Note that this approach is suitable for detecting and estimating parameters of a model with not more than one change point. However in our case, we would like to obtain more change points, because we focus on whether there is a change at the beginning of 2010, independently on whether the change is the most important. There are methods that allow to detect more change points in the time series (see e.g [Antoch et al., 2000] or [Eichinger and Kirch, 2018]), but they are based on very time-consuming calculations. Therefore, we use the following simplification: for all time series, we detect three change points using the basic algorithm three times – for whole observed period, and for the parts before and after the first detected change point. We are aware that we eliminate the situations when both the second and the third change points lie before or after the first change point, respectively, but this simplification is satisfactory for our purposes.

Numerical results

Change point detection is applied to all types of submissions. The results for Municipality 1 for the most important types, i.e. for type 1 and type 3, in the whole time period from January 2002 to December 2011 are shown in Fig. 2.9.

Further in this section, we deal only with the time period from January 2007 to December 2011, since we have not enough data for types 1, 2, 4 and 5 till December 2006 (see Fig. 2.1). For each type, we observe the proportions of municipalities (in %) in which the change point occurred in a given quarter, while we consider all municipalities together. As seen from Table 2.11 and Fig. 2.10, the most frequent change point appears in the third quarter of 2009 for types 1 and 3. It can be explained by establishing data boxes, because in this period, electronic submissions signed by an advanced electronic signature (type 1) started to repress the classical post (type 3) and personal submissions (type 4).

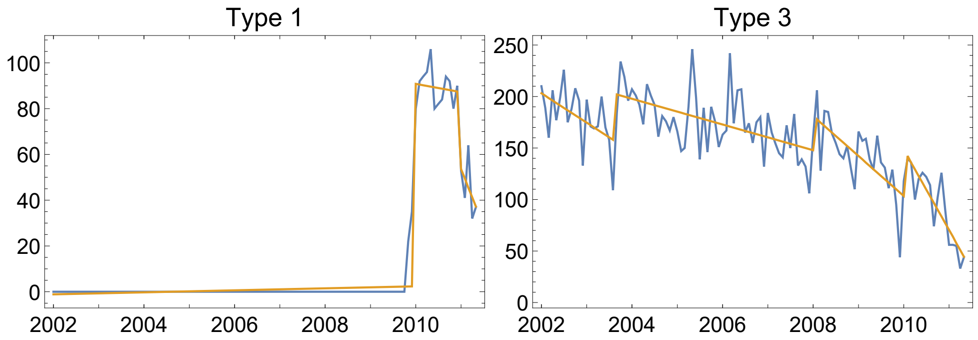


Figure 2.9: Linear approximation (orange line) of the number of submission of the types 1 and 3 (blue line) after application of change points detection.

Changes ratio	Type 1	Type 2	Type 3	Type 4	Type 5
1 quarter 2007	0 %	0 %	5 %	13 %	10 %
2 quarter 2007	14 %	7 %	14 %	12 %	0 %
3 quarter 2007	17 %	13 %	0 %	6 %	0 %
4 quarter 2007	0 %	0 %	13 %	7 %	17 %
1 quarter 2008	0 %	0 %	0 %	13 %	0 %
2 quarter 2008	14 %	0 %	17 %	6 %	8 %
3 quarter 2008	20 %	12 %	8 %	0 %	0 %
4 quarter 2008	0 %	7 %	0 %	0 %	0 %
1 quarter 2009	11 %	6 %	8 %	10 %	0 %
2 quarter 2009	0 %	0 %	12 %	0 %	0 %
3 quarter 2009	100 %	6 %	37 %	5 %	8 %
4 quarter 2009	11 %	5 %	13 %	17 %	7 %
1 quarter 2010	21 %	9 %	3 %	4 %	8 %
2 quarter 2010	7 %	0 %	6 %	3 %	0 %
3 quarter 2010	16 %	0 %	9 %	13 %	0 %
4 quarter 2010	12 %	0 %	15 %	14 %	12 %
1 quarter 2011	12 %	4 %	12 %	3 %	0 %
2 quarter 2011	3 %	0 %	6 %	6 %	7 %
3 quarter 2011	5 %	7 %	0 %	0 %	13 %
4 quarter 2011	0 %	0 %	0 %	0 %	0 %

Table 2.11: The proportions of municipalities (in %) in which a change point occurred in the given quarter (for all municipalities together) in the time period from January 2007 to December 2011.

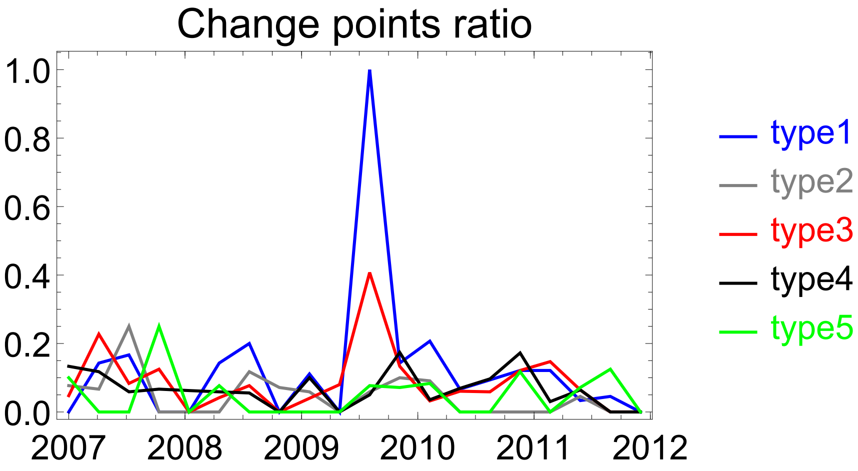


Figure 2.10: Ratios of change points in the data from January 2007 to December 2011.

2.4.2 Linear analysis of type dependencies

The second task is to find dependencies among the particular types. For this purpose, we first calculate the correlations and consequently, we try to make linear analysis in which different types are employed. All calculations are done for two time periods: for the whole time period and after the beginning of 2010, when all municipalities had to establish data boxes.

The correlation coefficient for each pair of types is calculated by the Spearman rank correlation (see [Zar, 2005]) instead of classical Pearson correlation because the data do not have normal distribution and may include outliers. The Spearman rank correlation is defined as the Pearson correlation coefficient between the ranked variables:

$$r_s = \frac{\operatorname{cov}(rg_X, rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}},$$

where X and Y are random variables, rg_X and rg_Y are corresponding ranked variables, $\operatorname{cov}(rg_X, rg_Y)$ and $\sigma_{rg_X}, \sigma_{rg_Y}$ are covariance and standard deviations, respectively, of the ranked variables.

Finally, we provide a model of dependencies among the types. As mentioned above, we use the linear regression. Since the most important changes came with the obligation of data boxes, which belong to the type 1, we focus on modelling dependence of the type 1 on the other types while we observe its dependence on the most correlated types.

Numerical results

The result from Subsection 4.1 led us to the idea that the type 1 and type 3 are dependent. However, it may lead to other dependencies among the types, too. Therefore, we first calculate the correlation coefficients. The coefficients for the whole time period from January 2002 to December 2011 are introduced in Table 2.12. It is seen that most of them are statistically significant according to the Spearman rank correlation test at 1% level. Indeed, the coefficient of correlation between the types 1 and 3 is negative which confirm our idea

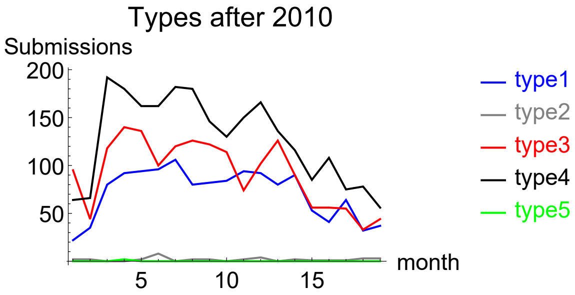


Figure 2.11: The number of all submission types per month from January 2010.

	Type 1	Type 2	Type 3	Type 4	Type 5
Type 1	1				
Type 2	0.627506	1			
Type 3	-0.556731	-0.516369	1		
Type 4	0.951819	0.592018	-0.588156	1	
Type 5	-0.261033	-0.196822	0.00386004	-0.144088	1

Table 2.12: Spearman's rank correlations of submission types in the time period January 2002 - December 2011. Significant correlations are marked by blue color.

that establishment of data boxes caused increase of the number of electronic submissions signed by an advanced electronic signature and decrease of the number of physical submissions. The coefficients for the time period from January 2010 to December 2011, i.e. for the period after establishing data boxes, are different, see Table 2.13. In this period, the significant correlation coefficients are only that ones between the types 1 and 3, 1 and 4, and 3 and 4. Moreover, comparing to the whole time period, the most of the correlation coefficients are positive which implies similar behaviour of the mentioned types.

For better imagination of the behaviour, we draw the scatter plots of significantly correlated pairs of types of submissions, see Fig. 2.12. Since the type of dependence is not very clear, we try to fit the model of linear dependence. We focus on modelling the dependence of type 1 on type 3, further on the dependence of type 1 on types 3 and 4 and finally, we compare it to the model of dependence of type 1 on all the types 2-5, i.e. we fit the models $t_1 = at_3 + b$, $t_1 = at_3 + bt_4 + c$ and $t_1 = at_2 + bt_3 + ct_4 + dt_5 + e$, respectively. The fitted models are

$$t_1 = 0.55t_3 + 23.24, \tag{2.7}$$

$$t_1 = 0.23t_3 + 0.45t_4 + 14.13, \tag{2.8}$$

$$t_1 = 1.30t_2 + 0.31t_3 + 0.25t_4 - 3.58t_5 + 9.98. \tag{2.9}$$

However, according to 95%-level confidence intervals, only the first model has the coefficients significantly different from zero, while the corresponding adjusted coefficient of determination is $R^2 = 0.68$. Comparing of the first model to the data is graphically shown in Fig. 2.13.

	Type 1	Type 2	Type 3	Type 4	Type 5
Type 1	1				
Type 2	0.0113386	1			
Type 3	0.824392	-0.206803	1		
Type 4	0.840155	-0.0509917	0.873085	1	
Type 5	0.187405	-0.234895	0.333468	0.271567	1

Table 2.13: Spearman's rank correlations of submission types in the time period January 2010 - December 2011. Significant correlations are marked by blue color.

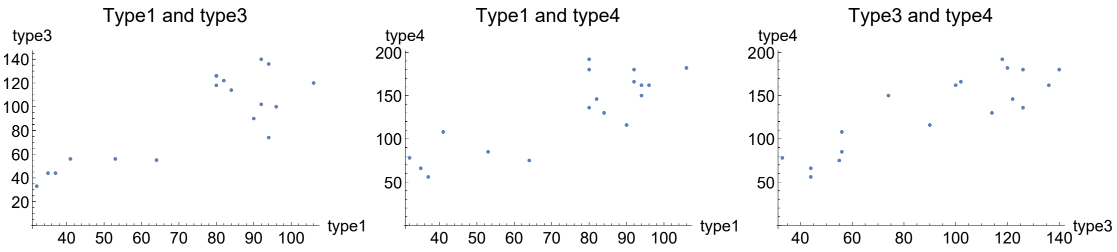


Figure 2.12: Scatter plots of all combinations of the correlated types 1, 3 and 4 in the time period January 2010 - December 2011.

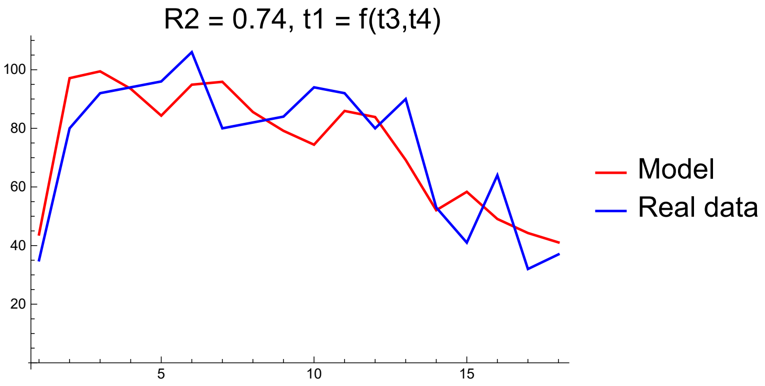


Figure 2.13: Comparing the data (blue line) and the fitted models (red line) of dependence of the type 1 on the type 3 in the time period January 2010 to December 2011.

2.4.3 Vector autoregressive models

Let us recall that the process is denoted as $\{Y_i = (t_1(i), t_2(i), t_3(i), t_4(i), t_5(i)), i = 1, \dots n\}$. The main goal is to build a suitable model for the submissions of the type 1, i.e. electronically guaranteed ones, using data on previous values of this and another submission types. We focus on the electronically guaranteed submissions because it is a new type, appeared after the law changes in 2010, and it is important to understand its dynamics itself and in connection with the other types.

In this section the vector autoregressive approach (see Definition 1.5.4) is applied to model the type 1 submissions. We are working with the first equation from the system of five equations of VAR(p) model, and denote aforementioned type 1 model as VAR1(p).

In our case, p is the number of months before the time point i from the model, and coefficients can be interpreted as impact indicators of lagged types 1-5 on the function $t_1(i)$. These coefficients are estimated using two methods. The first one is the ordinary least squares approach (OLS, see Definition 1.6.1), the second one is using the Dantzig Selector method for estimation of the sparse matrix of the coefficients (see Definition 1.6.3). The adjusted Akaike Information Criteria was used as the selection criteria for Dantzig Selector. Note, that the VAR-system can be rewritten as the linear regression system of equations as

$$y = \mathrm{vec}(Y) = (L' \otimes I_K)\beta + \mathrm{vec}(E), \tag{2.10}$$

where $\otimes$ is a Kronecker product, $Y := (Y_1, \dots, Y_T)$, $L_t := (Y_t, \dots, Y_{t-p+1})$, $L := (L_0, \dots, L_{T-1})$, $E := (E_1, \dots, E_T)$ is noise, and β is the vector of coefficients, that is to be estimated.

While using the VAR(p) models, the number of models parameters needed to be estimated is $K^2 p$ coefficients for complete VAR(p) model with K variables. Based on the data nature, it can be reasonable to try the VAR(1), VAR(3) and VAR(12) models (according to month, a quarter of a year, and year period in submission system). However, to estimate such number of parameters, the data should be much longer. Thus, we are working only with VAR(1) model.

As the model quality characteristics, the adjusted coefficient of determination is used:

$$R^2 = 1 - \frac{SS_{res}/(n-k)}{SS_{tot}/(n-1)},$$

where SS_{res} and SS_{tot} are residual and total sum of squares, respectively, n is the number of observations, and k is the number of explanatory variables in the model (not including the constant term). It is important to note, that the models build on the whole data, but the R^2 coefficient is calculated only for the period after establishing the data boxes.

Numerical results

Our dataset includes 113 time points and $K = 5$ variables. As mentioned above, the data are too short for estimation the VAR model with the order higher than 1. Thus, we calculate the VAR(1) model for 5 variables using OLS and Dantzig Selector+AICc (DSA) approach. The principal difference between the approaches is getting a dense matrix for the OLS and a sparse one for the DS. The obtained matrices of the coefficients are shown in Figure 2.14 as an image for better visualisation. The obtained models for all variables are:

$$t_1^{OLS}(i) = 0.47t_1(i-1) + 1.99t_2(i-1) - 0.001t_3(i-1) + 0.26t_4(i-1) - 0.32t_5(i-1),$$

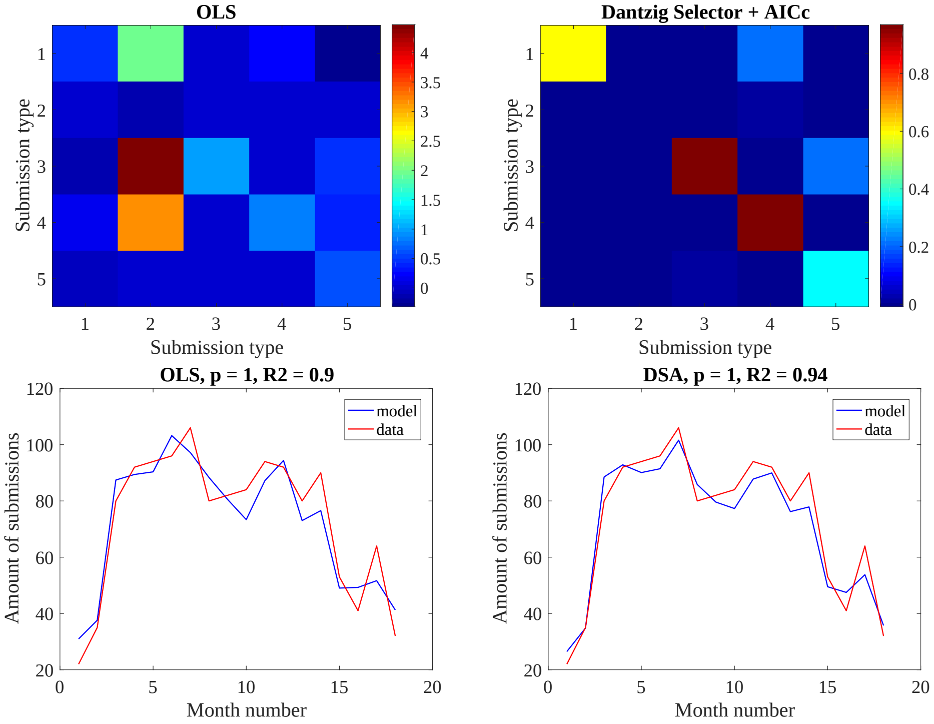


Figure 2.14: Type 1. VAR(1) model that includes all submission types. In the top panel, the matrix of coefficients are presented, in the lower ones, the corresponding comparison of the model and real data for type 1 are shown.

$$t_1^{DSA}(i) = 0.59t_1(i-1) - 0.01t_3(i-1) + 0.22t_4(i-1).$$

As seen from Figure 2.14, the main diversity between the models is the inclusion of types 2 and 5. The sparse model does not include them, while in the full model they are involved with relatively high coefficients. This can be explained by the low number of submissions of those types and their low impact on type 1, and other types as well. Thus, it makes sense to build a VAR model including only the 1st, 3rd, and 4th types and compare the results with a model including all the variables. For three variables (types 1,3,4) the matrices of coefficients are presented in Figure 2.15 and the obtained models are:

$$t_1^{OLS}(i) = 0.6t_1(i-1) - 0.005t_3(i-1) + 0.22t_4(i-1),$$
$$t_1^{DSA}(i) = 0.45t_1(i-1) - 0.008t_3(i-1) + 0.30t_4(i-1).$$

Note that the coefficient of determination for DSA is higher than for OLS for the model including all of the variables. It is because the model is constructed on all of the data, but the corresponding metrics is calculated only for the period, where type 1 is non-zero. As seen from the comparison of the coefficient of determination, it is useful to include in the model only three variables, instead of five. It does not make the model worse and the number of estimated parameters are much less. Also, we would like to stress, that an advantage of the VAR model is the possibility to model all of the types at the same time. An example of type 3 model is presented in Figure 2.16.

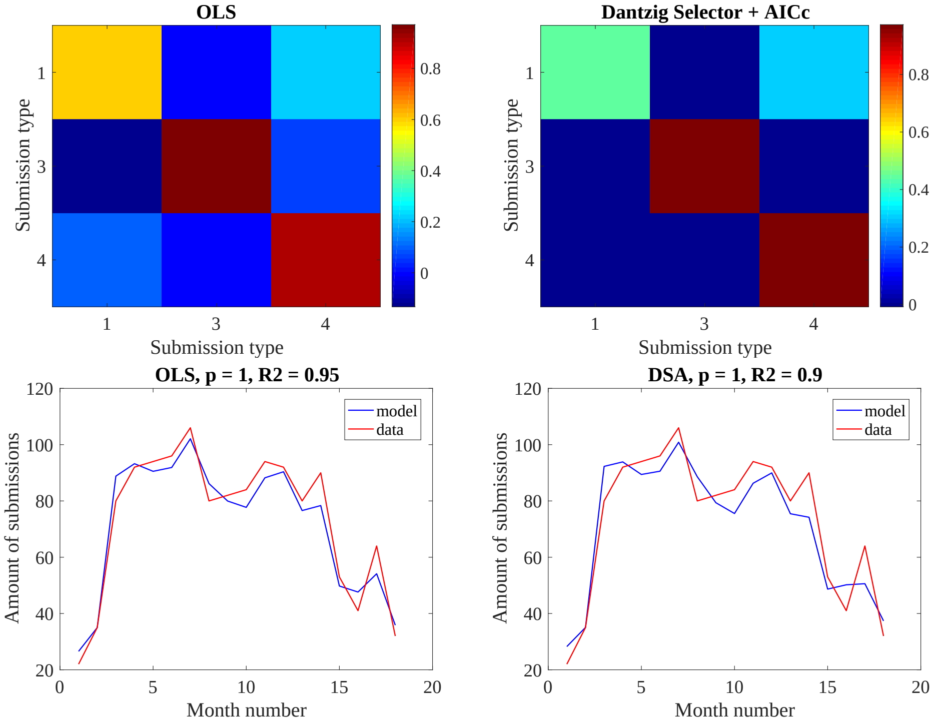


Figure 2.15: Type 1. VAR(1) model that includes 1, 3 and 4 submission types. In the top panel, the matrix of coefficients are presented, in the lower ones, the corresponding comparison of the model and real data for type 1 are shown.

Finally, we use the coefficient of determination to compare the linear regression model described in Section 2.4.2 with the VAR(1) model. The comparison of the linear and vector autoregressive models for the other municipalities is shown in Table 2.14. There are shown results for modelling type 1 using type 3 and type 4 (and type 1 for VAR). Only the datasets, that have at least 5 submissions of type 1 per month are included. Also, for VAR models we exclude too short datasets (less than 3 points for one parameter). Based on the results, presented in the table, we can speculate that VAR tends to give better models, even though we have too few datasets to make final conclusions.

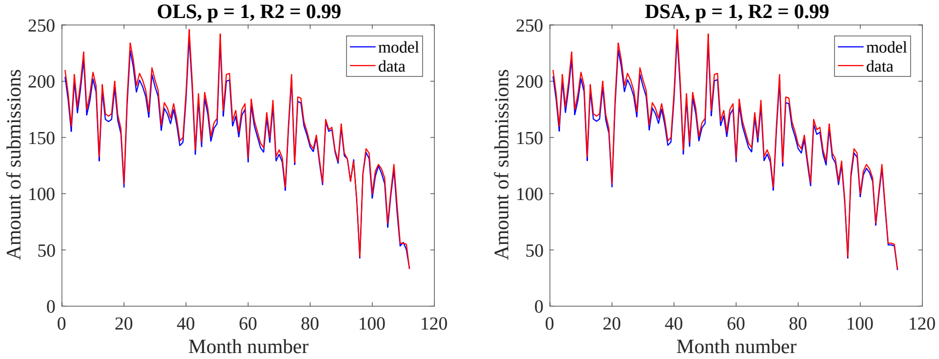


Figure 2.16: Type 3. VAR(1) model that includes all submission types, comparison of the model and real data for type 3 is shown.

	LR		VAR(1)	OLS	VAR(1)	DSA
Municipality	In	Out	In	Out	In	Out
municipality1	0.7830	0.12	0.91	0.77	0.91	0.77
municipality2	0.4939	0.20	0.62	0.70	0.62	0.70
municipality4	0.3038		0.21	-0.07	0.21	-1.88
municipality5	0.4235		0.15		0.15	
municipality6	0.2466	-0.01				
municipality8	-0.1095					
municipality9	-0.0040	0.08				
municipality10	0.3016		0.35		0.31	
municipality11	0.0105	0.65	0.58	0.83	0.58	0.83
municipality12	-0.2061	-0.24				
municipality14	-0.0675	0.16				
municipality15	0.3194	0.21	-0.26	-0.73	-0.26	-2.36
municipality16	0.1810	0.03				
municipality17	0.2337	-0.04				
municipality18	0.0484	0.01	-0.05	-0.13	-0.10	-0.20

Table 2.14: Comparison of the adjusted coefficient of the determination for VAR(1) models (OLS and DSA) and linear regression (LR). Only type 3 and 4 are included for modelling (and type 1 for VAR). In this table, the missing values mean that the corresponding time series were too short to fit the model.

Chapter 3

Brain activity analysis

3.1 Context of the problem

Several techniques can be used for the measurement of human brain activity in the neuroimaging field. The most known are electroencephalography (EEG), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), brain positron emission tomography (PET). All of these methods can be applied in specific conditions and have its benefits and restrictions. For example, the most popular and affordable is EEG that records directly the electrical activity of the brain and can detect changes over milliseconds. However, it has such disadvantages as, for instance, low spatial resolution, poor signal-to-noise ratio or poorly measured activity that occurs below the upper layers of the cortex. In contrast to EEG, the fMRI technique is a useful instrument for noninvasive imaging of brain activity that has a high spatial resolution (typically in mm), but a low temporal resolution (typically in seconds). The pictures of EEG recording setup and MRI scanner are shown in Figure 3.1.

In this chapter, we concentrate on data obtained with the fMRI. The most common approach to fMRI measurement is the use of the Blood Oxygenation Level Dependent (BOLD) contrast [Ogawa et al., 1990], which was shown to reflect the changes in the local brain activity due to human mental operations [Ogawa et al., 1992]. From an engineering point of view, BOLD fMRI provides an indirect measurement of neural activity. BOLD response can be modeled as a linear convolution of the hemodynamic response function

Figure 3.1: An EEG recording setup (the left panel) and MRI scanner (the right panel). National Institute of Mental Health, Klecany, Czech Republic.

(HRF) with an input signal corresponding to the underlying neuronal activity.

In many experimental situations, timing of the brain activation is assumed to be known *a priori*. In particular, when a reliable experimental stimulation of brain activity is applied. However, the *a priori* assumption of knowing the temporal profile of brain activation is a very strong one and in many cases unrealistic. This has motivated the development of methods that allow the estimation of activation time series without such prior information.

As long as the BOLD-signal can be viewed as a convolution of a neuronal signal and a hemodynamic response function, increases in neuronal activity can be expected to be represented in the BOLD-signal as delayed and smoothed activations. Therefore, to obtain an estimate of the underlying brain activity, one may attempt to revert the operation of the convolution kernel; in other words, to deconvolve the BOLD-signal in order to estimate the neuronal activity. Of course, this requires at least an approximate knowledge of the hemodynamic response function.

In [Gaudes et al., 2011] the authors present a method for mapping the brain's response to a single stimulus in space and time without prior knowledge of the paradigm timing, which they refer to as 'paradigm-free mapping' (PFM). This method is based on a deconvolution of the measured time series, which assumes a given linear hemodynamic response, and uses the ridge regression algorithm (see [Hoerl and Kennard, 1970]). In [Gaudes et al., 2013], an alternative sparse regression technique is proposed for the use in paradigm-free mapping. In particular, a promising new approach is based on the Dantzig selector estimator, solved via a homotopy procedure, together with the statistical model selection criteria. The methods presented in [Gaudes et al., 2011] and [Gaudes et al., 2013] are implemented in the software 3dPFM for AFNI, that is designed to identify brief BOLD events in resting state fMRI data. It is based on the linear convolution model and deconvolves a HRF for each fMRI voxel and estimates the neuronal-related signal.

Furthermore, in [Wu et al., 2013], the authors propose a blind hemodynamic deconvolution to recover the neuronal signal from the resting state of fMRI based on an extracted region-specific HRF. An improvement of the mentioned procedural data is reported in [Wu and Marinazzo, 2015]. In [Aggarwal et al., 2015], the authors propose a voxel-based method for joint estimation of the underlying activity signal and of the HRF. In [Sreenivasan et al., 2015] the authors present a nonparametric deconvolution method based on homomorphic filtering to estimate the neuronal signal from the fMRI signal.

So far, only the activity of one brain region was discussed. But measuring the whole brain opens wide opportunities to more general brain investigation. The common way to characterise brain activity is to estimate the connectivity of the neuronal system and analyse its graph properties. Brain connectivity in fMRI is a widely discussed topic. For the connectivity estimation, the functional connectivity (correlation of measured BOLD signals) is the most known approach. For recovering also the direction of the interactions, effective connectivity methods such as the Granger causality are recommended to be applied instead.

However, the connectivity structure estimated from the BOLD signal is not equivalent to the one of the original neuronal signal. Thus, to obtain the correct causality network, the analysis should be performed on the estimated source signal. The detailed discussion of this topic can be found in [Valdes-Sosa et al., 2011], [Friston, 2011], [Friston, 2009].

3.2 Data

In this chapter, we study the BOLD-signals measured using the fMRI. We assume that the data has the form of a collection of BOLD-signals, each of them is an average signal of one ROI. We estimate the neuronal activity for each ROI separately, and then use obtained data to estimate the effective brain connectivity.

3.2.1 Data model

We measure the BOLD signal denoted by y, which is represented as a vector of the length N. The BOLD signal is modeled as a result of a hemodynamic response to the local brain activity. We assume that the neuronal activity can be described by a sparse vector of activations. We do not assume any additional information about stimuli timing. In mathematical notation we assume that the measured signal is a result of the convolution of neuronal activity and the HRF:

$$y = h \star s + e, \quad y(n) = \sum_{i=0}^{L-1} h(i)s(n-i) + e(n), \quad n \in \{1,..,N\}. \tag{3.1}$$

Here $\star$ denotes the convolution, y, s, e are $N \times 1$ vectors denoting the original signal, the brain activity and the noise, respectively. The vector h denotes the HRF, its length is $L < N$. In the described model, we assume that e is a white noise, which is independent and identically distributed in each time point. In the matrix form, the model can be rewritten as

$$y = Hs + e, \tag{3.2}$$

where H is the Toeplitz convolution matrix $N \times N$ corresponding to the HRF h (see [Gray, 2006]).

In practice, the BOLD signal values output by the MRI scanner have a generally arbitrary scale and offset. Due to potential spontaneous neuronal activity fluctuations, establishing a proper baseline corresponding to 'no substantial activity' is a methodologically and even conceptually problematic task. Indeed, some level of ongoing 'baseline' activity can be expected, while we are aiming to detect the deviations from this baseline. As estimating the baseline value from the data itself also poses additional challenges, we have taken the heuristic approach of shifting the signal so that its relative range (minimum and maximum) around the zero value corresponds to that of the convolution kernel; alternatively just re-scaling the signal to have minimum 0 and maximum 1 gives comparable results. This is practically equivalent to assuming that the underlying (sparse) neuronal signal contains only *positive* activations (albeit of unknown amplitude). Note that such assumption is common in the field and often explicitly used in the model and optimization procedure definitions, see e.g. [Hernandez-Garcia and Ulfarsson, 2011, Bush and Cisler, 2013]. Of course, delineating the limits of both its conceptual and practical validity would require much deeper discussion of the nature of brain activity.

3.3 MCI

3.3.1 Basic idea: Bayes classifier

The main idea of our proposed approach is to utilize the information available in the observed data and the derived signals (namely the ordinal least square estimate of the brain activity and the family of regularized estimates) within a Bayesian inference tool that would separate the true activations from false positives due to observational noise.

The starting point is a basic estimate of the brain activity $\xi = (\xi_1, \dots, \xi_N)$, $\xi_j \in \mathbb{R}$ (think of e.g. the OLS solution $\hat{s}_{OLS}$), that is typically not a sparse vector. We assume, that its (non-zero) elements correspond to activations caused by one of two possible reasons: neuronal response (true activation) or noise (false activation). Then we try to apply a naive Bayes classifier for two classes on the vector ξ.

Now, let us discuss the Bayes classifier in more detail. Let $g : \mathbb{R} \to \{1, \dots, M\}$ be the function (classifier), which returns the number of mixture component (class) $g(\xi) \in \{1, \dots, M\}$ for each possible value of the observed characteristics $\xi \in \mathbb{R}$. In our case, there are two classes, i.e. $M = 2, \quad g(\xi_j) \in \{1, 2\}$, where $g(\xi_j) = 1$ corresponds to the conclusion 'j-th element falls to a true activation component', and $g(\xi_j) = 2$ corresponds to the conclusion 'j-th element falls to a false detection component'. The Bayes classifier is of the form

$$g^B(\xi_j) = \arg\max_{m=1,2} p_j^m f^m(\xi_j),$$

where $j \in \{1, \dots, N\}$ indexes the classified elements of the vector, $\xi_j \in \mathbb{R}$ are the basic (OLS) estimate values, p_j^m are the probabilities for j-th activation to fall to the m-th class, and f^m denotes the probability densities of the characteristic ξ, provided it falls within the class m. Thus, the classifier $g^B(.)$ for each j-th activation compares the posterior probabilities of the element belonging to a neuronal activation or noise. Details and properties of such a classifier can be found, for instance, in [Rish, 2001] and [Hand and Yu, 2001].

However, true values of the probabilities p^m and the densities $f^m(x)$ are usually unknown *a priori*. Instead, we may need to obtain some estimates $\hat{p}^m$ and $\hat{f}^m(x)$ from data. We suggest two approaches to obtain such estimates. First, by using the Gaussian mixture model, described in [McLachlan and Peel, 2000]. It is based on the assumption of the normal distribution of all mixture components. This approach is well-known, and its main advantage is a developed theory and existence of software tools for the automatic estimation of all the parameters of the model. For our purposes, it is important to estimate both of the needed elements, probability vectors $\hat{p}^m$ and component densities $\hat{f}^m(x)$.

The second approach is based on utilizing a model with varying concentrations (MVC) for estimating the component densities, and a heuristic procedure based on LASSO/Dantzig Selector algorithms for approximating the probabilities. Using this concept enables us, not only to disregard the assumption about the normality of the component distributions, but it also allows for utilizing element-specific prior probabilities – that are in the present scenario informed by the regularized regressions solutions. The detailed description of this concept is provided below and its block-scheme is shown in the left panel of Figure 3.2.

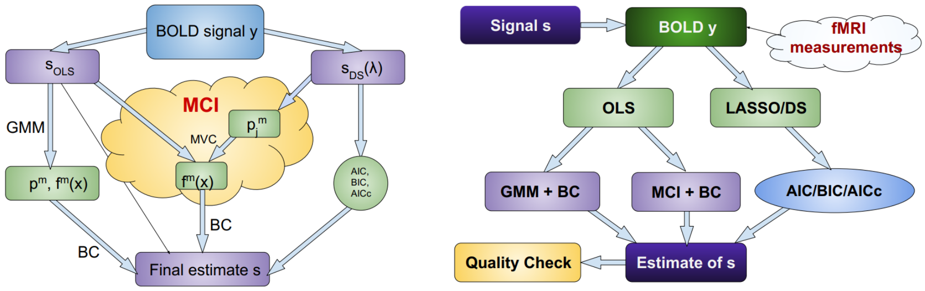


Figure 3.2: Left panel: Block-scheme of proposed MCI approach. Right panel: scheme of simulations.

3.3.2 MVC-based density estimates

In the classical finite mixture model, the mixing probabilities, and therefore the probability densities of all features ξ_j, are supposed to be the same for all the observed objects j ([McLachlan and Peel, 2000]):

$$f_j(\xi_j) = f(\xi_j) = \sum_{m=1}^{M} p^m f^m(\xi_j).$$

As a generalization of this approach, the model with varying concentrations (MVC) was proposed ([Maiboroda and Sugakova, 2012]). According to this model, mixing probabilities $p_j^m, m \in \{1,\dots,M\}, j \in \{1,\dots,N\}$, vary for different objects j, while still satisfying the condition: $\forall j : 0 \le p_j^m \le 1, \sum_{m=1}^{M} p_j^m = 1$. Then the distribution for the j-th object can be modelled as

$$f_j(\xi_j) = \sum_{m=1}^{M} p_j^m f^m(\xi_j).$$

In MVC, the mixing probabilities p_j^m are supposed to be known, whereas the densities of components f^m are assumed to be unknown and estimated from the data. For estimation of the parameters, we use the kernel estimate

$$\hat{f}^m(x) = \frac{1}{b_m N} \sum_{j=1}^{N} a_j^m K\left(\frac{x-\xi_j}{b_m}\right),$$

where $K(.)$ is a kernel function and b is the kernel width and a_j^m are the weights, aimed to distinguish the m-th component and suppress the influence of other components on it (see [Maiboroda and Sugakova, 2012] for details). The parameters are obtained using estimates developed in [Maiboroda, 2008] and [Doronin and Maiboroda, 2015]. Classification of components of mixture are discussed in [Sugakova, 2006] and in [Autin and Pouet, 2012].

In particular, when one denotes the inner product of two vectors $a = (a_1, ..., a_N)$, and $b = (b_1, ..., b_N)$ by $\langle a, b\rangle = \frac{1}{N}\sum_{i=1}^{N} a_i b_i$ and defines the matrix

$$\Gamma_N = (\langle p^l, p^m \rangle_N)_{l,m=1}^M.$$

Note that this matrix is further assumed to be invertible, with γ_{km} being the km-minor of Γ_N. Then the weights are given by:

$$a_j^k = \frac{1}{\det \Gamma_N} \sum_{m=1}^{M} (-1)^{m+k} \gamma_{km} p_j^m.$$

For the specific case of two-components, the weights a_j^m are given by:

$$a_j^1 = \frac{(1 - \langle p^1, 1 \rangle) p_j^1 + (\langle p^1, p^1 \rangle - \langle p^1, 1 \rangle)}{\langle p^1, p^1 \rangle - (\langle p^1, 1 \rangle)^2},$$

$$a_j^2 = \frac{\langle p^1, p^1 \rangle - \langle p^1, 1 \rangle p_j^1}{\langle p^1, p^1 \rangle - (\langle p^1, 1 \rangle)^2}, \quad j = 1, ..., N.$$

The kernel width b_m is estimated using the Silverman's rule-of-thumb (see, for example, [Hardle et al., 2006]):

$$b_m \approx 1.06 \hat{\sigma}_m N^{-1/5},$$

where $\hat{\sigma}_m$ and $\hat{\mu}_m$ are the estimated standard deviation and mean of the m-th component given by $\hat{\sigma}_m^2 = 1/(N-1) \sum_{j=1}^N a_j^m (\xi_j - \mu_m)^2$, and $\hat{\mu}_m = 1/N \sum_{j=1}^N a_j^m \xi_j$, respectively.

3.3.3 Probability estimates

As it was described above, for the classification of signal values s_j, we also need to set the probability p_j^m that it belongs to the true or false activations class. In this section, we propose a heuristic procedure based on the numerical solutions of the optimisation problems with a regularisation parameter. Particularly, we work with the homotopy algorithms for LASSO (1.2) and Dantzig Selector (1.3) estimators.

Let us denote $s(\lambda)$ such solution of (1.2) (or (1.3)) that corresponds to a parameter's value λ. Then, the support of $s(\lambda)$ and its sign vector can be defined as

$$\mathscr{B}(\lambda) = \operatorname{supp} s(\lambda) = \{j | 1 \le j \le N, s(\lambda)_j \ne 0\},$$

$$sgn(\lambda) = \{sgn(s(\lambda)_j), j = 1, \dots, N\}.$$

For both mentioned estimators, the particular solution $s(\lambda)$, its support $\mathscr{B}(\lambda)$ and sign vector $sgn(\lambda)$ are specified by the parameter value λ. For the LASSO problem, it was shown in [Efron et al., 2004], that for a given response vector y, there is a finite sequence of λ's, called transition points (see [Zou et al., 2007]),

$$\lambda_0 > \lambda_1 > \cdots > \lambda_K = 0, \quad \text{such that:}$$

- for all $\lambda > \lambda_0$, solution of the optimization problem (1.2) corresponds to the trivial solution, i.e. its support is the empty set $\mathscr{B}(\lambda) = \emptyset$;
- the value $\lambda_K = 0$ corresponds to the ordinary least squares solution;

- for λ values from the interval $I_m := (\lambda_{m+1}, \lambda_m)$, the solutions of (1.2) are constant with respect to λ in terms of their supports and sign vectors;

$$\mathscr{B}_m := \mathscr{B}(\lambda), \quad sgn_m := sgn(\lambda) \quad \forall \lambda \in I_m.$$

In other words, there is a sequence $\mathscr{B}_m$, $m = 0, \dots, K$ of subsets of indices $j, j = 1, \dots, N$ that correspond to nonzero coordinates of $s(\lambda)$ for $\forall \lambda \in I_m$. Thus, for each $j = 1, \dots, N$ there is a set of intervals I_m, where the j lies in the corresponding solution support $\mathscr{B}_m$:

$$\mathscr{I}_j := \cup_{\{m: j \in \mathscr{B}_m\}} I_m = \cup_{\{m: j \in \mathscr{B}_m\}} (\lambda_{m+1}, \lambda_m).$$

Further, let us denote the union of all intervals I_m as $\mathscr{I} := \cup_{m=0}^{K} I_m = (0, \lambda_0)$. Then we propose to estimate the probability ω_j that j-th activation comes from the true neuronal signal as

$$\hat{p}_j^1 = \frac{|\mathscr{I}_j|}{|\mathscr{I}|} = \frac{\sum_{m=0}^{K-1} (\lambda_m - \lambda_{m+1}) \mathbb{I}\{j \in \mathscr{B}_m\}}{\lambda_0}.$$

This approach corresponds to the idea, that more frequent appearance of non-zero activation estimate in the solution sets means higher probability for this moment to correspond to a true neuronal activation. While the probabilistic interpretation of this 'relative occurence in the solution set' is clearly heuristic, it is reasonably motivated as the LASSO operator is assumed to detect true variables among the noise.

Numerically, the set of transition points and corresponding sets of solutions for LASSO can be obtained by the homotopy algorithm LARS, presented in [Efron et al., 2004]. For Dantzig Selector, there is an available homotopy algorithm Primal-Dual Pursuit, which was presented in [Asif and Romberg, 2009]. Note that the Primal-Dual Pursuit produces a finite sequence of so-called critical values of the parameters, that includes the points of changes in the support or sign vectors of the solutions; without loss of generality, these critical values can be considered as the transition points.

Thus, the probability p_j^1 for j-th activation to come from true activation, can be approximated based on the numerical solution of the optimisation problem with regularisation parameters. Then, the probability $p_j^2 = 1 - p_j^1$, and the densities $f^1(.), f^2(.)$ can be estimated. Finally, this information is enough to build the classifier and separate the components of vector ξ on true and noisy activations.

3.3.4 Simulation study

To assess the standard methods as well as the newly proposed methods, we use numerical simulations. We generate a signal $y = s \star h + e$ using HRF h, while parametrically varying the the properties of the signal s and noise e. Then, LASSO and Dantzig Selector with AIC/AICc/BIC selection criteria are used to estimate the input signal s. These estimates compared with two approaches based on Bayes Classifier: GMM and the proposed method MCI. To assess the quality of the obtained estimates, several measures are used. General scheme of simulations is shown in right panel of Figure 3.2.

The HRF kernel is obtained using the function *spm_hrf(.)* in the SPM toolbox (The Wellcome Dept. of Cognitive Neurology, University College London), setting scan repeat time (RT) to 2.5s. To define the level of added noise e, we use the Signal-to-Noise Ratio

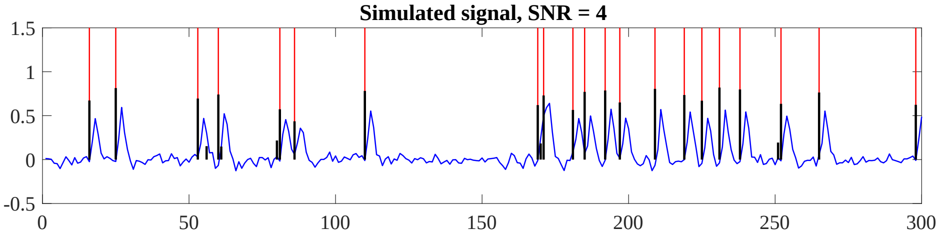


Figure 3.3: Example of a simulated and estimated signal. The simulated neuronal signal is shown in red, the simulated measured BOLD signal is shown in blue, the estimated neuronal signal is shown with thick black lines.

(SNR), $SNR = \frac{\sigma_{s\star h}}{\sigma_e}$, where $\sigma_{s\star h}$ and σ_e are standard deviation of BOLD signal and noise, respectively. For our simulations, we use several values of SNR in range $[1.5, 4.5]$.

The input signal s was simulated as a vector of the length N, that consists of K non-zero activations (value 1) and $N-K$ zero values. Positions of non-zero elements are chosen uniformly in the range $[1,\dots,N]$.

The value of the selection criteria depends on the length of the signal N and number of non-zero elements K. Thus, we explore the estimates for several different noise levels and multiple numbers of true non-zero coefficients.

We denote the number of true negatives as TN, true positives as TP, false negative and false positive as FN and FP, respectively. To assess the model quality, we use several measures: Jaccard index, sensitivity, and specificity. The Jaccard index ([Jaccard, 1901]) is a similarity measure, defined as

$$JI = \frac{TP}{TP+FP+FN}.$$

JI takes values in the range $[0,1]$, where 1 corresponds to perfect agreement. This similarity coefficient was also discussed in [Gower, 2004], [Kosub, 2016], or [Willett et al., 1998].

Sensitivity, or true positive rate, is the probability of a positive test decision given that the true value is positive. Specificity is a probability that a test result will be negative when the condition is not present (true negative rate):

$$SNS = TPR = \frac{TP}{TP+FN}, \quad SPC = 1 - FPR = \frac{TN}{FP+TN},$$

SNS and SPC take values in range $[0,1]$, where 1 corresponds to perfect specificity/sensitivity.

Numerical results

In this section, we present the numerical results of the simulation study. Figure 3.3 shows an example of a simulated signal and the obtained estimates. The red peaks are marks of the non-zero components of the simulated signal, the blue line is a noisy convolution of the corresponding simulated vector, thick black peaks are the obtained estimates (by the MCI DS method).

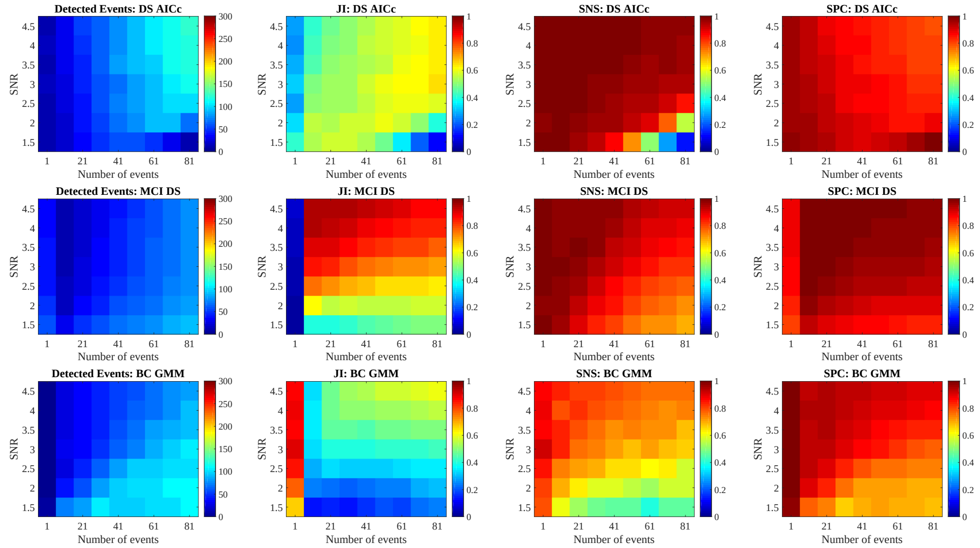


Figure 3.4: Comparison of the deconvolution methods on simulated data on the number of detected events, Jaccard index (JI), sensitivity (SNS), and specificity (SPC). Time series length is $N = 300$, the number of non-zero peaks is shown on the x-axis, the noise level on the y-axis. Estimation methods are shown in different rows: DS AICc, MCI DS and BC GMM.

Our simulations show that DS/LASSO with AIC/BIC tends to mark almost all possible elements as non-zero, in other words, extremely over-estimates the signal. Figures that demonstrate corresponding results can be found on Figure 3.5. Thus, further results in this section are presented for DS AICc, MCI DS, and Bayes classifier based on GMM components (BC GMM).

Figure 3.4 shows the averaged results for 100 simulations for varying noise levels and activation density. The matrix contains 7 noise levels, increasing from top to bottom on the y-axis, and a different number of peaks on the x-axis (from the smallest to the highest values).

Each row of the Figure corresponds to one of three compared methods: DS AICc, MCI DS and BC GMM. Every column shows one of the characteristics of the estimates: number of detected non-zero events, Jaccard index, sensitivity and specificity. Clearly, BC GMM works comparably good only for very sparse signals. Also, this method tends to over-estimate the signal for higher noise.

The other two methods show generally better results. DS AICc tends to slightly over-estimate the number of detected events for higher SNR and less sparse signals. Otherwise, the method performs relatively well and gives stable results for the presented parameters. The Jaccard index is typically about 0.6, which is much better than the results of BIC or uncorrected AIC.

Concerning MCI DS, it fails for extremely sparse signal (only one activation from 300 points) for any SNR, which happens due to inability to estimate the distribution of true activations using only one point. However, for other parameters settings it shows good results. In particular, the quality of obtained estimates by MCI DS is even better than that of DS AICc, mainly in terms of the Jaccard index.

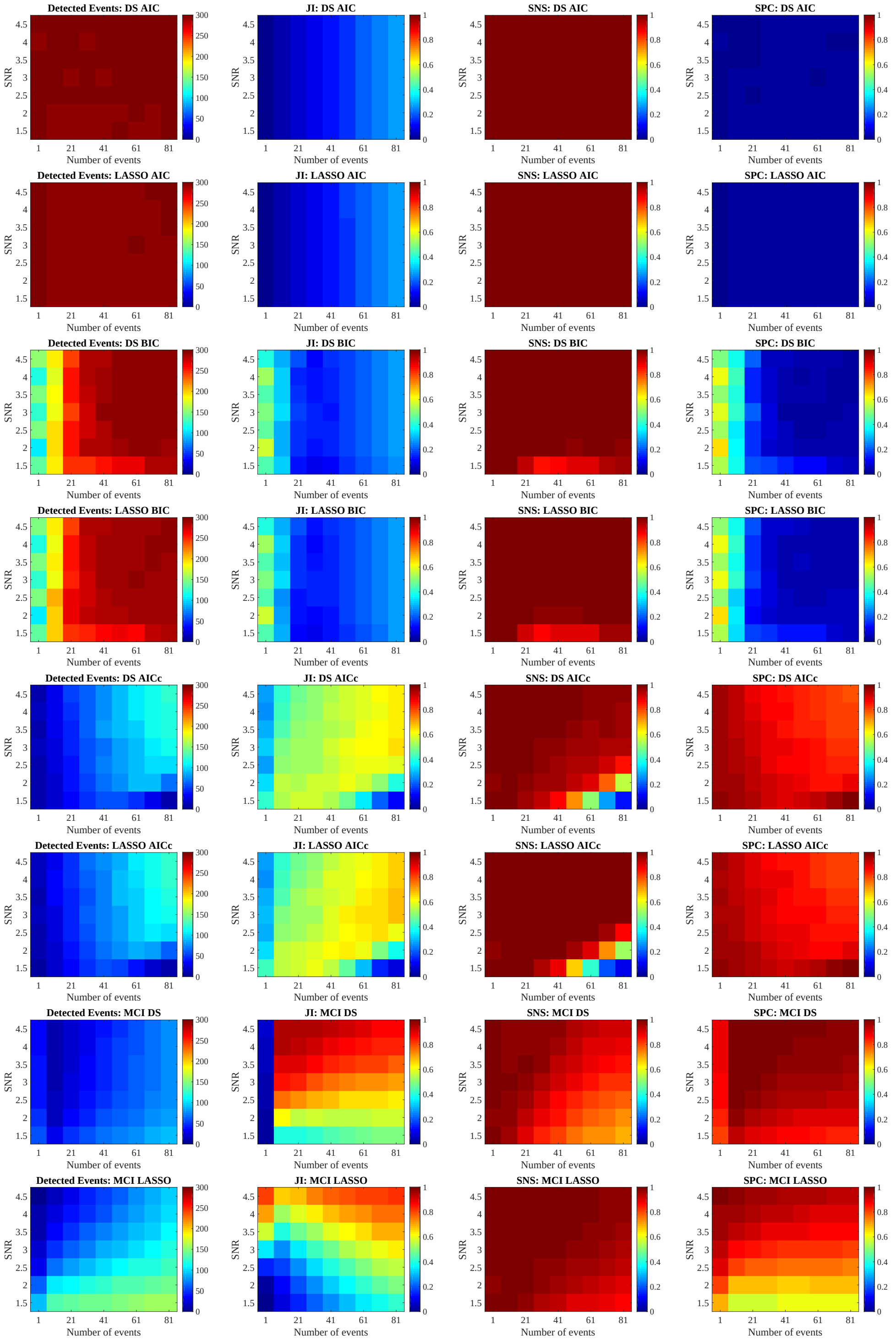


Figure 3.5: Comparison of DS and LASSO with standard selection criteria and MCI results.

3.3.5 N-back experiment

We included the data of 89 measurement sessions in total. For each measurement, we work with the single time series that corresponds to the BOLD-signal of the motor cortex. The detailed description of data acquisition and preprocessing is presented in Appendix 3.6.

According to the task, a subject should press the button during the experiment in the case of adherence to certain conditions. For each subject, we have the button presses timing, which is supposed to be the true activations of the motor cortex. All subjects had the same experimental design that give us a possibility to determine the general true activations signal and to average the measured BOLD signal in order to decrease noise. Also, the analysis of each subject apart was done.

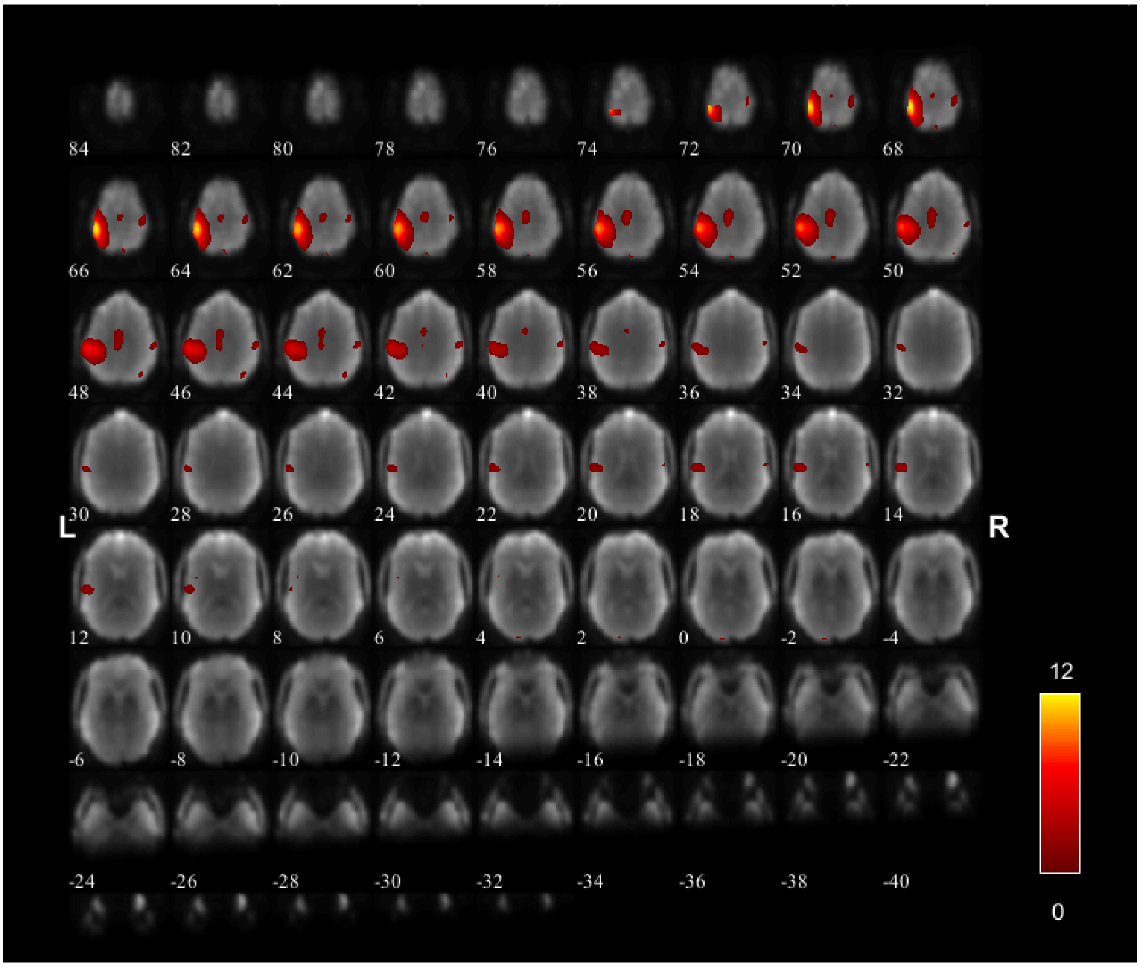


Figure 3.6: Primary motor cortex component.

In this section, we compare the standard approaches DS and LASSO with the selection criteria AIC, AICc, BIC, and our method MCI DS and MCI LASSO. To test that we use a suitable HRF kernel, we have first applied the forward model using the standard HRF function available in the SPM software, as well as its versions shifted by 1 TR backward and forward. The theoretical signal obtained using the forward shifted HRF was

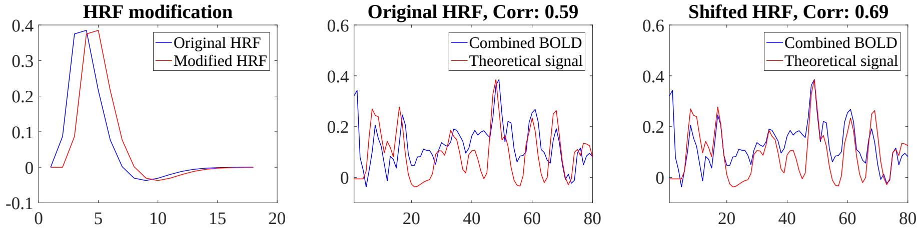


Figure 3.7: Selection of appropriate heamodynamic response function (HRF) using a forward model on real data from the N-back experiment. Left: the default and shifted HRF. Comparison of the averaged measured BOLD signal with the convolution of the true activations vector and the original HRF (middle) and shifted HRF (right).

more correlated with the real signal than the one using the original HRF (see Figure 3.7), suggesting that our data were more in line with this shifted HRF. Note that the shape and delay of HRF varies in principle across subjects, brain regions as well as acquisition protocols, causing extra degrees of freedom in fMRI analysis, however estimating the HRF is outside the scope of the current work, so we just fix it heuristically to the shifted version of the standard template.

In the next step, we applied all mentioned methods to compare the results using the quality measures discussed in Section 3.3.4. Then, Figure 3.8 shows the comparison of the random classifier with the discussed methods for original and shifted HRF. As can be seen, all of the methods for shifted HRF significantly differ from the random classifier, while for original HRF their results are comparable with it.

	JI	SNS	SPC
DS BIC	0.07	0.07	1.00
DS AIC	0.35	0.87	0.41
DS AICc	0.31	0.46	0.81
MCI DS	**0.51**	**0.61**	**0.92**
LASSO BIC	0.00	0.00	1.00
LASSO AIC	0.42	0.94	0.45
LASSO AICc	0.11	0.12	0.96
MCI LASSO	**0.49**	**0.78**	**0.73**

Table 3.1: Comparison of the deconvolution techniques on the group-level real data signal with shifted HRF.

The results of activations signal estimation for the combined signal is shown on Figure 3.9. DS BIC-estimate is almost a trivial vector. It contains only true activations, but only two of them. Conversely, the DS AIC-estimate includes a lot of true detections, which gives the high sensitivity, but also it contains a lot of false detections. Remain two estimates both are better, and MCI DS shows higher values of all three quality characteristics.

Figure 3.10 shows the receiver operating characteristic curve (ROC-curve) for DS and LASSO approaches run on the combined signal. Numerical results for the shifted HRF

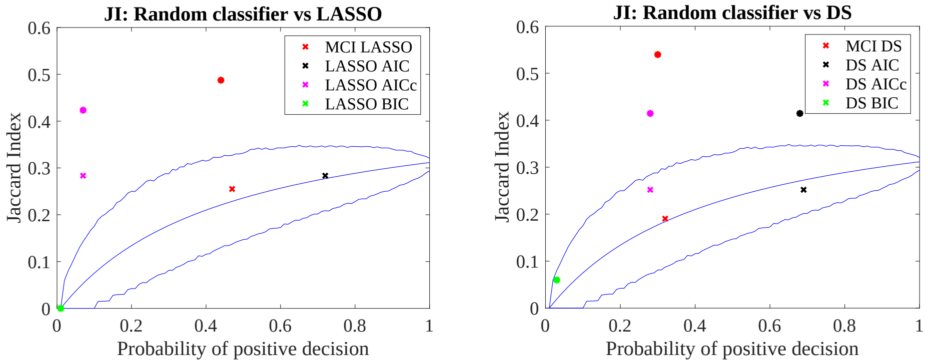


Figure 3.8: Comparison of the performance of the analysed methods with random classifier. For shifted HRF, the discussed methods show the results significantly different from the random one (confidence interval: [0.005,0.995]).

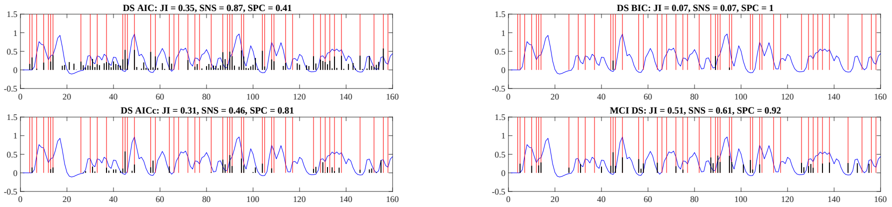


Figure 3.9: Application of the analysed methods to the group-level real data signal. The measured BOLD-signal is shown in blue, the theoretical activations in red, and the estimated activations in black.

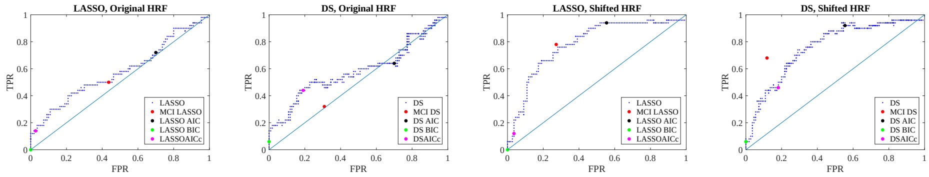


Figure 3.10: ROC curve of the Dantzig Selector and LASSO solutions for the group-level real data signal. Colored points specify the particular selection procedures.

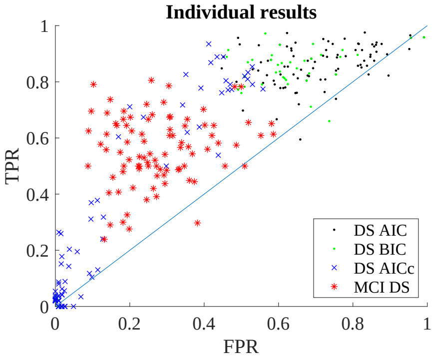


Figure 3.11: Comparison of deconvolution methods on individual-subject level.

are shown in Table 3.1. As it was described above, both of these methods generates the set of possible solutions on which the selection procedure should be applied. Thus, Figure 3.10 shows the ROC-curves with all discussed selection criteria. Also, the points corresponding to the MCI DS and MCI LASSO are marked on the graph, in order to compare the methods. The results demonstrate that the point corresponding to the MCI LASSO method lies close to optimal LASSO solution, and for DS the MCI performs much better than DS with the optimal selection technique.

Figure 3.11 shows the quality of estimates of each subject's time series. It contains the marks of different colors for analysed methods. The black and green points that mark DS AIC and DS BIC methods form almost the same clouds and demonstrate that most of all estimates contain too many positive detections. The blue crosses mark the DS AICc and show a lot of estimates with too few true positive detections. The rest of the blue crosses forms a cloud that has an intersection with all other methods. The red color is used for MCI DS estimates, and this cloud lies apart from the DS AIC and BIC and demonstrates more appropriate estimates in terms of true and false positive rates. Numerical results are shown in Table 3.2.

	JI	SNS	SPC
DS BIC	0.14(0.12)	0.37(0.38)	0.72(0.38)
DS AIC	0.14(0.11)	0.26(0.25)	0.86(0.24)
DS AICc	0.30(0.04)	0.86(0.12)	0.30(0.11)
MCI DS	**0.33(0.05)**	**0.56(0.12)**	**0.72(0.10)**

Table 3.2: Summary performance characteristics (Jaccard index, sensitivity, specificity) for subject-level analysis. Mean (std) is shown for each method.

3.4 Brain connectivity

Let's say that the brain's spatial domain is divided into a set of non-overlapping modules that show some homogeneity relative to information provided by image modalities, such as anatomic delineations, cytoarchitecture, or task-based fMRI activations. Such modules are commonly known as regions of interest (ROI). Under the assumption that regions of interest are linked, the connectivity matrix evaluation became an interesting task. Such a matrix is supposed to be a suitable approximation of interaction structure of the brain dynamics and can be used for finding the behaviourally or clinically relevant markers.

There are a few different approaches to define what does the ”connection” means. According to [Friston, 2011], three types of connectivity can be estimated: structural, functional and effective. The anatomical links are regarded as structural connection; the statistical dependencies define the functional connectivity, and the effective connectivity is the causal interactions between the ROIs. Usually, the functional connectivity is estimated by calculation of the correlation between the BOLD-signals corresponding to ROI, while the other methods that assume any type of model in regions connection, are commonly said to be effective connectivity. This section is devoted to the effective connectivity matrix estimation.

As shown in [Hlinka et al., 2011] and [Hartman et al., 2011], the linear Pearson correlation coefficient is sufficient as a functional connection estimate even when comparing to the nonlinear approaches. However, it is problematic to distinguish the direct and indirect links according to the correlation coefficient. Thus, we are looking for a method that estimates the connectivity matrix including only the direct links.

Since using of the Pearson correlation coefficient is said to be an acceptable approximation of the functional connectivity, we can assume that the linear model on a short time-scale can be sufficient for estimation of the effective connectivity. In this section, we apply the vector autoregressive model for measurement of the connectivity matrix.

Thus, for M ROIs, we measure the BOLD-signal of the length n. Obtained data can be represented as the multivariate time series $\{y_i \in \mathbb{R}^M, i = 1, \ldots, n\}$ where y_i^j is the value of BOLD-signal of the j-th ROI in the i-th moment, and each vector y^j, $j = 1, \ldots, M$ is assumed to fulfill (3.1). Therefore, we consider the existence of multivariate time series $\{s_i \in \mathbb{R}^M, i = 1, \ldots, n\}$, that directly describes the neuronal activity corresponding to measured $\{y_i \in \mathbb{R}^M, i = 1, \ldots, n\}$. Its true values are unknown, however using the technique described in Section 3.3, the estimated signals can be obtained from the BOLD-signals.

We assume that the time series $\{s_i \in \mathbb{R}^M, i = 1, \ldots, n\}$ can be fitted by VAR(1) model. In terms of equation (1.5.4), we handle the matrix B_1 as an estimate of effective connectivity. The matrix obtained in this way does not have the problem of the presence of the high coefficients that would correspond to the indirect links.

The healthy human brains are assumed to have a typical structure, that is relatively stable across people. Therefore, the estimated connectivity matrices can be used for the analysis of the brain network similarity. We demonstrate the classification of the group that consists of healthy subjects and schizophrenics, based on their estimated connectivity matrices.

3.4.1 Classification by effective connectivity

Schizophrenia is a serious mental disorder in which people interpret reality abnormally. Schizophrenia involves a range of problems with cognition, behaviour or emotions. It is characterized by a wide range of symptoms and connected with structural and functional brain changes. In this section, we study the data of 245 healthy subjects and 100 patients with the Schizophrenia diagnosis.

The description of data acquisition and preprocessing is presented in Appendix 3.7. Our analysis is done using a parcellation based on Craddock atlas with 30 ROIs. Thus, we work with the dataset (400 time points)×(30 ROIs) for each subject. The number of ROI is chosen because of the number of parameters to be estimated. The total length of the measurements for each patient is 400 points, thus it is impossible to estimate the connectivity matrix based on much more detailed brain parcellation.

For each subject j, $j = 1,\dots,345$, we estimate the neuronal activity $\{\hat{s}_i^j \in \mathbb{R}^M, i = 1,\dots,n\}$ using the MCI for deconvolution of the measured BOLD time series. Then, to obtain the connectivity matrix B_j, we run the VAR(1) model using the *arfit* package in MATLAB (see [Neumaier and Schneider, 2001]). Healthy controls are divided into two subgroups, $HC_0 = \{1,\dots,145\}$ and $HC = \{146,\dots,245\}$; the patients subgroup is formed as $P = \{246,\dots,345\}$. The first subgroup HC_0 is used for the estimation of the effective connectivity matrix B_{HC} that is meant to be typical for healthy people. This matrix is computed as an average over the mentioned subgroup of subjects,

$$B_{HC} = \frac{1}{|HC_0|} \sum_{i \in HC_0} B_i.$$

Average connectivity for groups HC and P are shown in Figure 3.12. The data from the rest of healthy subjects HC and patients P are used for the comparison of the similarity of the typical matrix B_{HC} and subjects' matrices B_j, $j \in HC \cup P$. By the similarity of two matrices M_1 and M_2, we mean the Pearson correlation between the two vectors $M_1(:)$, $M_2(:)$ formed by reshaping the matrices in the same way for both of them. Such similarity coefficient is calculated for both healthy and patients groups:

$$r_j = r(B_{HC}(:), B_j(:)), \quad j \in HC \cup P,$$

where $r(.,.)$ denotes the Pearson correlation coefficient. The groups $\{r_j, j \in HC\}$ and $\{r_j, j \in P\}$ are compared via the Wilcoxon test which shows the statistically significant difference between median values with $p < 0.05$. Also, the set of classifiers $g(., r_{thr})$ are build based on it according to the rule $g(r_j, r_{thr}) = \mathbb{I}\{r_j < r_{thr}\}$, $j \in HC \cup P$; if $g(r_j, r_{thr}) = 1$ then subject r_j is classified as a patient. Here, the set of thresholds r_{thr}

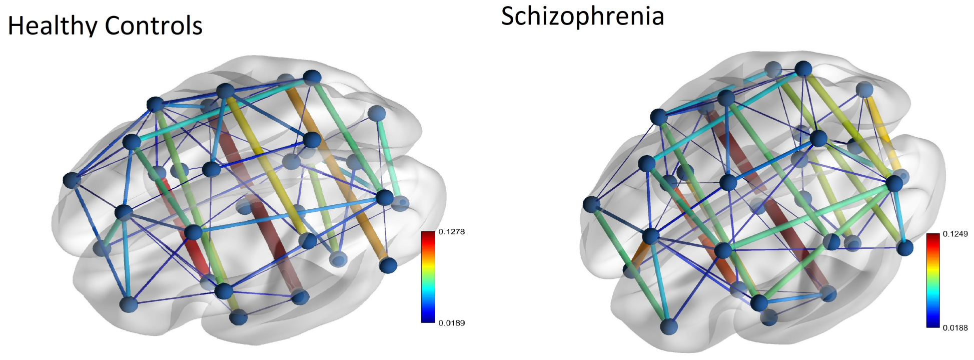


Figure 3.12: Real data example, Schizophrenia detection. Typical connectivity of healthy controls and patients.

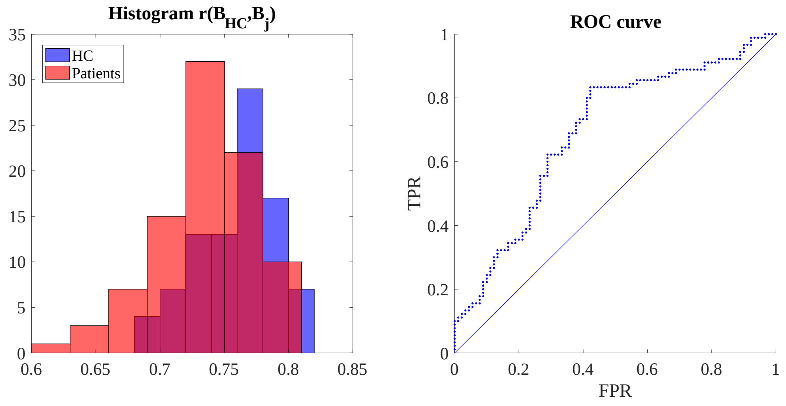


Figure 3.13: Real data example, Schizophrenia detection. Comparison of the similarity of connectivity matrix with the typical one in groups of healthy controls and patients.

is $\{r_j, j \in HC \cup P\}$. The corresponding histograms and obtained receiver operating characteristic (ROC) curve are shown in Figure 3.13. The best accuracy of this ROC curve is $ACC = \frac{TP+TN}{TP+TN+FP+FN} = 0.71$ with corresponding sensitivity $SNS = \frac{TP}{TP+FN} = 0.83$ and specificity $SPC = \frac{TN}{TN+FP} = 0.42$. Thus, the results show that the effective connectivity matrices, obtained by the described procedure, are meaningful and can be sensitive to the changes in brain neuronal activity that corresponds to Schizophrenia disease. Remark that also other machine learning methods can be used for classification, possibly with better results. However, even the proposed naive approach shows useful results.

3.5 BRAD: Software for brain activity detection

In this section, we describe the software BRAD with its user interface and examples of usage. The tool is written in MATLAB 2016b, the sources and few data examples can be downloaded here: $https://github.com/BRADsoftware/BRAD\_v1/$.

3.5.1 Noise filtering

The software includes an option to apply noise reduction before applying the deconvolution procedure. The Wiener filter is used to remove noise from the signal. It is a linear filter used for estimating the desired process from the observed noisy process. Application of the Wiener filter requires knowledge of the spectra of the noise and of the targeted signal. In our work, instead the spectra we use its data-driven estimates. Let us discuss this in more detail.

Formally, the Wiener filter provides an estimate of the signal z from the observed signal y by finding a function $g(n)$ such that $z(n) = (g \star y)(n)$. Let us denote the Fourier transformation of y as $Y(f)$, let $G(f)$ be the Fourier transformation of g, and let $Z(f)$ be the Fourier transformation of z. Let $S_{zz}(f) = |Z(f)|^2$ and $S_{yy}(f) = |Y(f)|^2$ be the power spectral densities (PSD) of z and measured signal y respectively. Also, let $S_{ee}(f)$ be the PSD of the noise e. Then $Z(f) = G(f)Y(f)$ and $G(f)$ can be found as

$$G(f) = \frac{S_{zz}(f)}{S_{zz}(f) + S_{ee}(f)}, \quad f \in [0, 2\pi).$$

In the Wiener filter application approach, $S_{ee}(f)$ and $S_{zz}(f)$ are supposed to be known, but they are principally unknown in our situation. Thus, we use the following procedure to estimate the mentioned spectra: due to the linearity of the Fourier transformation and the zero correlation between z and e, we can represent the PSD of the noisy signal y as a sum of the spectral densities of the noise e and of the clean signal z:

$$S_{yy}(f) = S_{zz}(f) + S_{ee}(f).$$

Here, $S_{yy}(f)$ is supposed to be known: it could be estimated from data or it could be uploaded by the user.

The vector e is supposed to be white noise; thus, its PSD is flat: $S_{ee}(f) = b$, $b \in \mathbb{R}$. Furthermore, we can estimate $S_{zz}(f)$ based on a assumed form of HRF: let h be the kernel of the HRF and $S_{hh}(f)$ be its PSD, then $\hat{S}_{zz}(f) = a * S_{hh}(f)$, where $a \in \mathbb{R}$ is the scaling parameter. In this way we obtain the expression

$$S_{yy}(f) = a * S_{hh}(f) + b, \quad a, b \in \mathbb{R}.$$

Then, the estimate of $G(f)$ can be calculated as

$$\hat{G}(f) = \frac{S_{hh}(f)}{S_{hh}(f) + b/a}, \quad f \in [0, 2\pi).$$

Finally, the filtered signal is estimated as the inverse Fourier transformation of $\hat{G}(f)Y(f)$.

3.5.2 User interface and data visualisation

The presented software BRAD (BRain Activity Detection) consists of functions and Graphical User Interface (GUI) written in MATLAB. The functionality of GUI is organised into several panels within the main frame (see Figure 3.15).

The panel 'Import data' allows uploading the input data. GUI also includes the panel 'Load' which can be used for loading any previously saved templates. The panel 'Save' includes three buttons. They allow to save the input settings ('Save input settings' button), to save all information ('Save calculations' button), and to save the deconvolved signal without complementary information ('Save deconvolved signal' button).

The panel 'Calculations' contains the settings of the requested calculations. Here the user can select which type of signal will be used (filtered, not filtered), and which methods of the deconvolution should be applied (OLS, LASSO, DS). After uploading the data and filling the settings, the user should press the button 'Calculate' to carry out the computations.

The graph located in the center of the frame contains a visualisation of the data. The BOLD-signal and the estimated signal are shown here in blue and red color respectively. The BOLD-signal is shown as a continuous line, the neuronal activity signal is shown as a dashed line. In the chart title there is a short description of the current signals. It contains the following information: whether is the BOLD-signal filtered, which deconvolution method is used, which selection criteria is used and how many peaks the estimated signal has. The panel 'Deconvolution methods' contains three buttons for deconvolution methods described above and a selection criteria for them. For each of the methods, it is possible to show the continuous estimate ('All Peaks') and various approaches to choose the sparse estimate. For all three methods it is possible to choose predefined amount of peaks ('Known Peaks Amount'). Also, for the LASSO and the Dantzig Selector the standard selection criteria AIC and BIC are implemented, together with the selection criteria based on the theory of mixtures of components (MCI).

The panel 'Signal type' allows selecting the use of the filtered or the non-filtered signal. The 'Whole Brain Level' button provides the possibility to analyse multivariate (voxel-level) time series.

The special panel 'Movie' allows to analyse the signal in terms of videoframes extracted from a stimulation material. It allows to show different sets of frames: the current one (corresponding to the cursor) using the button 'Show the frame'; the current and the previous one; the frame corresponding to a particular time index; all frames corresponding to positive peaks in deconvolved signal; all frames from a time interval. The images are shown in an additional window. The scheme of BRAD is shown in Figure 3.14.

3.5.3 User inputs

To analyse the data, the user needs to upload one or more types of data inputs. This can be done using the panel 'Import data'.

- Measured signal (necessary). It is a '.dat' file which contains one time series. The time series is supposed to be a measured BOLD-signal represented by a numerical column vector. It should be done using the button 'Import data'. The uploaded signal is rescaled to the range [0,1].

- Shape of the HRF (optional). It is also a '.dat' file which contains a time series defining the shape of the hemodynamic response function represented by a numerical column vector. Its length should be shorter than the length of the output signal. Moreover, the corresponding Toeplitz matrix should not be singular. Note that if the matrix is badly conditioned, some numerical problems may appear.
  The button 'Import HRF' should be used to upload the data. There is a possibility

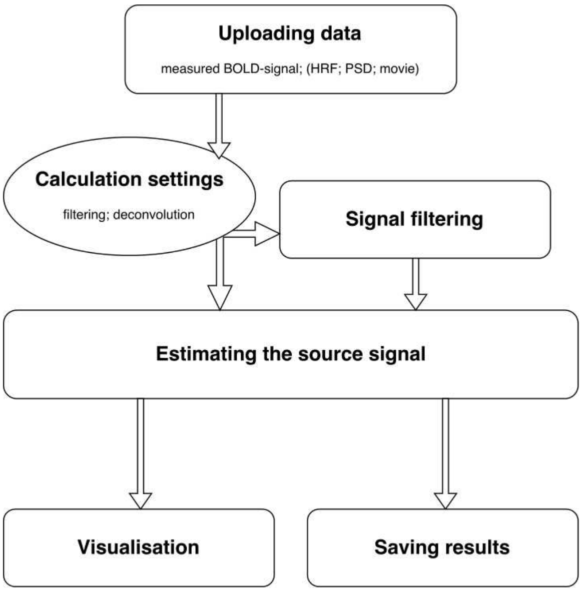


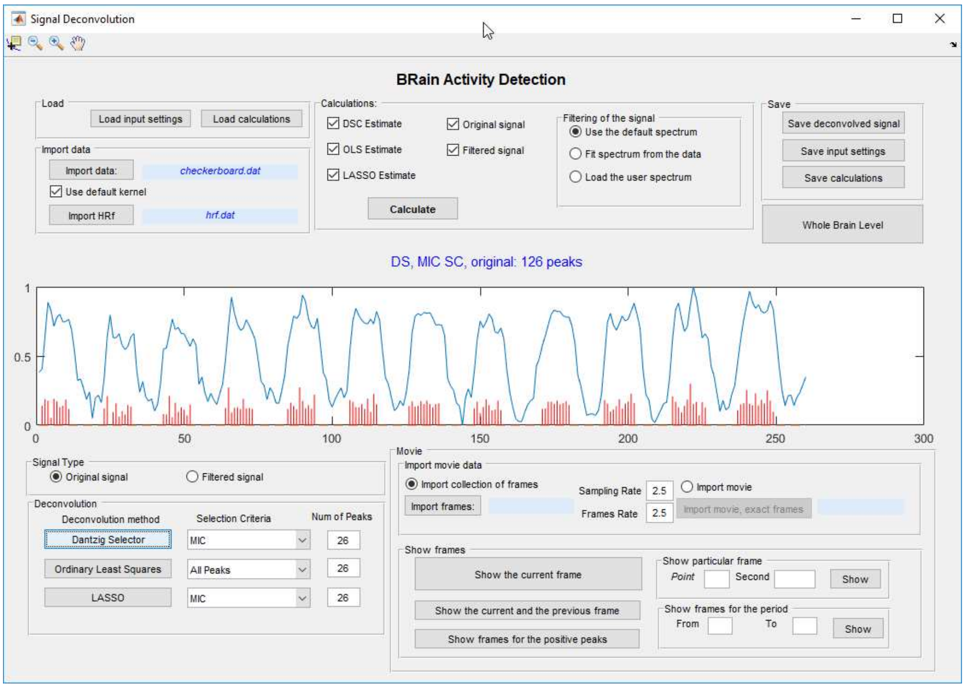
Figure 3.14: Block-scheme of the provided software BRAD. Firstly, the user should upload the data (measured BOLD-signal is necessary; hemodynamic response function, power spectral density and video are optional). Then the user should specify calculation settings (if a filtering of the BOLD-signal is needed, which filtering method should be used, which deconvolution methods should be used). After that, the estimates of the source signal are calculated. Obtained estimates can be visualized or/and saved.

to use the default kernel, which is saved in the file 'hrf.dat'. We include an useful example together with the software.

- Spectrum (optional). In case of the selection of filtering the signal, an estimate of its power spectral density is needed. The user should specify whether it is preferable to use the default power spectral density, to estimate the spectrum from the BOLD-signal, or if the power spectral density should be uploaded. In case of uploading, the file should be in the '.dat' format and should contain the numerical vector-column with length equal to the BOLD-signal length.

- Video (optional). In many cases, the BOLD signal is measured during visual stimulation of the subject. For the interpretation of the estimated signal s, it is useful to inspect the corresponding stimulus content. In BRAD we implemented two options for this: to upload the collection of frames or to upload the full video. It can be done using the buttons 'Import frames' or 'Import movie and exact frames' respectively. The sampling rate and frame rate is requested to be entered for analysis using visual information.

Further, the panel 'Load' could be used for uploading the data. For the whole brain level deconvolution, a 4D-matrix (3D + time) of real numbers in '.mat' file is requested. Moreover, a gray-matter mask is needed (3D-matrix in '.mat' file).



Import data: checkerboard.dat
☑ Use default kernel
Import HRf: hrf.dat

☑ DSC Estimate
☑ OLS Estimate
☑ LASSO Estimate
☑ Original signal
☑ Filtered signal

● Use the default spectrum
○ Fit spectrum from the data
○ Load the user spectrum

● Original signal
○ Filtered signal

Deconvolution method	Selection Criteria	Num of Peaks
Dantzig Selector	MIC	26
Ordinary Least Squares	All Peaks	26
LASSO	MIC	26

● Import collection of frames
○ Import movie
Sampling Rate 2.5
Frames Rate 2.5

Figure 3.15: Flashing checkerboard experiment. The figure shows the result of a BRAD run: the estimate of the neuronal activity during cycles of flashing checkerboard and black screen stimulation. BOLD-signal (blue line) shows 12 cycles, estimated neuronal signal (the red one) corresponds to the reaction to the flashing checkerboard condition. Average lengths of detected blocks of peaks is in line with the real length of the flashing checkerboard blocks.

3.5.4 Flashing checkerboard experiment

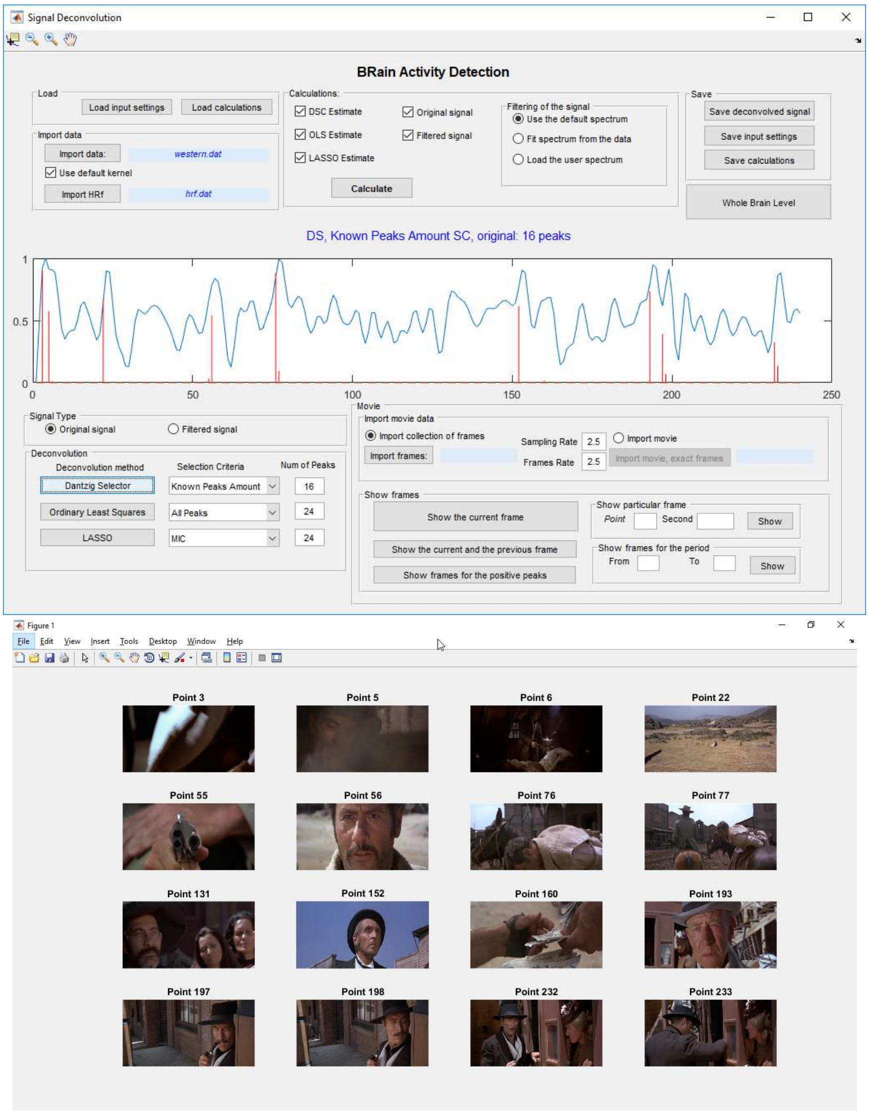
We use measurements from one healthy subject, female, age 39 years. The paradigm of the experiment consisted of alternating cycles of blocks of a flashing checkerboard and blocks with a black screen, (12 blocks of each). Duration of the rest blocks was 25 seconds; flashing blocks had duration 26s, 28s, and 32s (4 blocks of each). We study the BOLD time series that corresponds to the primary visual cortex. The detailed description of data acquisition and preprocessing is presented in Appendix 3.8.

We used the MCI for the estimation of the neuronal signal, with no filtering applied. The estimated signal clearly shows 12 blocks of peaks (which correspond to 12 waves in the BOLD-signal) and 12 calm periods. An average block of the detected peaks contains $10-13$ activations, corresponding to $25-32.5$s. The demonstration of measured and estimated signals, obtained using BRAD, is presented in the Figure 3.15.

3.5.5 Movie experiment

In this experiment 84 healthy subjects took part, mean age/±std: $30.83/\pm 8.48$ years. We study the BOLD signal that corresponds to the occipital fusiform gyrus. The detailed description of data acquisition and preprocessing is presented in Appendix 3.9. We measured the brain activity during watching a movie (part of classical western ”The Good, the Bad and the Ugly": uninterrupted 10-minute segment 16:48-26:48). The experiment setup including the movie selection was motivated by the work of Hasson et al. [Hasson et al., 2004], who proposed the use of a 'reverse-correlation' method to infer the functional relevance of a given region based on the content of a rich audiovisual stimulation material common in the moments corresponding to highest activation.

The resulting average time series of the occipital fusiform gyrus, selected from the Harvard-Oxford cortical atlas, was then used in the BRAD toolbox. For the underlying signal estimation, the Dantzig Selector was used; extraction of 16 peaks was selected. The result is shown in Figure 3.16. The fusiform face area is known to increase activity during presentation of visual stimuli containing faces (see [Kanwisher et al., 1997], [Haxby et al., 2000]). Therefore the detection of frames containing faces in many cases seems to be in line with the expectations, similarly as in the previous work using the reverse-correlation approach [Hasson et al., 2004]. However, as faces of various form appear relatively densely in the movie, proper statistical inference, that would probably include some permutation scheme (which should also take into account the autocorrelation of both the movie and brain activity time series) would be needed for more robust neuropsychological interpretation.



Import data: western.dat

☑ Use default kernel

Import HRf: hrf.dat

☑ DSC Estimate ☑ Original signal

☑ OLS Estimate ☑ Filtered signal

☑ LASSO Estimate

● Use the default spectrum

○ Fit spectrum from the data

○ Load the user spectrum

● Original signal ○ Filtered signal

Deconvolution method	Selection Criteria	Num of Peaks
Dantzig Selector	Known Peaks Amount	16
Ordinary Least Squares	All Peaks	24
LASSO	MIC	24

● Import collection of frames

Import frames:

Sampling Rate 2.5

Frames Rate 2.5

○ Import movie

Import movie, exact frames

Point Second

From To

Figure 3.16: Movie experiment. The left picture shows an estimate of the neuronal activity during watching the segment of western ”The Good, the Bad and the Ugly” when selection of 16 peaks was requested by the user. The measured BOLD-signal representing the occipital fusiform gyrus is shown by the blue line, the estimated signal is shown in red color. The right picture demonstrates frames from the movie which correspond to the detected peaks of neuronal activity.

Discussion

In studying of the socio-economical process of the submissions to municipalities in the Czech Republic, the main goals of the analysis are the characterisation of the data and building the model. Our analysis shows that the process of submissions to the municipalities can be described by the spatiotemporal point process with the discrete space and time domains, and the point of the spatiotemporal lattice of possible process value can be observed more than once with the non-zero probability. What actually happens, is the specific lattice points appear, corresponding for example to Prague or Brno municipalities, by tens and hundreds in the datasets.

We realised that the process of correspondence behaviour is very difficult in the sense that it can be described neither by classical models of point processes like Poisson process or cluster processes and their easy modifications, nor using Poisson distribution for modelling repetitions of points in the same time and spatial coordinates. Therefore we had to focus on empirical distributions which according to our opinion seems to be a good way. It could be potentially interesting, to compare the empirical distributions between the data from different municipalities. However, the available datasets are not large enough for such an analysis.

The change point analysis is proposed for the detection of the process changes. The analysis shows, that the proposed modified classical technique of change point detection, gives a reasonable result for such type of data. However, the assumption about the partial linearity of the data is not always the best one. Additionally, using the non-linear trends can be very useful. Moreover, there are approaches for the change point detection in real time. We did not try such methods, but we suppose that such a technique, can be also useful for effective administration.

Regarding the submissions subgroups analysis, we show that the autoregressive models give reasonable results in modelling of the number of submissions of different types. We provide such analysis with the lag 1 month, but based on the origin of the data, we suppose that bigger lags, like 3 months or 12 months, can also be very useful. However, such analysis would need much longer observations, than we have. The second potentially weak spot of this model is big differences in the number of submissions of the particular type. Because of that, the coefficients of the VAR model can be poorly interpreted; to avoid this problem, the types can be normalised before the analysis.

Finally, the rejection of the hypothesis about the independence between spatial and temporal coordinates suggests that the analysis of the dependencies between the municipalities or the cumulative geographical regions can be potentially interesting. To make this analysis better, it would be preferable to estimate the VAR model, or at least apply the Granger causality, but this also needs much more data.

The second part of the work is dedicated to neuroimaging data analysis. This thesis is devoted to the application of the mathematical instruments to the data, thus we do not discuss the correctness, assumptions, and problems of the used models, specifically, the modelling of the BOLD signal and the hemodynamic response itself. For interested readers, we refer to [Ogawa et al., 1992].

In this thesis, we introduce a method for improvement of the estimates of neuronal activity from the BOLD signal. It works by applying additional postprocessing on the result of standard regularization-based deconvolution methods, such as LASSO and Dantzig Selector. Notably, it outperforms both the standard application of the Bayes information criterion and Akaike's information criterion, as well as its improved variant. We demonstrate the advantages of this method both in numerical simulations as well as in experimental data, where the activation of the motor cortex occurring during a complex cognitive task is studied. More accurate estimates, obtained with the proposed method, suggest that the approach may further improve the detection of neuronal activation that can not be directly recorded or controlled in other datasets.

The method is dealing with the challenges of the common overfitting or underfitting of the data that occurs to regularization methods combined with standard model selection criteria. It is combining the information contained in the solutions rather than select single one, which shows a good performance both in simulations and in a real dataset example. The future work would involve attempting its efficient combination with some other directions recently taken up in the field, namely spatial regularization ([Karahanoğlu et al., 2013], [Farouj et al., 2017] or [Hernandez-Garcia and Ulfarsson, 2011]), change sparsity, HRF estimation ([Aggarwal et al., 2015]), and further processing.

Sparse linear regression methods are not the only way for deconvolution in this field. Here are few other families of approaches described in the literature: the cubature Kalman filtering [Havlicek et al., 2010], activelets (wavelets designed using prior knowledge on the hemodynamic response function form [Khalidov et al., 2011]), spatiotemporal HRF deconvolution [Aquino et al., 2014] or nonparametric hemodynamic deconvolution of fMRI using homomorphic filtering [Sreenivasan et al., 2015]. However, as the outlined MCI approach relies on being combined with an initial method providing a parameterized family of solutions (in this thesis, we used the LASSO and Danzig Selector as examples of such families), we leave open the question of how it relates and could synergize with them. We can speculate that the key idea of combining model outputs using e.g. Bayesian approaches instead of crisp model selection within a constrained parametric family may be applicable to other schemes, as long as these involve selection among model parameterized model families.

The timing of local neuronal activations can be important on its own, for example, in studying of cortical response to unpredictable mental events [Gaudes et al., 2011], or can be used for obtaining meta-analytic interpretation of the brain activation patterns [Tan et al., 2017]. In [Allan et al., 2015] has been proposed that functional connectivity in MRI is driven by spontaneous BOLD events detectable by deconvolution methods. Also, the use of deconvolution as an intermediate step to estimate effective connectivity has been discussed in [Wu et al., 2013]. Similarly, the dynamics of deconvolution-based brain activation patterns can be further explicitly modelled to fit sparse coupled hidden Markov models [Bolton et al., 2018] characterizing the large-scale brain dynamics. Last but not least, the development of deconvolution approaches is relevant also in the context of functional near-infrared spectroscopy [Santosa et al., 2019, Seghouane and Ferrari, 2019] and quantitative susceptibility mapping [Costagli et al., 2019].

With respect to the evaluation of the effective connectivity in our work, there are alternative approaches discussed in the literature. The effective connectivity estimate using DCM is proposed and discussed in [Friston et al., 2011] and [Lohmann et al., 2012]. In [Frässle et al., 2017], the newest regressive DCM is presented. Another alternative model, based on Granger causality, is discussed, for example, in [Marinazzo et al., 2011] and in [Friston et al., 2014]. A comparison of algorithms, including a newly developed method for causal inference in large networks, is presented in [Kořenek and Hlinka, 2020].

Regarding the last part of the thesis describing the software for deconvolution, there are currently several limitations of the presented tool BRAD. For instance, the list of the included deconvolution methods is not meant to be complete. They were selected as typical examples of the existing methods. We present the standard method (OLS), the sparse approaches discussed earlier (LASSO and Dantzig Selector) with AIC and BIC. Also, MCI is included in the software as a newly developed approach to estimate construction. Another methods also may be useful, and the modular architecture of the software tool allows their easy and efficient inclusion in the tool at a later stage of development in future or by other labs.

The current version of BRAD only allows presentation of the visual content of the stimuli. However, as an example extension, the module for stimulation using different modalities, audio stream, or visualisation of more general stimuli can be included. There are several noise filtering choices implemented in the presented software. However, the list is not complete. Other options, for example, such as band-pass filtering with a predefined pass-band, could be added.

Conclusion

This thesis is dedicated to the application of statistical analysis and mathematical modelling to real-data tasks. We present two projects from very different fields, but both of them can be described as the spatiotemporal point processes. The analysis is done using a wide range of methods that differ from task to task and depend on the question to be answered. Some of the approaches, specifically, the vector autoregressive model, are applied for both processes, but with different goals and results.

The first studied phenomenon is the process of submissions to municipalities in the Czech Republic that can be described by a spatiotemporal point process, where the geographical coordinates of sender and day of submission are the spatial and temporal coordinate of the point, respectively. For the analysis, we use the classical as well as modified data analysis methods. Every time, we formulate the reason for the particular method's usage and show the problematic moments of its application.

Our main practical results of the analysis are: we show that the VAR(1) model can be a good approximation of the number of submissions per month; the approach allows modelling the total number of submissions or to focus on the selected subgroup. Also, we propose the change point detection technique that is applicable for this data and easy for the implementation.

The second studied process is brain neuronal activity, measured by blood oxygenation level-dependent (BOLD) signal measured by functional MRI. For this neuroimaging data, we present a new deconvolution method, the Mixture Component Inference (MCI) approach that allows extracting the timing of neuronal activations from the measured fMRI signal. This result is promising as a general exploratory technique and can be useful in various tasks, for example, for filtering the fMRI data on voxel-level or for estimation of neuronal activity in experiments with complicated design. Further, we describe the technique that allows estimating the effective brain connectivity from the fMRI signal based on the neuronal activations obtained using the proposed MCI approach. Finally, we present a software tool BRAD for the estimation and analysis of human brain neuronal activity.

Appendix

3.6 N-back experiment

Subjects

The real data example includes the data of 56 subjects (24 males and 32 females, age = 33.66 ± 10.97) that performed fMRI recording with n-back task as a part of a complex study with repeated psychiatric, neuroanatomical and neuropsychological evaluation. All subjects have been recruited for the study based on the following exclusion criteria: history of mental disorder, history of neurological disorder, presence of any artificial objects that would interfere with the magnetic resonance imaging. All volunteers signed Informal consent approved by the Ethic committee of National Institute of Mental Health (Czech republic). The data set of 89 MRI sessions in total was analyzed, including data from repeated sessions (6 months apart) in 25 subjects (2 visits for 17 subjects and 3 visits for 8 subjects).

N-back task

The n-back task is a continuous performance task commonly used for assessment of working memory performance [Gazzaniga, Michael S.; Ivry, Richard B.; Mangun, George R. (2009). Cognitive Neuroscience: The Biology of the Mind (2nd ed.).]. The subject is presented with a sequence of stimuli, and the task requires indicating when the current stimulus matches the one from n steps earlier in the sequence. If the task load factor n equals 2 or more, the working memory buffer needs to be updated continuously to keep track of what the current stimulus must be compared to. The spatial n-back task paradigm applied in this study presents the stimulus (blue square presented on black background) that appears in one of the 9 possible positions (3x3 matrix) on the screen during each turn. The tasks consisted of two alternating conditions: 0-back and 2-back. Each condition was presented in four sessions (eight sessions in total, sequence order: 0-2-0-2-0-2-0-2). The paradigm required the subject to recall the spatial position of the blue square a) in initial position for the 0-back session and b) two turns back for 2-back session, and press the button during the experiment based on these rules specific for each of the n-back conditions. Importantly, for the purpose of this study, we did not distinguish between the two conditions (0- and 2-back) or the correctness of the subjects' responses, we simply considered/analysed all events of button presses recorded for each subject. For each subject, we have the button presses timing, which is assumed to correspond to the true activations of the motor cortex. As all subjects had the same experimental design, we can also carry out a group level analysis, setting as group ground-truth as button presses present in at least 80% of subjects (71 of 89)] and to average the measured BOLD signal in order to decrease noise. Also, the analysis of each subject separately was done. On average the subjects responded by button press 44.63 times (SD = 6.18, min = 27, max = 58). Most of the responses were correct, on average 5 of the responses were incorrect.

fMRI acquisition and preprocessing

Brain images were obtained using Siemens Prisma 3T MR machine. Functional T2*-weighted images with BOLD contrast were acquired with voxel size 3x3x3 mm, slice dimensions 64x64 voxels, 32 axial slices, repetition time 2000 ms, echo time 30 ms, flip angle 70°.

The data pre-processing was done using SPM8 toolbox (The Wellcome Dept. of Cognitive Neurology, University College London) in Matlab (The MathWorks, Inc.). In order to minimise head motion effects, functional volumes were spatially realigned, and the slice-time correction was used to fix acquisition delays. This was followed by normalization of functional volumes into the standard anatomical space using a template provided by SPM toolbox and spatial smoothing with 8 mm FWHM kernel.

An independent component analysis was performed in order to extract time series corresponding to the primary motor cortex, which was done using GIFT toolbox (MIALAB, Mind Research Network) in Matlab. The number of independent components was estimated using the minimum description length criteria. Independent components were decomposed with an Infomax algorithm (see [Bell and Sejnowski, 1995]) with default settings. A component representing the primary motor cortex was selected by visual inspection, see Figure 3.6. Its time series was further analysed. Moreover, on each of the mentioned time series the outlier correction was carried out using Tukey's test (see [Tukey, 1977]) and the re-scaling discussed in Section 2.1 was applied. To check that good correspondence of the stimuli and the estimated neuronal activations was not caused by any motion artefact related to the motor activity during the button press, we did the deconvolution for all the ICA components with all the methods and compared the results with stimuli. A significant match of the estimate to the real stimuli was found only for two components that overlapped with motor cortex; the unilateral component, that showed higher agreement with the stimulus, was selected for subsequent analysis.

3.7 Schizophrenia

We study the data of 245 healthy subjects and 100 patients with the Schizophrenia diagnoses. Scanning was performed with a 3T MR scanner (Siemens; Magnetom Trio) located at the Institute of Clinical and Experimental Medicine in Prague, Czech Republic. Functional T2*-weighted images with BOLD contrast were acquired with voxel size $3\times3\times3$mm. GE-EPIs (TR/TE = 2000/30 ms) comprised axial slices acquired continuously in descending order covering the entire cerebrum. A three-dimensional high-resolution T1-weighted image (TR/TE/TI = 2300/4.6/900 ms, voxel = $1\times1\times1\text{mm}^3$) covering the entire brain was used for anatomical reference.

Initial data preprocessing was done using the SPM12 toolbox (Wellcome Department of Cognitive Neurology, London, UK), CONN toolbox (McGovern Institute for Brain Research, MIT, USA) and FSL routines (FMRIB Software Library v5.0, Analysis Group, FMRIB, Oxford, UK). CONN's default preprocessing pipeline (defaultMNI) contains the functional realignment and unwarping, slice-timing correction, structural segmentation into white matter and cerebrospinal fluid and structural normalization to the MNI space, functional normalization to the MNI space, outlier detection, and smoothing with 8mm kernel size.

3.8 Flashing checkerboard

We use measurements from one healthy subject, female, age 39 years. The brain images were obtained using a Siemens Prisma 3T MR machine. Functional T2*-weighted images with BOLD contrast were acquired with a voxel size 3x3x3 mm, slice dimensions 64x64 voxels, 44 axial slices, repetition time 2500 ms, echo time 30 ms and flip angle 80°.

All preprocessing steps were carried out using the SPM8 toolbox (The Wellcome Dept. of Cognitive Neurology, University College London) in the Matlab (The MathWorks, Inc.) environment. Functional volumes were spatially realigned to minimise head motion effects and slice-time corrected to fix any acquisition delays. This was followed by normalization of functional volumes into the standard anatomical space using a template provided by the SPM toolbox and a spatial smoothing with 8 mm FWHM kernel.

In order to extract the time series corresponding to primary visual cortex, an independent component analysis was performed using the GIFT toolbox (MIALAB, Mind Research Network) in Matlab. The number of independent components was estimated using the minimum description length criteria. Independent components were decomposed with the Infomax algorithm (see [Bell and Sejnowski, 1995]) with the default settings. We selected a component representing the primary visual cortex by visual inspection. Its time series was futher analysed.

3.9 Movie experiment

In this experiment 84 healthy subjects took part, mean age/±std: $30.83/\pm 8.48$ years. The brain images were collected using a Siemens Trio 3T MR machine. The high-resolution 3D anatomical T1-weighted image which consisted of 224 sagittal slices was acquired with TR = 2300 ms, TE = 4.63 ms, flip angle = 10°, FOV 256x256, image matrix size 256x256, voxel size = 1x1x1 mm. The functional T2*-weighted images with blood oxygenation level-dependent (BOLD) contrast were acquired with voxel size 3x3x3 mm, slice dimensions 64x64 voxels, 44 axial slices, repetition time 2500 ms, echo time 30 ms and flip angle 90°.

The preprocessing pipeline consisted of the same blocks as in the flashing checkerboard experiment but in a slightly different order. The preprocessing was performed using the CONN toolbox (The Gabrieli Lab. McGovern Institute for Brain Research, MIT) which uses standard SPM modules. Functional volumes were slice-time corrected, spatially realigned, normalised into MNI space and spatially smoothed with 8 mm FWHM kernel. The subject-specific anatomical T1 image was used for the purpose of the indirect normalization of the functional volumes into the MNI space and to create a mask of the white matter (WM) and the cerebrospinal fluid (CSF). These masks were used as regions from which average signals were extracted.

The CONN toolbox performed denoising of the original fMRI signal using the average signal of WM, CSF as well as six motion parameters to minimise their possible confounding effect. The signal was band-pass filtered using a Butterworth filter (0.004-0.1 Hz). The resulting average time series of the occipital fusiform gyrus, selected from the Harvard-Oxford cortical atlas, was then used in the BRAD toolbox.

3.10 BRAD

Supplementary materials (tool BRAD, documentation and data examples) can be found online at *https : //github.com/BRADsoftware/BRAD_v2/*.

Bibliography

[Aggarwal et al., 2015] Aggarwal, P., Gupta, A., and Garg, A. (2015). Joint estimation of activity signal and HRF in fMRI using fused LASSO. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 829–833.

[Allan et al., 2015] Allan, T. W., Francis, S. T., Caballero-Gaudes, C., Morris, P. G., Liddle, E. B., Liddle, P. F., Brookes, M. J., and Gowland, P. A. (2015). Functional Connectivity in MRI Is Driven by Spontaneous BOLD Events. *PLOS ONE*, 10(4).

[Anderson and Ray, 1975] Anderson, C. W. and Ray, W. D. (1975). Improved maximum likelihood estimators for the gamma distribution. *Communications in Statistics*, 4(5):437–448.

[Antoch et al., 2000] Antoch, J., Hušková, M., and Jarušková, D. (2000). Change point detection. *5th ERS IASC Summer School.*

[Aquino et al., 2014] Aquino, K., Robinson, P., Schira, M., and Breakspear, M. (2014). Deconvolution of neural dynamics from fMRI data using a spatiotemporal hemodynamic response function. *NeuroImage*, 94:203 – 215.

[Armstrong, 2014] Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5):502–508.

[Asif and Romberg, 2009] Asif, M. S. and Romberg, J. (2009). Dantzig selector homotopy with dynamic measurements. In *Computational Imaging VII*, volume 7246 of *Proceedings of SPIE*. IS&T - Soc Imaging Sci & Technol; SPIE.

[Autin and Pouet, 2012] Autin, F. and Pouet, C. (2012). Adaptive test on components of densities mixture. *Mathematical Methods of Statistics*, 21(2):93–108.

[Bell and Sejnowski, 1995] Bell, A. J. and Sejnowski, T. J. (1995). An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.

[Bickel et al., 2009] Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732.

[Bolton et al., 2018] Bolton, T. A. W., Tarun, A., Sterpenich, V., Schwartz, S., and Van De Ville, D. (2018). Interaction Between Large-Scale Functional Brain Networks are Captured by Sparse Coupled HMMs. *IEEE Transactions On Medical Imaging*, 37(1):230–240.

[Bonferroni, 1936] Bonferroni, C. E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*.

[Bowman and Shenton, 1988] Bowman, K. O. and Shenton, L. R. (1988). *Properties of Estimators for the Gamma Distribution*. Taylor Francis Inc, United States.

[Burnham and Anderson, 2004] Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33:261—-304.

[Bush and Cisler, 2013] Bush, K. and Cisler, J. (2013). Decoding neural events from fMRI BOLD signal: A comparison of existing approaches and development of a new algorithm. *Magnetic Resonance Imaging*, 31(6):976 – 989.

[Candes and Tao, 2007] Candes, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Annals of Statistics*, 35(6):2313–2351.

[Chiu et al., 2013] Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). *Stochastic Geometry and Its Applications, 3rd Edition.* Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ.

[Choi and Wette, 1969] Choi, S. C. and Wette, R. (1969). Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics*, 11(4):683–690.

[Costagli et al., 2019] Costagli, M., Lancione, M., Cecchetti, L., Pietrini, P., Cosottini, M., Ricciardi, E., and Tosetti, M. (2019). Quantitative Susceptibility Mapping of Brain Function During Auditory Stimulation. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(4):516–522.

[Daley and Vere-Jones, 2003] Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*. Springer-Verlag New York.

[Doronin and Maiboroda, 2015] Doronin, A. and Maiboroda, R. (2015). Testing hypotheses on moments by observations from a mixture with varying concentrations. *Modern Stochastics: Theory and Applications*, 1(2):195–209.

[Efron et al., 2004] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least Angle Regression. *Annals of Statistics*, 32(2):407–451.

[Eichinger and Kirch, 2018] Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526–564.

[Etzel et al., 2009] Etzel, J. A., Gazzola, V., and Keysers, C. (2009). An introduction to anatomical roi-based fmri classification analysis. *Brain Research*, 1282:114 – 125.

[Farouj et al., 2017] Farouj, Y., Karahanoglu, F. I., and Van de Ville, D. (2017). Regularized spatiotemporal deconvolution of fMRI data using gray-matter constrained total variation. In *2017 IEEE 14th International Symposium On Biomedical Imaging (Isbi 2017)*, pages 472–475. IEEE; EMB; IEEE Signal Proc Soc.

[Friston, 2009] Friston, K. (2009). Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLOS Biology*, 7(2):1–6.

[Friston, 2011] Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36. PMID: 22432952.

[Friston et al., 2014] Friston, K. J., Bastos, A. M., Oswal, A., van Wijk, B., Richter, C., and Litvak, V. (2014). Granger causality revisited. *NeuroImage*, 101:796 – 808.

[Friston et al., 2011] Friston, K. J., Li, B., Daunizeau, J., and Stephan, K. E. (2011). Network discovery with dcm. *NeuroImage*, 56(3):1202–1221.

[Frässle et al., 2017] Frässle, S., Lomakina, E. I., Razi, A., Friston, K. J., Buhmann, J. M., and Stephan, K. E. (2017). Regression dcm for fmri. *NeuroImage*, 155:406 – 421.

[Gaudes et al., 2011] Gaudes, C. C., Petridou, N., Dryden, I. L., Bai, L., Francis, S. T., and Gowland, P. A. (2011). Detection and characterization of single-trial fMRI BOLD responses: paradigm free mapping. *Human Brain Mapping*, 32(9):1400–1418.

[Gaudes et al., 2013] Gaudes, C. C., Petridou, N., Francis, S. T., Dryden, I. L., and Gowland, P. A. (2013). Paradigm free mapping with sparse regression automatically detects single-trial functional magnetic resonance imaging blood oxygenation level dependent responses. *Human Brain Mapping*, 34(3):501–518.

[Giroux and McLelland, 2003] Giroux, G. and McLelland, A. J. (2003). Governance structures and accounting at large municipalities. *Journal of Accounting and Public Policy*, 22(3):203 – 230.

[Gower, 2004] Gower, J. C. (2004). *Measures of Similarity, Dissimilarity and Distance*. John Wiley & Sons, Inc.

[Gray, 2006] Gray, R. M. (2006). Toeplitz and circulant matrices: a review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239.

[Green, 2011] Green, W. H. (2011). *Econometric Analysis, 7th Edition*. Pearson Education, New York.

[Hand and Yu, 2001] Hand, D. J. and Yu, K. (2001). Idiot's Bayes - Not So Stupid After All? *International Statistical Review*, 69(3):385–398.

[Hardle et al., 2006] Hardle, W. K., Muller, M., Sperlich, S., and Werwatz, A. (2006). *Nonparametric and Semiparametric Models*.

[Hartman et al., 2011] Hartman, D., Hlinka, J., Paluš, M., Mantini, D., and Corbetta, M. (2011). The role of nonlinearity in computing graph-theoretical properties of resting-state functional magnetic resonance imaging brain networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):013119.

[Hasson et al., 2004] Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664):1634–1640.

[Havlicek et al., 2010] Havlicek, M., Jan, J., Brazdil, M., and Calhoun, V. D. (2010). Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *NeuroImage*, 53(1):65 – 77.

[Hawkins and Deng, 2010] Hawkins, D. M. and Deng, Q. (2010). A nonparametric change-point control chart. *Journal of Quality Technology*, 42(2).

[Haxby et al., 2000] Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233.

[Heeger and Ress, 2002] Heeger, D. and Ress, D. (2002). What does fmri tell us about neuronal activity? *Nature reviews. Neuroscience*, 3:142–51.

[Hernandez-Garcia and Ulfarsson, 2011] Hernandez-Garcia, L. and Ulfarsson, M. O. (2011). Neuronal event detection in fmri time series using iterative deconvolution techniques. *Magnetic Resonance Imaging*, 29(3):353 – 364.

[Hlinka et al., 2011] Hlinka, J., Paluš, M., Vejmelka, M., Mantini, D., and Corbetta, M. (2011). Functional connectivity in resting-state fmri: Is linear correlation sufficient? *NeuroImage*, 54(3):2218–25.

[Hoerl and Kennard, 1970] Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

[Hulianytskyi and Omelianchyk, 2012] Hulianytskyi, L. and Omelianchyk, D. (2012). Development and applications of integral budget index. *Problems of Computer Intellectualization*, page 227–241.

[Hurvich and Tsai, 1989] Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

[Jaccard, 1901] Jaccard, P. (1901). Etude comparative de la distribution florare dans une portion des apes et des jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, pages 547–579.

[James et al., 2009] James, G. M., Radchenko, P., and Lv, J. (2009). DASSO: connections between the Dantzig selector and LASSO. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):127 – 142.

[Johnson et al., 1994] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1, 2nd Edition*. Wiley.

[Kanwisher et al., 1997] Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, 17(11):4302–4311.

[Karahanoğlu et al., 2013] Karahanoğlu, F. I., Caballero-Gaudes, C., Lazeyras, F., and Ville, D. V. D. (2013). Total activation: fmri deconvolution through spatio-temporal regularization. *NeuroImage*, 73:121 – 134.

[Khalidov et al., 2011] Khalidov, I., Fadili, J., Lazeyras, F., Ville, D. V. D., and Unser, M. (2011). Activelets: Wavelets for sparse representation of hemodynamic responses. *Signal Processing*, 91(12):2810 – 2821. Advances in Multirate Filter Bank Structures and Multiscale Representations.

[Kosub, 2016] Kosub, S. (2016). A note on the triangle inequality for the Jaccard distance. *arXiv:1612.02696*.

[Kořenek and Hlinka, 2020] Kořenek, J. and Hlinka, J. (2020). Causal network discovery by iterative conditioning: comparison of algorithms. *Chaos: An Interdisciplinary Journal of Nonlinear Science*.

[Lechnerová and Lechner, 2010] Lechnerová, R. and Lechner, T. (2010). Aplikace bodových procesu při analýze veřejné správy v ČR. *Information Bullentin of the Czech statistical society*, 22(3):81–88.

[Lechnerová and Lechner, 2013] Lechnerová, R. and Lechner, T. (2013). Analýza časových řad formální komunikace obcí. *Information Bullentin of the Czech statistical society*, 24:63–70.

[Lohmann et al., 2012] Lohmann, G., Erfurth, K., Mueller, K., and Turner, R. (2012). Critical comments on dynamic causal modelling. *NeuroImage*, 59:2322–9.

[López and Peters, 2010] López, D. M. and Peters, G. F. (2010). Internal control reporting differences among public and governmental auditors: The case of city and county circular a-133 audits. *Journal of Accounting and Public Policy*, 29(5):481 – 502.

[Mahmoud et al., 2007] Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007). A change point method for linear profile data. *Quality and reliability engineering international*, 23:247–268.

[Maiboroda, 2008] Maiboroda, R. (2008). *Otsiniuvannia ta klasyfikatsiia za sposterezhenniamy iz sumishi. Monohrafiia.* Publishing and printing center "Kyjiv University".

[Maiboroda and Sugakova, 2012] Maiboroda, R. and Sugakova, O. (2012). Statistics of mixtures with varying concentrations with application to DNA microarray data analysis. *Journal of Nonparametric Statistics*, 24(1):201–215.

[Marinazzo et al., 2011] Marinazzo, D., Liao, W., Chen, H., and Stramaglia, S. (2011). Nonlinear connectivity by Granger causality. *NeuroImage*, 58(2):330 – 338.

[McLachlan and Peel, 2000] McLachlan, G. and Peel, D. (2000). *Finite Mixture models*. Wiley-Interscience.

[Moller and Waagepetersen, 2004] Moller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC.

[Neumaier and Schneider, 2001] Neumaier, A. and Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.*, 27(1):27–57.

[Ogawa et al., 1990] Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic-resonance-imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872.

[Ogawa et al., 1992] Ogawa, S., Tank, D. W., Menon, R. S., Ellermann, J. M., Kim, S.-G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation - functional brain mapping with magnetic-resonance-imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13):5951–5955.

[Parkinson and Roseland, 2002] Parkinson, S. and Roseland, M. (2002). Leaders of the pack: An analysis of the canadian 'sustainable communities' 2000 municipal competition. *Local Environment*, 7(4):411–429.

[Poldrack, 2007] Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, 2(1):67–70.

[Pradhan and Kundu, 2011] Pradhan, B. and Kundu, D. (2011). Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, 81(9):1187–1198.

[Prokešová and Dvořák, 2014] Prokešová, M. and Dvořák, J. (2014). Statistics for Inhomogeneous Space-Time Shot-Noise Cox Processes. *Methodology and Computing in Applied Probability*, 16(2):433–449.

[Rhoad et al., 1991] Rhoad, R., Milauskas, G., and Whipple, R. (1991). *Geometry for Enjoyment and Challenge*. McDougal Littell; New Edition edition.

[Rish, 2001] Rish, I. (2001). An empirical study of the naive Bayes classifier. Technical report, IBM Research Division Thomas J. Watson Research Center.

[Santosa et al., 2019] Santosa, H., Fishburn, F., Zhai, X., and Huppert, T. J. (2019). Investigation of the sensitivity-specificity of canonical- and deconvolution-based linear models in evoked functional near-infrared spectroscopy. *Neurophotonics*, 6(2):1 – 10.

[Schoenberg, 2005] Schoenberg, F. P. (2005). Consistent parametric estimation of the intensity of a spatial–temporal point process. *Journal of Statistical Planning and Inference*, 128(1):79 – 93.

[Seghouane and Ferrari, 2019] Seghouane, A. and Ferrari, D. (2019). Robust Hemodynamic Response Function Estimation From fNIRS Signals. *IEEE Transactions on Signal Processing*, 67(7):1838–1848.

[Smith et al., 2011] Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., and Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage*, 54(2):875 – 891.

[Sreenivasan et al., 2015] Sreenivasan, K. R., Havlicek, M., and Deshpande, G. (2015). Nonparametric hemodynamic deconvolution of fMRI using homomorphic filtering. *IEEE Transactions on Medical Imaging*, 34(5):1155–1163.

[Stacy, 1973] Stacy, E. W. (1973). Quasimaximum likelihood estimators for two-parameter gamma distributions. *IBM Journal of Research and Development*, 17(2):115–124.

[Stoyan et al., 1995] Stoyan, S., Kendall, W. S., and Mecke, J. (1995). *Stochastic Geometry and its Applications. Second Edition.* John Wiley & Sons.

[Sugakova, 2006] Sugakova, O. (2006). Classification of components of a mixture. *Theory of Probability and Mathematical Statistics*, 72:157–166.

[Tan et al., 2017] Tan, F. M., Caballero-Gaudes, C., Mullinger, K. J., Cho, S.-Y., Zhang, Y., Dryden, I. L., Francis, S. T., and Gowland, P. A. (2017). Decoding fMRI events in sensorimotor motor network using sparse paradigm free mapping and activation likelihood estimates. *Human Brain Mapping*, 38(11):5778–5794.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society Series B-methodological*, 58(1):267–288.

[Tukey, 1977] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Publishing Company.

[Valdes-Sosa et al., 2011] Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J., and Friston, K. (2011). Effective connectivity: Influence, causality and biophysical modeling. *NeuroImage*, 58(2):339 – 361.

[Willett et al., 1998] Willett, P., Barnard, J. M., and Downs, G. M. (1998). Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, 38(6):983–996.

[Wu et al., 2013] Wu, G.-R., Liao, W., Stramaglia, S., Ding, J.-R., Chen, H., and Marinazzo, D. (2013). A blind deconvolution approach to recover effective connectivity brain networks from resting state fMRI data. *Medical Image Analysis*, 17(3):365 – 374.

[Wu and Marinazzo, 2015] Wu, G.-R. and Marinazzo, D. (2015). Retrieving the hemodynamic response function in resting state fMRI: methodology and applications. *PeerJ PrePrints*, 3:e1317v1.

[Yanagimoto, 1988] Yanagimoto, T. (1988). The conditional maximum likelihood estimator of the shape parameter in the gamma distribution. *Metrika*, 35(1):161–175.

[Yao et al., 2018] Yao, Y., Raman, S. S., Schiek, M., Leff, A., Frässle, S., and Stephan, K. E. (2018). Variational bayesian inversion for hierarchical unsupervised generative embedding (huge). *NeuroImage*, 179:604 – 619.

[Zar, 2005] Zar, J. H. (2005). Spearman rank correlation. *Encyclopedia of Biostatistics*, 7.

[Zhang, 2013] Zhang, J. (2013). Reducing bias of the maximum likelihood estimator of shape parameter for the gamma Distribution. *Computational statistics*, 28(4):1715–1724.

[Zou et al., 2007] Zou, H., Hastie, T., and Tibshirani, R. (2007). On the ”degrees of freedom” of the LASSO. *Annals of Statistics*, 35(5):2173–2192.

LIST OF PUBLICATIONS

IMPACTED JOURNALS

- **Pidnebesna, A.**, Fajnerová, I., Horáček, J., and Hlinka, J. (2020+). Recovering Neuronal Signal from Hemodynamic Response: the Mixture Components Inference Approach. Human Brain Mapping. Submitted.
  www.biorxiv.org/content/10/1101/2019.12.19.876508v1
- **Pidnebesna, A.**, Tomeček, D., and Hlinka, J. (2018). BRAD: Software for brain activity detection from hemodynamic response. Computer Methods and Programs in Biomedicine, 156:113 – 119.

PEER-REVIEWED JOURNALS

- **Pidnebesna, A.**, Helisová, K., and Staněk, J. (2018). Statistical analysis of dependencies among submissions to municipalities in the Czech Republic. Information Bulletin of the Czech statistical society, 29(3):1–19.
- **Pidnebesna, A.**, Helisová, K., Dvořák, J., Lechnerová, R., and Lechner, T. (2016). Statistical analysis and modelling of submissions to municipalities in the Czech Republic. Information Bulletin of the Czech statistical society, 27(4):1–18.

PAPERS IN PROCEEDINGS

- **Pidnebesna, A.** (2017). Autoregressive models of submissions to municipalities in Czech Republic. In Proceedings of the XII IEEE International Conference CSIT-2017 & International Workshop on Inductive Modeling. Vezha&Co.
- **Pidnebesna A.** (2017). Statystychnyy analiz zalezhnostey mizh specialnymy typamy zvernen' do municipalitetiv Cheskoji Respubliky. Intellectual Systems for Decision Making and Problems of Computational Intelligence: Conference Proceedings. – Kherson: PP Vyshemirsky V. S., 2017. S. 130-132.
- **Pidnebesna A.**, Helisová K., Dvořák J., Lechnerová R., Lechner T. (2016). Statistical Modelling of Submissions to Municipalities in the Czech Republic. Proceedings of the 31st International Workshop on Statistical Modelling, Vol.2, Rennes, July 4-8, 2016. S. 113-116.
- **Pidnebesna A.** (2016). Statystychnyy analiz i modeliuvannia chysla zvernen' do poshtovych ustanov v Cheskiy Respublici. Intellectual Systems for Decision Making and Problems of Computational Intelligence: Conference Proceedings. – Kherson: PP Vyshemirsky V. S., 2016. S. 136-138.

The participation of all authors is equivalent.