Vilnius University

Faculty of Mathematics and Informatics

MSc Thesis

Registration of Functional Data

Funkcinių Duomenų Registravimas

Andrius Petronis

Vilnius 2017

Ekonometrinės analizės katedra

Darbo vadovas prof. Alfredas Račkauskas
Darbo recenzentas lekt. dr. Jurgita Markevičiūtė

Darbas apgintas 2017 m. sausio 11 d.
Darbas įvertintas _____

Registravimo Nr. _____
2017-01-03 _____

Registration of Functional Data

Abstract

This paper investigates both landmark and continuous registration methods for functional data that are expected to reduce phase variation between observations. The empirical part is dedicated to the analysis of IT system issues within a company, which satisfies all necessary conditions of the non-homogeneous Poisson process. Results reveal the existence of two clusters that could be registered separately using a two-step approach. This leads to lower variance and more precise representation of the underlying process, compared to unregistered cases, while there are no significant changes in pointwise probability density functions. Also, more than 95% of all data points lies within a 2-σ range away from the mean function, while only less than 40% of functions are fully covered with these boundaries. Moreover, registration reduces maximum deviations from the mean functions. Therefore, these results correspond to the existing literature and extend knowledge about potential benefits of applied registration.

Key words: landmark registration, continuous registration, time-warping functions, k-means clustering, non-homogeneous Poisson process.

Funkcinių duomenų registravimas

Santrauka

Šiame darbe apžvelgiami funkcinių duomenų registravimo, pagal konkrečius taškus ar visą intervalą, metodai, kurie turėtų padėti sumažinti stebinių tarpusavio variaciją. Empirinė dalis yra skirta IT sistemos sutrikimų dienos metu, atitinkančių nehomogeninį Puasono procesą, tyrimui. Gauti rezultatai padeda atskleisti du duomenų klasterius, kurie gali būti modeliuojami atskirai, naudojantis dviejų žingsnių metodu. Tai leidžia sumažinti duomenų variaciją bei geriau atspindi tikrąjį procesą lyginant su neregistruotais stebiniais, tačiau tai neturi reikšmingos įtakos turimų duomenų skirstiniui konkrečiuose taškuose. Taip pat, daugiau, nei 95% stebinių patenka į 2-σ intervalą nuo vidurkio funkcijos, nors tik mažiau, nei 40% visų funkcijų yra pilnai padengiami šių rėžių. Taip pat, registracija sumažina maksimalius skirtumus lyginant su vidurkio funkcija. Gauti rezultatai sutampa su kitų empirinių darbų išvadomis bei padeda praplėsti žinias apie funkcinių duomenų registracijos naudą.

Raktiniai žodžiai: registravimas pagal konkrečius taškus, registravimas per visą intervalą, laiko deformacijos funkcijos, k-vidurkių klasterizavimas, nehomogeniškas Puasono procesas.

Contents

List of Tables

List of Figures

1 Introduction

Functional data analysis has long provided an opportunity to study intrinsically smooth processes that occur in various fields of science or everyday situations. However, the subject receives relatively less attention compared to the usual multivariate settings, mostly because it requires maintaining a strict ordering on dimensions. In other words, functional processes are usually investigated over a given time period or some specific development phases. Despite these limitations, the area has expanded rapidly in the recent years due to an increasing amount of available information worldwide, which offers more options for modelling functional processes or other unstructured data sources, such as images, sounds or shapes.

In many cases, multiple functional observations are available for the analysed period, but those are not necessarily aligned properly in terms of phase variation. For instance, one might need to investigate maximum values which do not occur at the exact same moment. Therefore, various registration techniques are applied in such situations so that researchers could focus on amplitude variations only. As a result, more accurate representation of the mean value is achieved while having less noise between functional observations.

Despite different registration methods have been introduced in the empirical literature, their behaviour vastly depends on the particular problem at hand. Hence, the main purpose of this paper is to investigate both landmark and continuous registration methods in situations, when there are several clusters within a given sample. Also, this paper introduces an option to model the underlying process using its mean value and variance only, given that there is no exact causal relationship for the occurrence of random events.

More specifically, the case study analyses a volume of IT system issues that occur during the day in which an arrival rate could be characterized as the non-homogeneous Poisson process. In other words, it counts a total number of requests within each 5-minute interval that has a location-dependent Poisson parameter. Also, data has been smoothed using Fourier basis functions which reveals two large peaks in the underlying process. These are then registered using landmark and continuous methods.

Empirical findings suggest that continuous registration leads to the largest reduction in variance for a given sample of functions, while the result could be improved using a two-step approach. It means that one could register all observations, check for possible clusters and then choose a new registration method for those distinct groups. In addition, this paper concludes that

there are no significant changes in pointwise probability density functions before and after the registration, as the normality is rejected in both cases. Finally, more than 95% of all data points are within 2-σ range away from the mean function, while less than 40% of functions are fully covered by these boundaries. Also, maximum deviations from the mean function are reduced using registration which gives an improved representation of the underlying process.

The following section provides an overview of the existing literature in this field, including some history of the functional data analysis, registration methods and applied case studies. Then, the methodology part introduces some mathematical notations and formulas necessary for the analysis, while case study gives an explicit description of the available dataset. This paper is concluded with the main empirical findings as well as provides some possible options for further discussions in this field.

2 Literature Review

The following section is dedicated to the related literature, which introduces the main ideas behind the topic of functional registration and its development throughout the years. First, it presents functional data in general and then focus on papers with proposed theoretical registration methods. Finally, the section is concluded with several recent examples of actual cases studies in this field.

2.1 History of functional data analysis

Functional data analysis (FDA) is a field of statistics that considers each element in a sample being some function. Hence, the topic is closely related to the registration literature, because it provides a theoretical background for data smoothing techniques, estimation of derivatives or principal components. A general overview has been provided in Wang et al. (2016) [21], where authors summarize the current progress in this area. Also, there is a proposed split between the *first* and *next* generation of functional data, where the latter is described as some complex data objects that include shapes and images.

The term *functional data analysis* has been initially introduced in Ramsay (1982) [14] and Ramsay and Dalzell (1991) [15], while the general subject is much older and its origins could be found in Grenander (1950) [6] or Rao (1958) [19]. According to the literature, functional data has intrinsically infinite number of dimensions, which causes issues for both theoretical and empirical development of this area. On the other hand, it offers various opportunities for more advanced data analysis given its complex nature. Also, an increase in computational capacity has made it easier to apply these theoretical techniques in practise over the last few decades.

An explicit review of key mathematical concepts of functional data analysis has been provided in Hsing and Eubank (2015) [8]. This incorporates foundations for the probability theory from the perspective of random elements in Hilbert spaces as well as continuous time stochastic processes. In addition, inferential methods have been investigated in Horvath and Kokoszka (2012) [7], where authors present different aspects of statistical hypothesis testing for functional data settings. These include model specification, change points and functional principal component tests.

On the other hand, complicated theoretical literature might be less suitable for researchers with a limited statistical background. Therefore, much less mathematical notation is used in Ramsay and Silverman (2005) [18],

which allows introducing the topic among people, who are not directly linked to the discipline of statistics. This is particularly important given that lots of functional data is generated in various fields of science such as biology, medicine, physics or economics. As a result, authors offer a review that could build knowledge for many new empirical researches.

2.2 Review of existing registration methods

As it has been mentioned before, registration methods have some direct dependency on functional data analysis, because it deals with situations when phase variation exists among functional observations. In other words, the underlying process could sometimes be represented more accurately once the registration is applied. The idea is to manipulate domain values in order to get a better phase alignment within a given sample of functions.

One of the main building blocks of these methods is known as landmark registration, which has been studied extensively in Bookstein (1991) [1], Kneip and Gasser (1992) [10], Gasser and Kneip (1995) [3]. It allows detecting some specific touch-points in a set of functions using both manual and automated techniques. Despite it is a relatively straightforward task to identify key landmarks, there might be situations when this becomes complicated due to some unexpected functional shapes.

For this reason, an algorithm to register functional data, based on some baseline functions rather than specific landmarks, has been proposed in Kneip and Ramsay (2008) [11]. The main idea is that functional observations should deviate only in terms of amplitude variation and it is possible to minimize eigenvalue of some underlying matrix [16]. In addition, authors provide an explicit theoretical justification for functional registration, which includes mathematical derivations for consistency as well as quantitative separation between phase and amplitude variations.

An increasing popularity of this subject has led to a development of several new registration techniques that are provided in both theoretical and empirical literature. For instance, a pairwise curve synchronisation has been introduced in Tang and Müller (2008) [20], which exploits all the available information in given a sample in order to obtain robust estimates for individual time-warping functions.

Other methods involve curve alignment by moments (James, 2008) [9], self-modelling warping functions (Gervini & Gasser, 2004) [4], alignment using dynamic time-warping (Wang & Gasser, 1997) [22] or some non-parametric estimations (Gervini & Gasser, 2005) [5]. Despite all these suggested methods have different theoretical justifications, the main idea

behind functional registration is to identify the most suitable set of domain transformations for a dataset at hand.

2.3 Applied cases studies in this field

Statistical methods described in the literature of functional data analysis are applicable for both registered and unregistered observations, which provides many different opportunities for data modelling. One of these situations is presented in Račkauskas and Laukaitis (2002) [12], where authors incorporate functional autoregressive models to predict a total number of financial transactions during some specific weekdays. Hence, it sometimes possible to analyse financial data using a measure of intensity and then predict future volumes with relative accuracy.

Possible applicability of functional registration methods has also been presented in Ramsay and Silverman (2002) [17], which is later exploited with working examples using R and MATLAB statistical packages (Ramsay & Silverman, 2009) [16]. More specifically, authors investigate children growth functions, yearly volumes of non-durable goods, handwriting recognition or even the juggling process. As a result, it is possible to see opportunities for applying registration techniques with datasets that are less frequently analysed in the empirical literature.

Another advantage of functional registration is the opportunity to analyse images or 3-dimensional objects. These situations have been presented in Zhang et al. (2015) [23], where authors consolidate the knowledge about both registration and usual shape analysis. So, applied cases are not limited to observations over some time period only, but could also be considered with other types of data as well. This definitely offers many theoretical and empirical challenges for future researches.

3 Methodology

The following section is dedicated to the review of relevant theories in this field of statistics, which is later applied in the proposed case study. So, the main focus is on both mathematical notations and definitions that are necessary for the empirical part of this paper.

3.1 Functional data analysis techniques

Basic definitions for the functional data analysis are provided in Müller (2005) [13], where author identifies it as *a sample of random curves that represents i.i.d realizations of an underlying stochastic process.* In general, functions are usually smooth and integrable that allows investigating their derivatives (i.e. velocity or acceleration). More precisely:

$$y_{ij} = x_i(t_{ij}) + \epsilon_{ij} \tag{3.1}$$

where t represents some continuum (i.e. time) so that $x_i(t)$ is smooth.

Functional data analysis provides several advantages over the multivariate approach, where neighbourhood and order of observations are less relevant. For instance, one could easily change a position of some component in a multivariate data vector without any effect on the further statistical analysis, while functional process would necessarily be impacted.

In general, functional data usually consists of some high-frequency data that represents complex underlying processes. So, the applied smoothing techniques allow achieving generalization that is needed for the investigation of derivatives. This becomes even more complicated when functions are representing many repeated observations (i.e. daily temperature) or have multiple dimensions (i.e. shapes in 3D). As a result, functional principal component analysis, clustering, time-warping and other methods are used to reduce dimensions and achieve more reliable empirical estimates. Further sections provide a brief summary of these techniques and their interpretation.

3.1.1 Basis functions

Basis functions are invoked to model processes that are both complicated and unpredictable [16]. The main objective is to have parameters that could be estimated easily while maintaining a complex functional form. Despite various methods have been proposed in the empirical literature, this paper

emphasizes B-Spline and Fourier basis functions that are most frequently used in similar studies.

A linear combination of basis functions $\phi_k,\ k = 1, ..., K$ could be expressed in the mathematical notation:

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) = \mathbf{c}'\phi(t) \tag{3.2}$$

which is called a basis function expansion with coefficients $c_1, c_2, ..., c_K$.

Here, $\mathbf{c}$ represents a vector of K coefficients and ϕ is a vector of length K that contains basis functions. However, if a sample consists of N functions then this notation becomes:

$$\mathbf{x}(t) = \mathbf{C}\phi(t) \tag{3.3}$$

where $\mathbf{x}(t)$ is a vector of length N incorporating functions $x_i(t),\ i = 1, ..., N$ while the coefficient matrix $\mathbf{C}$ has N rows and K columns. Also, t is not included in the notation when one defines functions in a general sense rather than at a specific time value.

Then, assuming functions are smooth, one could define derivatives as:

$$\frac{d^m x(t)}{dt^m} = \mathbf{c}\ \frac{d^m \phi(t)}{dt^m} \tag{3.4}$$

with $m = 1$ and $m = 2$ representing *velocity* and *acceleration* respectively. Derivatives could reveal important dynamics of the underlying process and are studied extensively in the related empirical literature.

Also, the selection of some particular basis functions usually depends on a given dataset. For example, Fourier methods are more suitable for periodic functions in which observations are equally spaces. More specifically, these basis functions are sine and cosine pairs of growing frequency:

$$1,\ sin(\omega t),\ cos(\omega t),\ sin(2\omega t),\ cos(2\omega t),\ \ldots,\ sin(m\omega t),\ cos(m\omega t) \tag{3.5}$$

where constant ω and period T has a relationship $\omega = 2\pi/T$.

One the other hand, B-Spline basis functions tend to perform better given some complex data structures and constraints in which separate poly-

nomial segments are expected to be smooth at joins known as knots. So, the system could then be represented as:

$$number\ of\ basis\ functions = order + number\ of\ interior\ knots \quad (3.6)$$

where knots define a number of matching derivatives at each break point and determine smoothness for a given sample of functions. So, a large volume of basis functions leads to a better fit to the data, but this could also imply overfitting. In other words, the representation of some underlying process could deviate according to the selection of basis functions.

3.1.2 Least-squares estimates

Estimation of coefficients for basis functions is essential in order to get smooth representation of any noisy data. For this reason, one could apply the ordinary least-squares method.

First of all, the minimization of the sum of squared errors:

$$SSE(x) = \sum_{j=1}^{n} [y_j - x(t_j)]^2 \quad (3.7)$$

could be defined as a basis function expansion:

$$SSE(\mathbf{c}) = \sum_{j=1}^{n} [y_j - \sum_{k=1}^{K} c_k \phi_k(t_j)]^2 = \sum_{j=1}^{n} [y_j - \phi(t_j)'\mathbf{c}]^2 \quad (3.8)$$

which is based on the error model:

$$y_j = x(t_j) + \epsilon_j = \mathbf{c}'\phi(t) + \epsilon_j = \phi'(t_j)\mathbf{c} + \epsilon_j \quad (3.9)$$

Suppose $\mathbf{y}$ is a vector $(y_1, ..., y_n)'$, ϵ contains the corresponding residual values and $\mathbf{\Phi}$ is N by K matrix with basis functions $\phi_k(t_j)$, then:

$$\mathbf{y} = \mathbf{\Phi c} + \boldsymbol{\epsilon} \quad (3.10)$$

so that the coefficient vector $\mathbf{c}$ is estimated using the least-squares:

$$\hat{\mathbf{c}} = (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{y}, \quad (3.11)$$

Then, one could find a vector of fitted values using:

$$\hat{\mathbf{y}} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{y} \tag{3.12}$$

The resulting data fit depends on a number of basis functions selected so that a large volume will lead to lower smoothness. On the other hand, it is possible to introduce a penalty parameter that could adjust these estimates to find some optimal fit. This option is presented in the next section.

3.1.3 Smoothing penalties

The selection of a smoothing parameter defines curvature of the data fit. So, this could lead to different local minimum values that might be relevant for further registration purposes. Let, $J[x]$ measure the roughness of x, so that the penalized least-squares estimate becomes:

$$PENSSE = \sum_{j=1}^{n}[y_j - x(t_j)]^2 + \lambda J[x] \tag{3.13}$$

where an increase in λ leads to more smooth data fit, while the opposite situation reduces the total penalty.

Penalties are defined differently for both B-Spline and Fourier basis functions. The first one often measures a total curvature as an *integrated squared second derivative*:

$$J[x] = \int [D^2x(t)]^2 dt \tag{3.14}$$

while Fourier penalty usually is defined as *harmonic acceleration operator* [18] that uses shifted sinusoidal variation as some baseline behaviour:

$$x(t) = c_0 + a_1 \sin \omega t + b_1 \cos \omega t \tag{3.15}$$

then, if $\omega = 2\pi/T$ it is possible to define the *differential operator* as $L = \omega^2 D + D^3$, given that $\omega^2 Dx + D^3 x$ for this functions is equal to zero. As a result, an integral of this squared harmonic acceleration operator is:

$$J[x] = \int [Lx(t)]^2 dt \tag{3.16}$$

After defining $x(t) = \mathbf{c}'\phi(t) = \phi'(t)\mathbf{c}$, one could use:

$$\int [Lx(t)]^2 dt = \int \mathbf{c}'[L\phi(t)][L\phi(t)]'\mathbf{c}dt = \mathbf{c}'\mathbf{R}\mathbf{c} \tag{3.17}$$

where the penalty matrix is $[R]_{jk} = \int [L\phi_j(t)][L\phi_k(t)]'dt$ and the penalized least squares estimate for $\mathbf{c}$ could be defined as:

$$\hat{\mathbf{c}} = [\mathbf{\Phi}'\mathbf{\Phi} + \lambda\mathbf{R}]^{-1}\mathbf{\Phi}'\mathbf{y} \tag{3.18}$$

which corresponds to a linear smoother:

$$\hat{\mathbf{y}} = \mathbf{\Phi}[\mathbf{\Phi}'\mathbf{\Phi} + \lambda\mathbf{R}]^{-1}\mathbf{\Phi}'\mathbf{y} = S(\lambda)\mathbf{y} \tag{3.19}$$

Degrees of freedom for the smoothed data are equal to a number of basis functions K in the least-squares estimation. So, for the penalized version it is possible to have $K > n$, meaning that a penalty parameter reduces flexibility of the actual data fit. As degrees of freedom are determined by λ, the actual measure becomes:

$$df(\lambda) = trace[S(\lambda)], \tag{3.20}$$

$$S(\lambda) = \mathbf{\Phi}[\mathbf{\Phi}'\mathbf{\Phi} + \lambda\mathbf{R}]^{-1}\mathbf{\Phi}' \tag{3.21}$$

Therefore, the smoothing parameter λ could be estimated using a method of *generalized cross validation*:

$$GCV(\lambda) = \frac{\sum(y_i - x_\lambda(t_i))^2}{[trace(\mathbb{I} - S(\lambda))]^2} \tag{3.22}$$

In general, one could achieve an optimal fit using the least squares estimates together with some smoothing penalty. Then, it is possible to apply registration methods that are described in the next section.

3.2 Registration of functional data

Registration of the functional data is a pivotal part of this paper, given that all previous steps are needed only for the initial data preparation. Despite, there are many different options available in the empirical literature, the main focus is on landmark and continuous registration methods. However, the first step requires defining time-warping functions in general.

3.2.1 Time-warping functions

The main idea behind the registration process is to perform some transformation of time that is achieved using time-warping functions. However, a formal definition has been provided in Tang and Müller (2008) [20].

Suppose $Y_1, Y_2, ..., Y_n$ are n continuous curves, defined over some bounded interval $\mathcal{T} = [0, T]$. Given that these curves are observed at discrete time points $t_j,\ j = 1, ..., m$, and observed data for each curve is (t_j, y_{ij}), then one could define a model:

$$y_{ij} = Y_i(t_j) + \epsilon_{ij} = X_i\{h_i^{-1}(t_j)\} + \epsilon_{ij}, \quad t_j \in \mathcal{T} \tag{3.23}$$

where X is a random function defined over $\mathcal{T}$ with independent realizations X_i, while error terms are defined as independently distributed random noise with properties $E(\epsilon) = 0$ and $E(\epsilon^2) = \sigma^2 < \infty$. Then time-warping random functions X_i could be modelled using:

$$X_i(t) = \mu(t) + \delta Z_i(t), \quad \text{for } t \in \mathcal{T} \tag{3.24}$$

where μ is some fixed function, while Z_i represents independent random trajectories. These are realizations of a process Z with $E\{Z(t)\} = 0$ and $E\{Z^2(t)\} < \infty$ for $t \in \mathcal{T}$. Also, δ is some small constant converging to zero together with an increasing sample size. The functions $h_i : \mathcal{T} \to \mathcal{T}$ are time-warping functions that measure realizations of random synchronization function h. Hence, the internal time scale of a trajectory is defined by the inverse mapping h^{-1}. Also, there are several conditions that these random time-warping functions h have to satisfy, i.e. common endpoints:

$$h(0) = 0, \quad h(T) = T \tag{3.25}$$

Then, strict monotonicity, which requires to obey then fact that time always moves forwards as well as ensures invertibilty of h functions:

$$h(t_1) < h(t_2), \quad \text{for } 0 \leq t_1 \leq t_2 \leq T \tag{3.26}$$

Finally, an average identity, defined as:

$$E\{h(t)\} = t \quad \text{for } t \in \mathcal{T} \tag{3.27}$$

In general, this is a formal definition of time-warping functions. On the other hand, an illustrative example provided in Ramsay et al. (2008) [16] might give some better understanding of these time transformations. Authors investigate pubertal growth spurts for a sample of girls in which the focus is on acceleration functions crossing zero values (Figure 1). The goal is to align curves according to the mean point at around 11-12 years:

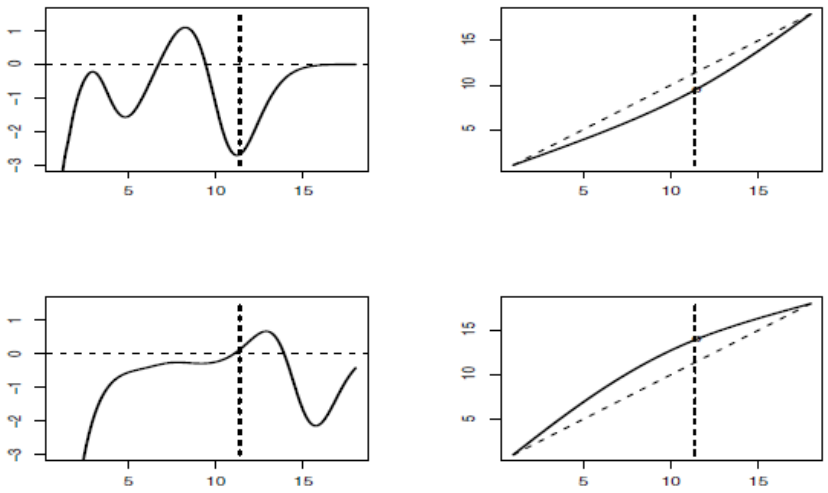


Figure 1: An example of time-warping functions for growth data of girls (Ramsay et al. (2008)) [16]

Two examples of growth acceleration curves are provided on the left-hand-side, while their corresponding time-warping functions are presented on the right-hand-side. In this situation, both functions would be shifted to the left or right in order to match the dotted line at the zero crossing value (Figure 2).

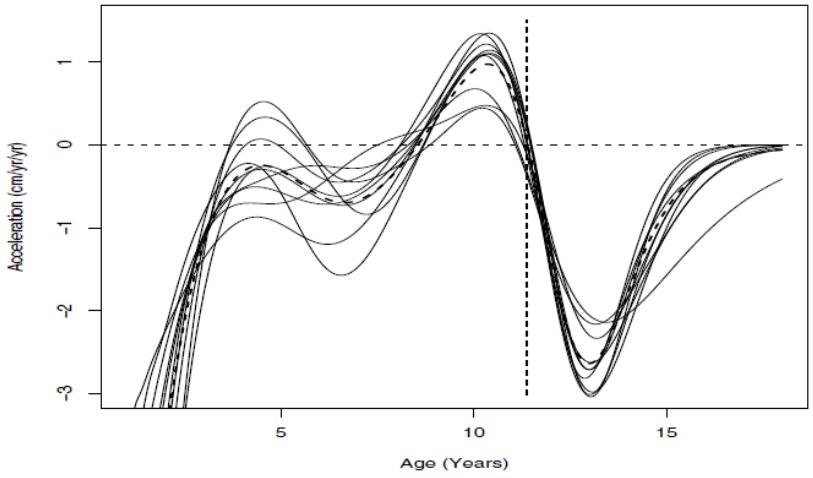


Figure 2: A full sample of registered functions (Ramsay et al. (2008)) [16]

As a result, it would be possible to measure an extent of these growth spurts which otherwise suffers from phase variation. On the other hand, results largely depend on the applied registration method. The example above uses the landmark approach, but it is also possible to introduce continuous registration as well.

3.2.2 Registration methods

Landmark registration offers a relatively straightforward approach that is often used as the first option for the analysis. Its main objective is to detect some specific data points for each observation (i.e. maximum, minimum, zero-crossing, etc.) and align them to the corresponding mean value. For instance, functions of yearly temperature could be synchronized based on a maximum value throughout this period. As a result, these new functions would provide a better representation of the mean function in terms of amplitude variation.

More specifically, one needs to select some specific points $t_{i1}, ..., t_{iK}$ for each function $x_i(t)$ and then decide on a general reference point $t_{01}, ..., t_{0K}$, given the constrains:

$$w_i(0) = 0, \quad w_i(T) = T \tag{3.28}$$

so that:

$$w_i(t_{ij}) = t_{0j} \tag{3.29}$$

Usually, it is possible to perform automated landmark registration for each function within a pre-defined interval. However, issues could occur in situation when the functional shape is highly inconsistent.

Continuous registration does not suffer from such issues, because it uses information of the entire curves rather then specific landmarks. The main idea behind this method is that registered functions are expected deviate only in terms of amplitude, which is measured as a proportion between registered curve $x[h(t)]$ and target curve $x_0(t)$ over a range of t values. More specifically, plotting values of these functions should provide a straight line that goes through the origin. As a result, principal component analysis of the order two matrix $\mathbf{T}(h)$ of integrated products of these values should essentially detect one component, and the smallest eigenvalue close to zero:

$$\mathbf{C}(h) = \begin{bmatrix} \int \{x_0(t)\}^2 dt & \int x_0(t)x[h(t)]dt \\ \int x_0(t)x[h(t)]dt & \int \{x[h(t)]\}^2 dt \end{bmatrix} \tag{3.30}$$

Therefore, it is necessary to estimate h so that it minimizes the lowest eigenvalue of $\mathbf{C}(h)$. As a result, continuous registration tends to capture the underlying mean function better compared to landmark registration. Also, it is possible to use different target curves, while mean function is often applied as a primary choice for such registration.

3.2.3 Phase and amplitude separation

A method to evaluate an impact of registration in terms of phase and amplitude variation has been proposed in Kneip and Ramsay (2008) [11]. The idea is to compare actual data fit before and after registration is applied and identify the most suitable option for a given dataset.

For instance, suppose that for a sample of N observations, x_i represents i^{th} unregistered observation, while y_i and h_i are registered and time-warping functions respectively. Also, $\bar{x}$ and $\bar{y}$ indicate some mean values. Hence, the *total mean square error* could be defined as follows:

$$MSE_{total} = N^{-1} \sum_{i=1}^{N} \int [x_i(t) - \bar{x}(t)]^2 dt \tag{3.31}$$

In order to quantify the amount of separation between phase and amplitude variation, one needs to define a constant C_R, which indicates covariation between deformed Dh_i and squared registered functions y_i^2:

$$C_R = 1 + \frac{N^{-1} \sum_i^N \int [Dh_i(t) - N^{-1} \sum_i^N Dh_i(t)][y_i^2 - N^{-1} \sum_i^N y_i^2(t)]dt}{N^{-1} \sum_i^N \int y_i^2(t)dt} \tag{3.32}$$

So, mean square errors for both amplitude and phase variations could be separated using formulas:

$$MSE_{amp} = C_R N^{-1} \sum_{i=1}^{N} \int [y_i(t) - \bar{y}(t)]^2 dt \tag{3.33}$$

$$MSE_{phase} = C_R \int \bar{y}^2(t)dt - \int \bar{x}^2(t)dt \tag{3.34}$$

which leads to a general definition:

$$MSE_{total} = MSE_{amp} + MSE_{phase} \tag{3.35}$$

and

$$R^2 = \frac{MSE_{phase}}{MSE_{total}} \tag{3.36}$$

In general, a value of MSE_{phase} represents the amount of phase variation that is reduced by the applied registration method. Therefore, higher R^2 number suggests a better fit for the unregistered data and should be considered for the actual registration. Also, it is possible to have MSE_{phase} as a negative value, which means that the selected registration method does not improve the fit. Hence, it might be an indicator for overfitting the data as well as allow comparing different registration options.

3.3 Descriptive statistics for functional data

Functional data analysis has similar descriptive statistics compared to multivariate cases. So, one could easily calculate a value of the mean function:

$$\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t) \tag{3.37}$$

as well as functional variance:

$$s(t) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t) - \bar{x}(t)]^2 \tag{3.38}$$

On the other hand, it is sometimes more valuable to study covariance between different points rather than specific t values. Hence, bivariate covariance function $\sigma(s,t)$ is estimated by:

$$v(s,t) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(s) - \bar{x}(s)][x_i(t) - \bar{x}(t)] \tag{3.39}$$

where $x_i(s)$ and $x_i(t)$ are curve values at s and t respectively. Resulting estimates correspond to the variance-covariance surface which identifies areas with the largest deviations within a given sample of observations.

The vast majority of these theoretical concepts have been incorporated in the empirical part of this paper. Hence, the following two sections includes a description of a given dataset and actual results based on the functional registration techniques.

4 Case Study

4.1 Description of the available dataset

The next section describes an available dataset and its preparation for the analysis. This involves splitting and summarizing records, exclusion of weekends or public holidays, data smoothing and registration. Also, the main graphs with descriptive statistics, including mean functions and variance-covariance surfaces, are provided at the end of this paper (Appendix C).

4.1.1 Data collection process

The case study investigates a volume of IT system issues that occurred during a period of fifteen weeks (Figure 3). Data has been provided by a real company, which monitors its performance from the perspective of information technologies. More specifically, raw data file contains records as dates, specifying an exact time of the event, which are then split into 5 minute intervals. Hence, this corresponds to an intensity measure of the non-homogeneous Poisson process (Appendix D). On other hand, interesting patters are revealed only after smoothing procedures have been applied to this data using Fourier basis functions.

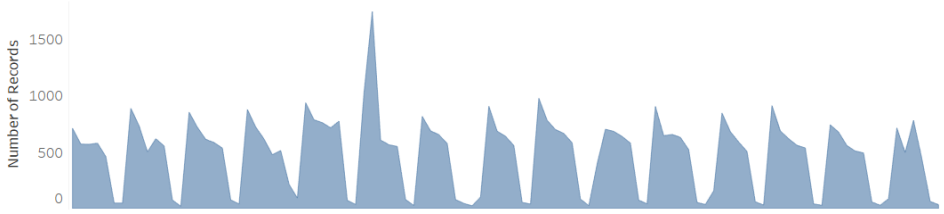
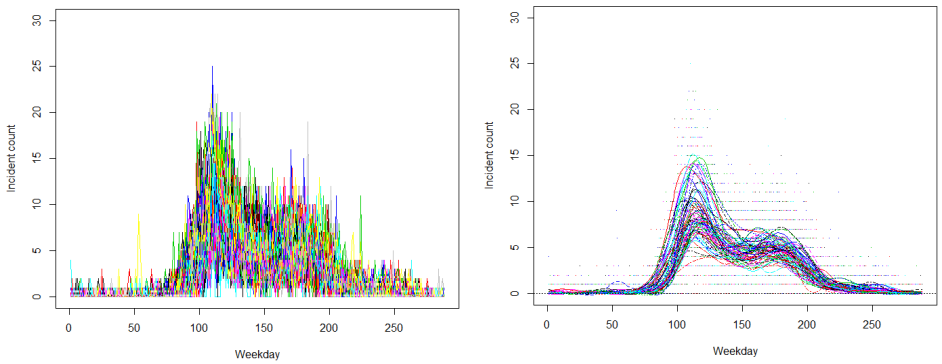


Figure 3: A number of daily requests during the period of fifteen weeks

Applied data smoothing indicates two large peaks during usual working hours, while significantly lower activity is observed at night or the lunch time. Also, weekends and public holidays have been excluded from the analysis, because there are almost no requests during these days. Moreover, one hour is added to a half of dates due to an official time shift change in the region. Hence, this provides a more natural alignment of the underlying process that could otherwise bias the registration procedure. So, the final dataset consists of 71 functions (working days) suitable for the further analysis.

4.1.2 Applied smoothing techniques

All daily observations have been smoothed using 65 Fourier basis functions together with some penalty parameter. So, resulting values show the intensity of events that occurred during the last 24 hours (Figure 4).



(a) Non-smoothed (b) Smoothed

Figure 4: A volume of daily observations divided into 5-minute intervals

In general, one could argue that daily activity increases starting from 7 a.m. (85^{th} interval) and then drops at around 3 p.m. (180^{th} interval), while other periods have a much lower volume of IT issues. Hence, the primary goal of this paper is to correctly capture those active periods, especially peak values, and their deviations from the mean function.

Also, the *generalized cross-validation* method has been used in order to estimate a smoothing parameter λ. It determines a roughness of the analysed sample of functions and could distort the registration process if selected inappropriately.

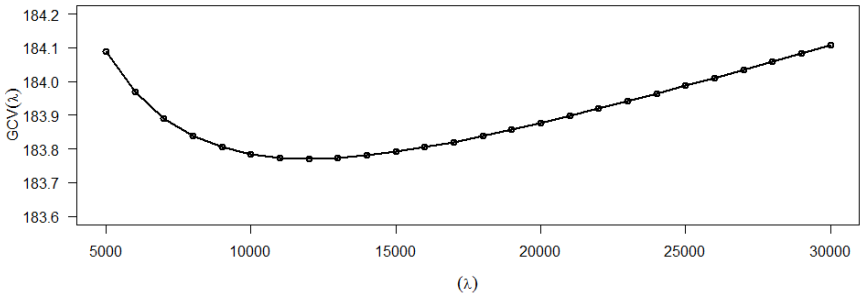
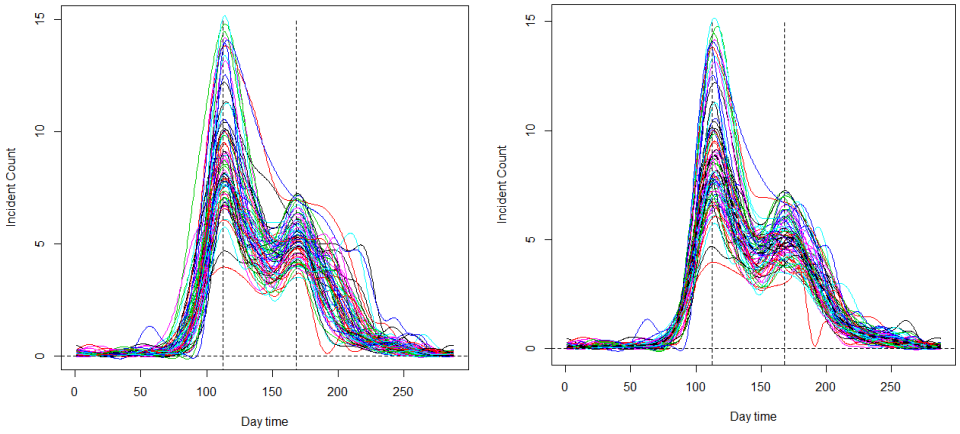


Figure 5: Estimated λ values using the generalized cross-validation method

According to the graph, $\lambda = 12,000$ provides the best fit to the underlying data and is used for the analysis (Figure 5).

4.1.3 Comparison of registered functions

Both landmark and continuous registration methods have been applied to the event data after the initial pre-processing steps (Figure 6). Hence, the landmark registration approach leads to a relatively strict alignment of curves for both maximums that are set exactly to their mean values. Unfortunately, several functions are not captured properly, especially during the second peak.



(a) Landmark (b) Continuous

Figure 6: Comparison of the two registration methods

On the other hand, the continuous registration approach does not possess these limitations and provides a better alignment for the second half of the day. One could see an exact quantitative comparison between the two methods in terms of amplitude and phase separation (Table 1), while the difference is relatively small in absolute terms. Also, it might be interesting to check actual time-warping functions which are available at the end of this paper (Appendix E). Hence, it is possible to conclude that the provided synchronization is not so significant compared to the example in the methodology section. This is because unregistered peaks are relatively close to each other and phase variation is rather small. More formally, this is measured by the R^2 value in the table below. On the other hand, registration is still relevant in this case, because it allows reducing variation in a

given sample and leads to an improved representation of the mean function.

Table 1: Phase and amplitude separation for the $[85^{th} - 180^{th}]$ interval

Measure	Landmark	Continuous
MSE_{phase}	18.60	36.10
$MSE_{amplitude}$	230.54	205.73
MSE_{total}	249.14	241.83
R^2	0.075	0.149
C	0.981	1.001

All previous proposals are supported with estimated MSE_{total} values, meaning that continuous registration leads to a better alignment of all sample functions. Moreover, landmark registration reduces phase variation less during the active period when compared to unregistered cases. Therefore, continuous registration is selected as a primary option for the further analysis. In addition, detailed mathematical definitions for this approach have already been provided in the methodology section.

5 Results

This section provides an overview of the main empirical findings with an emphasis on benefits for using functional registration. Also, it offers a method to analyse a given dataset that does not have any underlying causal relationship for further modelling.

5.1 Clustering

First of all, functional observations have been clustered into two groups based on their maximum values. Therefore, it allows separating functions that have a large volume of daily requests from those with a relatively lower intensity. Also, these observations have different probability density functions and such separation could improve actual representation of the underlying process. More specifically, the k-means method has been applied for clustering:

$$J(V) = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(j)} - v_j||^2 \tag{5.1}$$

where $||x_i^{(j)} - v_j||$ is the Euclidean distance between x_i and v_j, n is a number of data points in i^{th} cluster, and k counts centres. As two centroids have been chosen for clustering, there is an evident distinction between Mondays and the rest of weekdays (Figure 7).

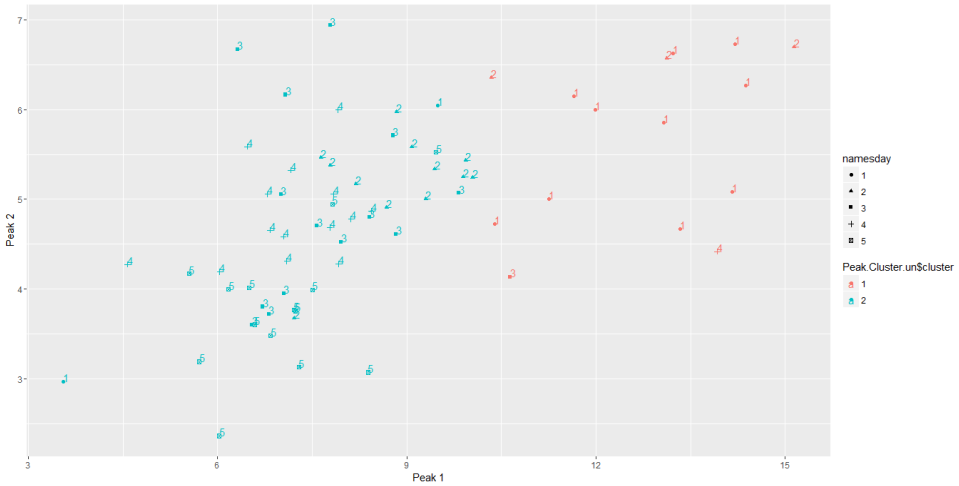


Figure 7: K-mean clustering plot (K=2)

This separation reveals that a vast majority of Mondays are included in the first cluster together with several days after public holidays (Table 2). Hence, it is likely that people are prone to raise more requests to the system after some longer periods without an access.

Table 2: Results of k-means clustering by weekdays

Cluster	Mon	Tue	Wed	Thu	Fri
Group 1	10	3	1	1	0
Group 2	2	12	14	14	14

However, these results do not necessarily imply a statistically significant difference, which could be tested using permutations t-test:

$$T(t) = \frac{|\bar{x}_1(t) - \bar{x}_2(t)|}{\sqrt{\frac{1}{n_1}Var[x_1(t)] + \frac{1}{n_2}Var[x_2(t)]}} \tag{5.2}$$

The idea is to re-sample observations 200 times and then calculate pointwise t-test values (Figure 8).

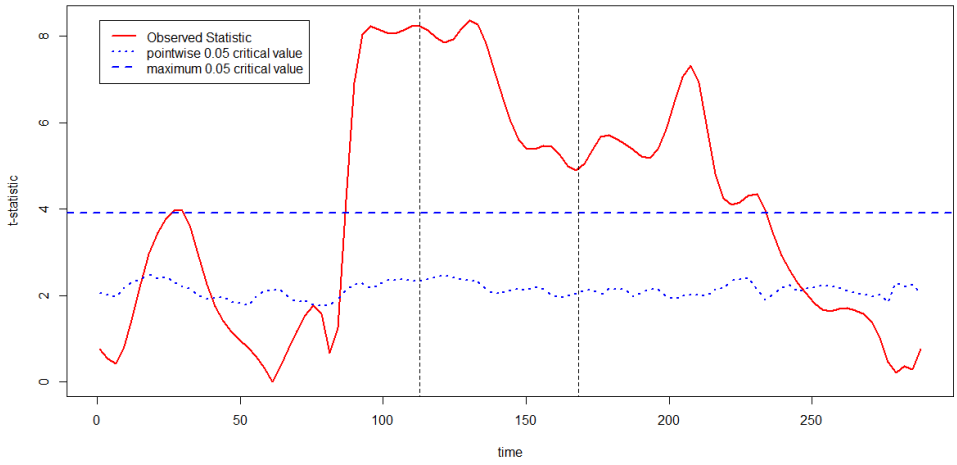


Figure 8: Permutations t-test

Results provide a strong proposition that the request volume is more intense on Mondays compared to the rest of weekdays. So, it is reasonable to register observations within these two clusters separately and then compare it to the initial estimates.

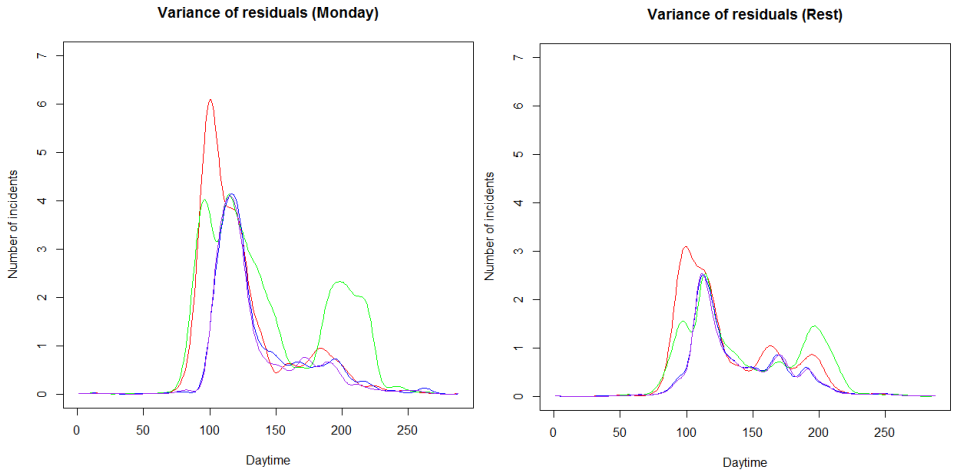
5.2 Two-step registration

Given the existence of two distinct groups of functions, one might expect that application of different registration methods could lead to more accurate estimates for both mean and variance. So, there is an option is to calculate integrated mean-squared-errors for these functions separately and choose method with the lowest value (Table 3):

Table 3: IMSE for different registration methods

Registration	Mondays (IMSE)	Rest (IMSE)
Unregistered	237.00	161.34
Landmark	304.63	150.27
Continuous	158.34	104.09
Two-step	148.95	101.07

As a result, both continuous registration methods tend to provide the most appropriate fit for a given set of observations, while the two-step approach is slightly more accurate and should be chosen over the usual method. So, an improved alignment of functional data suggests considering cluster analysis before the actual modelling.



(a) Mondays (b) Rest

Figure 9: Pointwise variation for different registration methods

In addition, landmark registration performs relatively worse, especially within a sample of Monday functions. The method suffers in terms of accuracy due to some unusual shapes around the second peak, which sometimes have more than one intense shock. So, it becomes difficult to properly capture the behaviour of these functions during the second half of the day, while continuous registration does not encounter such issues and has the overall lowest IMSE value for both groups. On the other hand, this is a likely conclusion given that continuous registration uses a mean function from the landmark approach as some baseline value.

Estimates of pointwise variances provide another useful way to compare different registration methods. Also, it corresponds to the previous conclusions that continuous method is the most suitable option for the dataset at hand. The idea is that registration should reduce phase variations and improve calculations of the mean function (Figure 9). So, one could compare these estimates with unregistered cases and decide if registration is actually needed for these observations.

In conclusion, both continuous registration methods (blue/purple) are rather similar, and they both outperform the landmark registration (green). Also, these empirical findings correspond to the previous IMSE estimates and indicate specific intervals when registration fails to reduce variance for the second peak.

5.3 Probability density functions

After the most suitable registration method is chosen, it is interesting to analyse changes in pointwise probability density functions. The main goal is to investigate if there are any significant shifts that could lead to some well-known distributions, which are modelled using mean and variance. However, it is not possible to investigate functions as an entity, so it is necessary to pre-select some specific points that could be analysed further.

Despite two-step registration provides more accurate alignment of sample observations, there are no significant changes in probability density functions for both Mondays (Figure 10) and the rest of weekdays (Figure 11). Graphs compare unregistered functions (orange) against their registered equivalents (purple). One could see that those differences are relatively minor, especially during the most active period of the day.

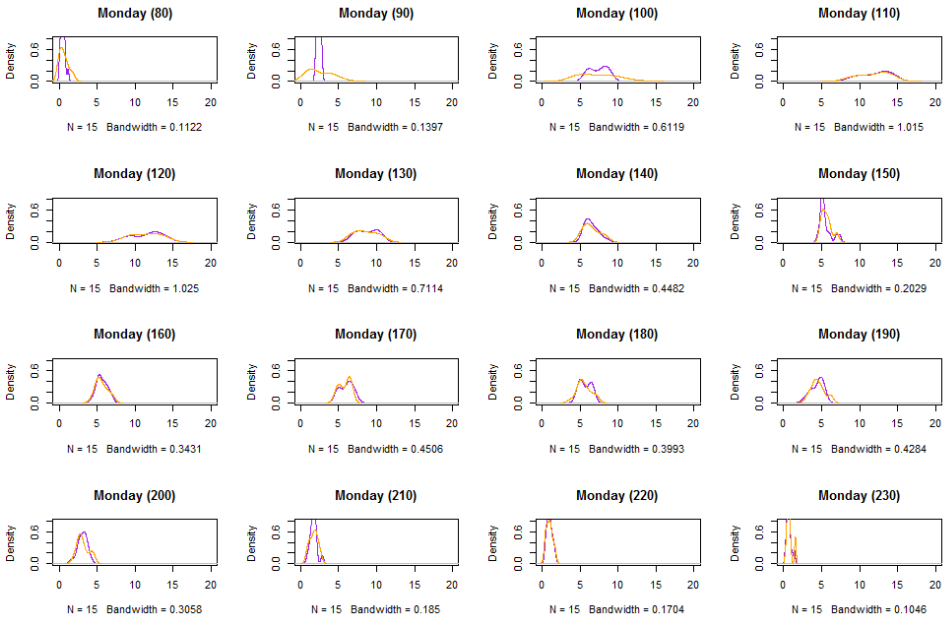


Figure 10: Pointwise probability density functions for Mondays

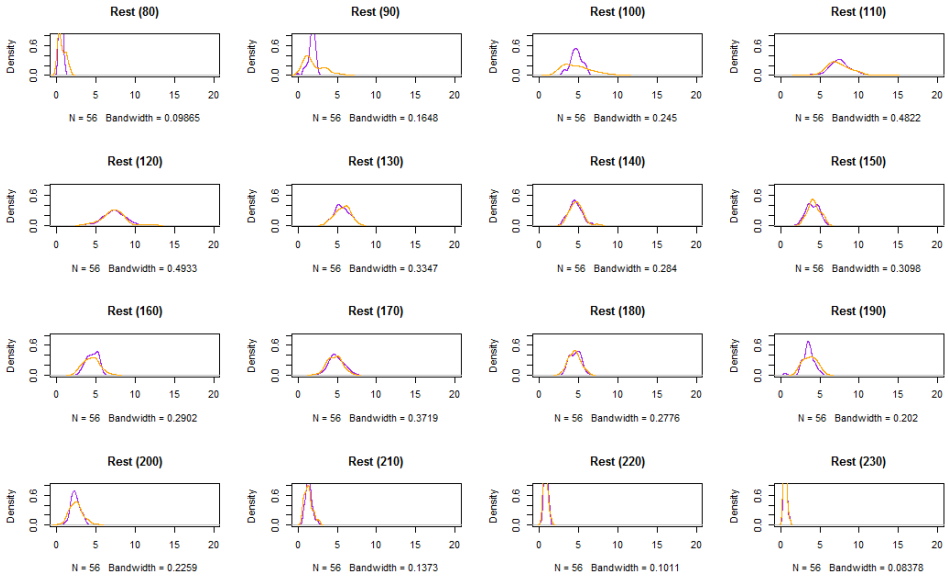


Figure 11: Pointwise probability density functions for Rest

The hypothesis of normality has been rejected using Kolmogorov–Smirnov test for each of these data points. Therefore, it is not possible to apply standard statistical techniques to model distribution of sample observations. On the other hand, one might consider some non-parametric approach that considers the arrangement of all data points.

5.4 Modelling an average behaviour using σ boundaries

There is a limited possibility to model the request intensity due to relatively random nature of the underlying process. More specifically, there are no apparent predictors that could explain this process for any given day. Hence, the following part investigates the behaviour of sample observations when compared to the estimated mean function. For instance, one could check an average number of points within different σ boundaries (Figure 12):

$$Pr\{\mu(t) - k\sigma(t) \leq x(t) \leq \mu(t) + k\sigma(t)\} = p(t) \tag{5.3}$$

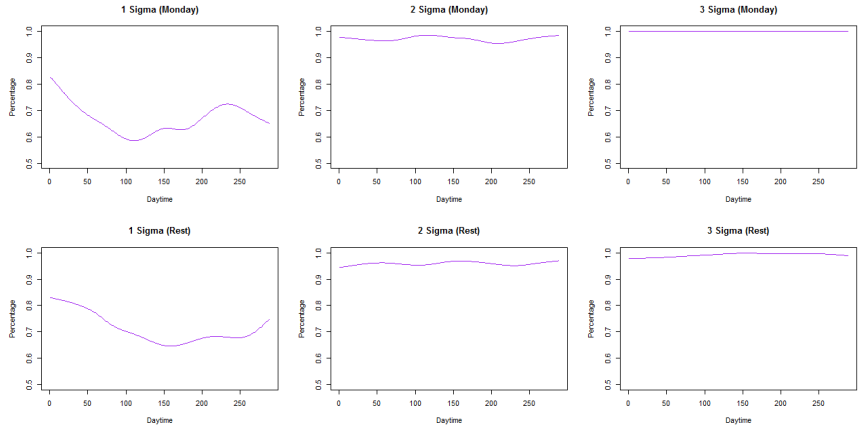


Figure 12: Average percent of values within σ boundaries (Mondays)

Results reveal that more than 95% of all values are within 2-σ intervals for both Mondays and the rest of weekdays, while the actual estimates could be found at the end of this paper (Appendix A). On the other hand, this is a generalized average value of all data points, while it is more relevant to know if functions are within these boundaries as a whole:

$$Pr\{\mu(t) - k\sigma(t) \leq x(t) \leq \mu(t) + k\sigma(t), \ \ \forall\, t \in [0;T]\} \approx$$
$$\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\left(\hat{\mu}(t) - k\hat{\sigma}(t) \leq x_j(t) \leq \hat{\mu}(t) + k\hat{\sigma}(t), \ \ \forall\, t \in [0;T]\right) = \frac{M}{N} \tag{5.4}$$

In this situation, results are less optimistic, with less then 40% of all functions within 2-σ range. So, one would be more confident using at least 3-σ interval, because it usually covers both data points and functions. On the other hand, Mondays indicate less variation in a given sample of observations compared to the mean function. Hence, it could be modelled using a narrower interval while the rest of weekdays values are more unpredictable.

5.5 Density functions for maximum deviation from the mean

Another interesting option is to analyse a probability of some large outliers compared to the mean function. More specifically, this could be represented by the following density function:

$$F(a) = Pr\Big\{ \max_{0 \leq t \leq T} \big[x(t) - \mu(t)\big] \leq a \Big\} \tag{5.5}$$

where a is identifies maximum deviations in terms of request volume.

Results are presented in following the graph (Figure 13), which compares both registered (orange) and unregistered (purple) cases. Summary of these estimations is presented at the end of this paper (Appendix B).

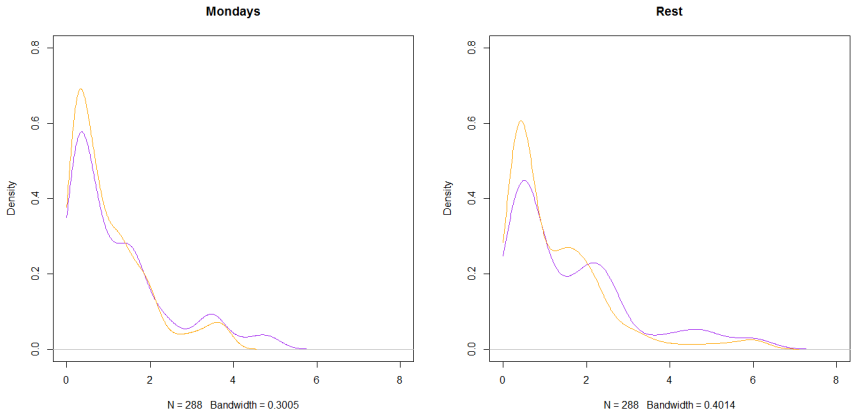


Figure 13: Distribution of maximum values (kernel method)

Hence, the two-step continuous registration leads to lower maximum deviations from the mean function for both clusters. Empirical findings suggest that 75% of these points are not larger than 3.41 and 5.31 for Mondays and rest of the weekdays respectively. Also, maximum values are 4.55 and 7.09, so one should not expect significant departures from mean functions in terms of absolute volumes with relative confidence.

One the other hand, one might also be interested in standardized maximum deviations as well:

$$F(a) = Pr\Big\{ \max_{0 \leq t \leq T} \Big[\frac{x(t) - \mu(t)}{\sigma(t)} \Big] \leq a \Big\} \tag{5.6}$$

which leads to rather different results (Figure 14).

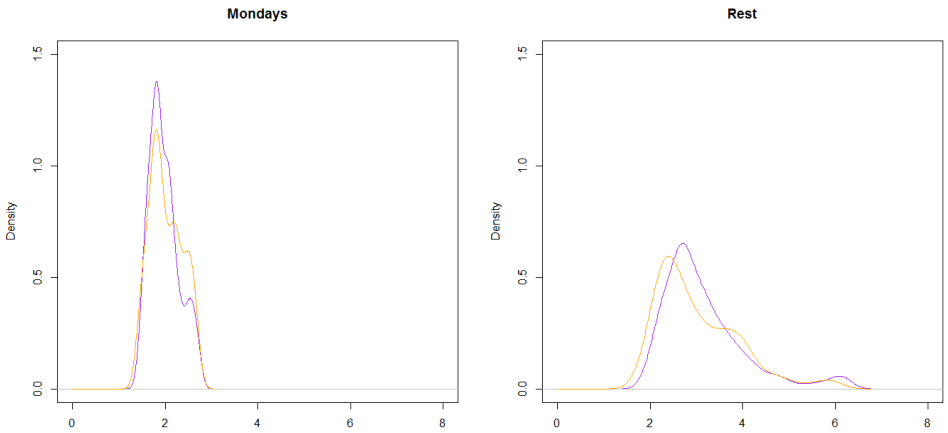


Figure 14: Distribution of standardized maximum values (kernel method)

Standardized maximum deviations from the mean function are similar for both registered and unregistered cases. Also, values are much more concentrated around the mean for Mondays when compared to the rest of weekdays. In general, one might expect registered values to have lower maximum deviations from its corresponding mean functions when compared to unregistered cases. On the other hand, it is rather difficult to make any conclusions about the true mean function for both clusters.

5.6 Density functions for the L^2 norm

One of the main issues with a given sample of functions is to measure their closeness to the corresponding mean function. In other words, two-step continuous registration should lead to a reduction of phase variation, which then allows getting an improved mean estimate. So, there is a possibility to evaluate these functions in the L^2 space:

$$F(a) = Pr\Big\{\Big(\int_0^T [x_j(t) - \mu(t)]^2 dt\Big)^{\frac{1}{2}} \leq a\Big\} \tag{5.7}$$

Again, results are quite similar to both registered (orange) and unregistered (purple) cases (Figure 15).

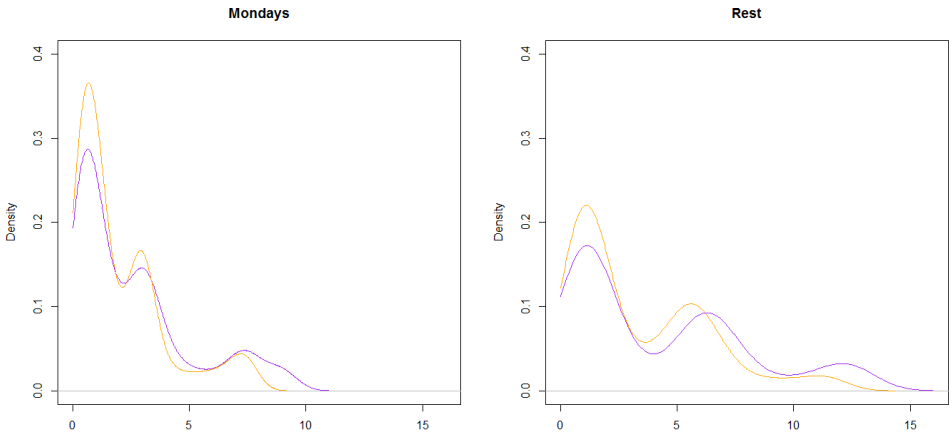


Figure 15: Distribution of values within L^2 space (kernel method)

On the other hand, it is possible to conclude that functions registered using the two step approach are more likely to be included within a specific L^2 space, because it is shifted more to the left compared to the unregistered case. This is applicable for both Mondays and rest of weekdays.

6 Conclusions

Functional data analysis provides various techniques to model datasets that could be described as functional observations. The topic is gaining more popularity due to some new sources of information available worldwide as well as opportunity to model shapes or 3-D objects. However, there situations when functions suffer from phase variation, which does not allow getting a clear view of the underlying process. Hence, different registration methods could be applied in order to solve these issues and get some more appropriate mean function.

This paper investigates a specific case study of daily IT system issues that could be presented as the non-homogeneous Poisson process. A sample consists of 71 functional observations that have been constructed using least-square estimates with a penalty parameter and Fourier basis functions. Later these functions are registered using both landmark and continuous methods. So, it is possible to show that functional registration reduces phase variation for a given sample of observations.

Empirical findings reveal the existence of two functional clusters that could be modelled separately. More specifically, these are Mondays and the rest of weekdays, meaning that different volumes of requests are observed for between those groups. Also, there are some evidence that two-step continuous registration provides even better fit to the data. Therefore, one could use any registration method, detect clusters and then apply some different approach for these separate groups of functions. On the other hand, there are no findings that would indicate any significant changes in pointwise probability density functions.

Finally, it has been revealed that more than 95% of all data points are within 2-σ range, while less than 40% of all functions are fully covered with these boundaries. Hence, it might be more suitable to choose at least 3-σ interval when modelling the actual volume of issues. In addition, registration methods allow reducing maximum deviations from the mean function as well as increases the inclusion of functions in the L^2 space. These results correspond to the empirical literature and provide some new options to model functional datasets.

7 References

[1] F. L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

[2] R. G. Gallager. *Discrete Stochastic Processes*. Springer Science & Business Media, 2012.

[3] T. Gasser and A. Kneip. Searching for structure in curve sample. *Journal of the American Statistical Association*, 90(432):1179–1188, 1995.

[4] D. Gervini and T. Gasser. Self-modelling warping functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):959–971, 2004.

[5] D. Gervini and T. Gasser. Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. *Biometrika*, 92(4):801–820, 2005.

[6] U. Grenander. Stochastic processes and statistical inference. *Arkiv för matematik*, 1(3):195–277, 1950.

[7] L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer Science & Business Media, 2012.

[8] T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, 2015.

[9] G. M. James. Curve alignment by moments. *The Annals of Applied Statistics*, 1(2):480–501, 2007.

[10] A. Kneip and T. Gasser. Statistical tools to analyze data representing a sample of curves. *Annals of Statistics*, 20(3):1266–1305, 1992.

[11] A. Kneip and J. O. Ramsay. Combining registration and fitting for functional models. *Journal of the American Statistical Association*, 103(483):1155–1165, 2008.

[12] A. Laukaitis and A. Račkauskas. Functional data analysis of payment systems. *Nonlinear Analysis: Modeling and Control*, 7(2):53–68, 2002.

[13] H. G. Müller. Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32(2):223–240, 2005.

[14] J. O. Ramsay. When the data are functions. *Psychometrika*, 47(4):379–396, 1982.

[15] J. O. Ramsay and C. J. Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539–572, 1991.

[16] J. O. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and MATLAB*. Springer Science & Business Media, 2009.

[17] J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. New York: Springer, 2002.

[18] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Science & Business Media, 2005.

[19] C. R. Rao. Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17, 1958.

[20] R. Tang and H. G. Müller. Pairwise curve synchronization for functional data. *Biometrika*, 95(4):875–889, 2008.

[21] J. L. Wang, J. M. Chiou, and H. G. Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

[22] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *The Annals of Statistics*, 25(3):1251–1276, 1997.

[23] Z. Zhang, Q. Xie, and A. Srivastava. *Elastic Registration and Shape Analysis of Functional Objects*, pages 218–238. 2015.

A
Empirical estimates of σ boundaries

The bootstrapping procedure has been used to select a sample of functions (200 - Mondays and 700 - Rest), while this cycle is repeated 10 times:

Table 4: σ estimates for Mondays

Sample	1-σ (P)	2-σ (P)	3-σ (P)	1-σ (F)	2-σ (F)	3-σ (F)
Set 1	66.85%	97.17%	100%	0%	40.00%	100%
Set 2	65.80%	96.40%	100%	0%	32.00%	100%
Set 3	67.46%	96.58%	100%	0%	38.50%	100%
Set 4	65.48%	96.23%	100%	0%	30.00%	100%
Set 5	66.58%	96.90%	100%	0%	38.00%	100%
Set 6	67.60%	97.11%	100%	0%	43.00%	100%
Set 7	67.11%	97.18%	100%	0%	41.00%	100%
Set 8	66.91%	96.94%	100%	0%	39.50%	100%
Set 9	67.61%	96.93%	100%	0%	40.00%	100%
Set 10	67.19%	96.63%	100%	0%	36.00%	100%
Average	**66.86%**	**96.81%**	**100%**	**0%**	**37.80%**	**100%**

Results with (P) indicate an overall pointwise average value for each set of observations, while (F) estimate a percentage of functions that have all data points within a given σ range. This has been calculated using the two-step continuous registration method which provides the lowest variance for the available sample of functions.

Table 5: σ estimates for Rest

Sample	1-σ (P)	2-σ (P)	3-σ (P)	1-σ (F)	2-σ (F)	3-σ (F)
Set 1	71.16%	95.90%	99.18%	0%	34.00%	81.71%
Set 2	71.30%	95.81%	99.09%	0%	35.14%	80.29%
Set 3	71.31%	95.80%	99.10%	0%	32.14%	79.57%
Set 4	70.44%	95.35%	98.89%	0%	29.57%	76.00%
Set 5	72.12%	95.81%	99.09%	0%	35.00%	79.71%
Set 6	72.28%	96.02%	99.19%	0%	34.71%	82.57%
Set 7	71.21%	95.87%	99.14%	0%	36.43%	79.86%
Set 8	71.93%	95.96%	99.21%	0%	34.57%	82.00%
Set 9	72.07%	95.96%	99.14%	0%	34.00%	81.00%
Set 10	71.25%	95.74%	98.98%	0%	34.29%	78.00%
Average	**71.51%**	**95.82%**	**99.10%**	**0%**	**33.99%**	**80.07%**

B
Empirical estimates of maximum deviations

The following tables present the distribution of maximum deviations away from the mean function in which the largest pointwise differences are calculated in absolute terms. A short summary of the main quartiles from the corresponding density estimates are provided below:

Table 6: Quartiles of maximum deviations for Mondays

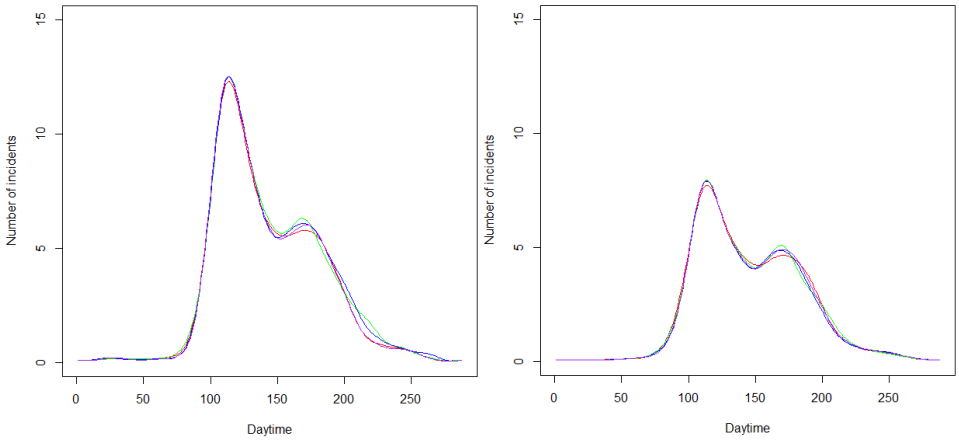
Quantile	Unregistered	Registered
Min.	0	0
1st Qu.	1.46	1.14
Median	2.92	2.28
Mean	2.92	2.28
3rd Qu.	4.38	3.41
Max.	5.85	4.55

Table 7: Quartiles of maximum deviations for Rest

Quantile	Unregistered	Registered
Min.	0	0
1st Qu.	1.86	1.77
Median	3.72	3.54
Mean	3.72	3.54
3rd Qu.	5.58	5.31
Max.	7.43	7.09

C
Graphs with descriptive statistics

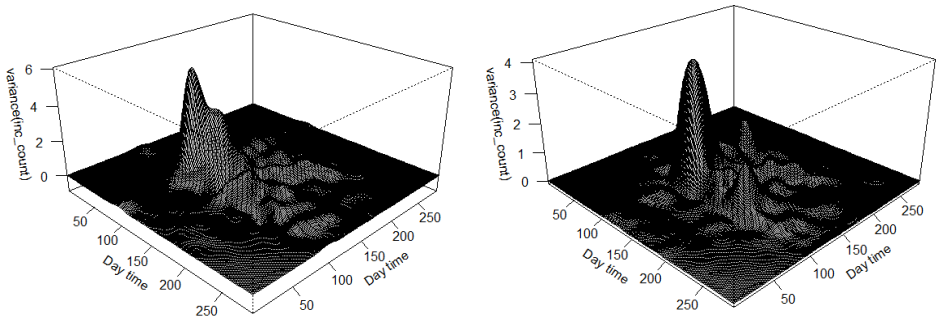
These graphs indicate estimated mean functions for unregistered (red) as well as registered using landmark (green), continuous (blue) and two-step continuous (purple) methods.



(a) Mondays (b) Rest

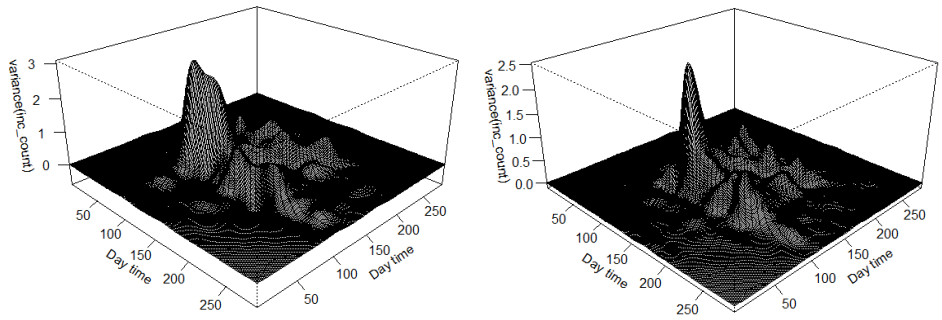
Figure 16: Estimated mean functions using different methods

Variance-covariance surfaces compare unregistered observations with those registered cases using the two-step continuous approach:



(a) Unregistered (b) Two-step registration

Figure 17: Comparison of variance-covariance surfaces for Mondays



(a) Unregistered (b) Two-step registration

Figure 18: Comparison of variance-covariance surfaces for Rest

D
Non-homogeneous Poisson process

This section is based on a book chapter in Gallager (2012) [2], where author provides a theoretical background for the non-homogeneous Poisson process.

First of all, the Poisson process is usually characterized using some constant arrival measure λ. However, in some situations it is more reasonable to loosen this assumption and consider a flexible arrival rate that changes as a function of time. So, it is defined as the *non-homogeneous Poisson process* with a measure $\lambda(t)$, where counting process $\{N(t); t > 0\}$ could be characterized by independent increment property, when for all $t \leq 0, \delta > 0$, it satisfies the following equations:

$$\begin{aligned} Pr\Big\{\tilde{N}(t, t+\delta) = 0\Big\} &= 1 - \delta\lambda(t) + o(\delta) \\ Pr\Big\{\tilde{N}(t, t+\delta) = 1\Big\} &= \delta\lambda(t) + o(\delta) \\ Pr\Big\{\tilde{N}(t, t+\delta) \geq 2\Big\} &= o(\delta) \end{aligned} \tag{D.1}$$

where $\tilde{N}(t, t+\delta) = N(t+\delta) - N(t)$. Hence, the property of stationary increments is not applicable for the non-homogeneous Poisson process.

Also, one could use "shrinking Bernoulli process" as an approximation, but first $\lambda(t)$ must be bounded away from zero. Later, domain values are divided into separate increments with length δ, which varies inversely with $\lambda(t)$, in order to maintain the probability of the arrival $p = \delta\lambda(t)$ fixed for any interval value. Hence, once ignoring variation in $\lambda(t)$ for the moment, it leads to:

$$\begin{aligned} Pr\Big\{\tilde{N}\Big(t, t+\frac{p}{\lambda(t)}\Big) = 0\Big\} &= 1 - p + o(p) \\ Pr\Big\{\tilde{N}\Big(t, t+\frac{p}{\lambda(t)}\Big) = 1\Big\} &= p + o(p) \\ Pr\Big\{\tilde{N}\Big(t, t+\frac{p}{\lambda(t)}\Big) \geq 2\Big\} &= o(\epsilon) \end{aligned} \tag{D.2}$$

It might be easier to define $m(t)$ as:

$$m(t) = \int_0^t \lambda(\tau) d\tau \tag{D.3}$$

Let $\{Y_i; i \geq 1\}$ be some sequence of independent and identically distributed (i.i.d) random variables with $Pr\{Y_i = 1\} = p$ and $Pr\{Y_i = 0\} = 1 - p$. Then, consider a counting process $\{\mathrm{N(t);\ t} > 0\}$, where Y_i for each $i \geq 1$, indicates a volume of arrivals during the interval $(t_{i-1}, t_i]$ so that t_i satisfies $m(t_i) = ip$. Hence, $N(t_i) = Y_1 + Y_2 + ... + Y_i$. Given that p decreases as 2^{-j}, each interval could be divided into the pair of increments:

$$Pr\Big\{N(t) = n\Big\} = \frac{[1 + o(p)][m(t)]^n \exp[-m(t)]}{n!} \tag{D.4}$$

Also, for any interval $(t, \tau]$ given that $\tilde{m}(t, \tau) = \int_t^T \lambda(u)du$ and $t = t_k$, $\tau = t_i$ for a set of k, i, we find:

$$Pr\Big\{\tilde{N}(t, \tau) = n\Big\} = \frac{[1 + o(p)][\tilde{m}(t, \tau)]^n \exp[-\tilde{m}(t, \tau)]}{n!} \tag{D.5}$$

As $p \to 0$ in the limit, the counting process $N(t); t > 0$ approaches some non-homogeneous Poisson process, providing the following theorem:

Theorem D.1 *For a non-homogeneous Poisson process with some right continuous arrival measure $\lambda(t)$ bounded away from zero, the distribution of $\tilde{N}(t, \tau)$, the number of arrivals in $(t, \tau]$, satisfies:*

$$Pr\Big\{\tilde{N}(t, \tau) = n\Big\} = \frac{\tilde{m}(t, \tau)]^n \exp[-\tilde{m}(t, \tau)]}{n!} \tag{D.6}$$

where $\tilde{m}(t, \tau) = \int_t^T \lambda(u)du$.

In conclusion, given a non-linear time scale, one could see a homogeneous Poisson process as its non-homogenous counterpart. For instance, if there is a (homogeneous) Poisson process $\{N^*(s); s \geq 0\}$ with rate 1, then non-homogeneous Poisson process could be represented as $N(t) = N^*(m(t))$ for each t.

E

Graphs with time-warping functions

The following graphs present three sample curves and their corresponding time-warping functions using the continuous registration method. Given that there are 71 observations available, only curves number 16, 33 and 49 have been chosen for illustration.

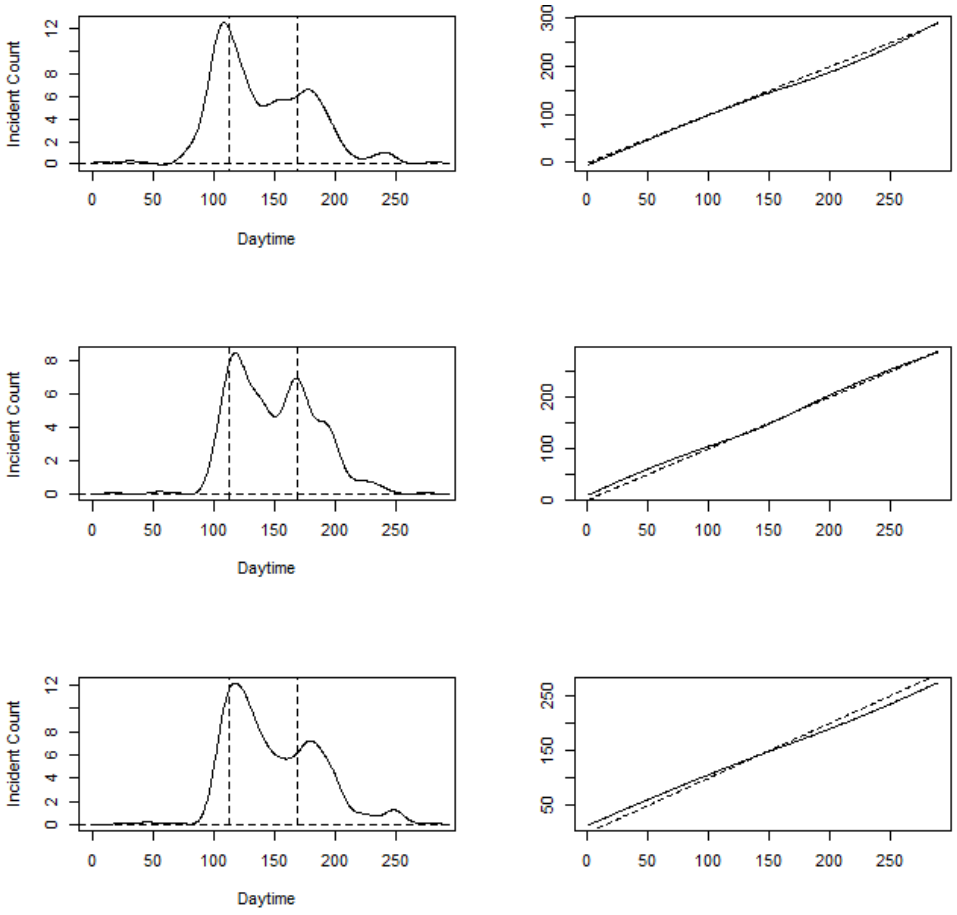


Figure 19: Functions before registration (left) and their corresponding time-warping functions (right) using the continuous registration method