Global analysis of the COVID-19 outbreak profiles*

Tadas Danielius, Jovita Gudan, Alfredas Račkauskas
Vilnius university

Abstract

In this paper functional data analysis techniques are used to study worldwide COVID-19 outbreak profiles. Indicators under investigation include quarantined cases $Q(t)$, mortality rate $\kappa(t)$ and cure rate $\lambda(t)$. All functions are interpreted as random processes with paths in the space $L_2([0,1])$ by choosing the interval $[0,1]$ for different time periods. Exploratory analysis includes profile parameters such as mean, variance, depth, derivatives, principal components. Comparison of average profiles of different parts across the world is provided by means of a simple trend model. Functional comparison of two COVID-19 outbreak waves is suggested.

1 Introduction

The first suspected case of coronavirus disease-2019 (COVID-19) was fixed on December 1st, 2019, in Wuhan, the capital of Hubei Province in China. By March 2020, the World Health Organization declared the situation as pandemic. Since that time confirmed cases and deaths have been reported almost at every location over the world, many countries and regions have been locked-down and applied strict social distancing measures to stop the virus propagation. Understanding the propagation patterns of the disease and prediction of its spread over time is of great importance and has attracted a large scientific community (see, e.g., [11] and references therein).

Many classical models such as SIR [[14], [15]], SEIR [17], SEIJR [19] and their modifications have been applied to describe the outbreak of SARS-CoV-2 in various countries (we refer to [13] for detailed exposition of the state-of-art). These models include various indicators of disease: the number of the susceptible cases $S(t)$, exposed cases (infected but not yet be infectious, in a latent period) $E(t)$, infectious cases (with infectious capacity and not yet be quarantined) $I(t)$, recovered cases $R(t)$, insusceptible cases $P(t)$, quarantined cases (confirmed and infected) $Q(t)$, closed cases (or death) $D(t)$ at time t, and

---

*The research supported by the Research Council of Lithuania, grant Nr. S-COV-20-4

are based on the average behavior of a population under study. Actually all these indicators are random functions in time depending as well on location of coronavirus outspread. Corresponding data sets usually contain daily observations. We aim to study variation among the observed data and to compare sets of data corresponding to different parts of the world with respect to certain types of variation. To this aim we use functional data analysis methodology which have been applied to various areas including human growth (see, Chen and Müller [6]), gene expression (see, Tang and Müller [7]) geomagnetic activity (see, Gromenko and Kokoszka [8]), medicine (see, Bircan et all. [9], psychology (see, Levitin et all., [10]) etc.

Based on the publicly available data [20], we provide analysis of the epidemiological indicators $Q(t), R(t), D(t)$ and particularly we are interested in derived parameters: the mortality rate $\kappa(t)$ and cure rate $\lambda(t)$ defined by the relations

$$D'(t) = \kappa(t)Q(t) \quad \text{and} \quad R'(t) = \lambda(t)Q(t), \tag{1}$$

where $D'(t)$ and $R'(t)$ denotes derivatives (or intensity) of $D(t)$ and $R(t)$ respectively. The questions we are seeking to answer include the following.

- What is the typical mortality rate (respectively cure rate) pattern for European countries? For Baltic region?
- What countries are most similar to Lithuania with respect to different indicators, e.g., mortality and cure rates?
- In what period of pandemics there was most variable mortality and cure across Europe?
- In what period of pandemics there was most similar mortality and cure across Europe?
- Do some and which countries tend to have extreme mortality (respectively cure) rate? How long periods of extreme rates last?
- How do the shapes of the mortality and cure rates patterns differ among the East and West parts of Europe? among Europe and rest of the world?
- How similar successive epidemic waves are compared with previous waves?

The paper is organized as follows. Section 2 is devoted to preparation of functional sample to be analysed. Exploratory analysis of the functional sample is given in Section 3. Section 4 contains analysis of epidemic waves by comparing grows rates. Finally, in Section 5 we compare different regions of the world with respect to behaviour of epidemiological indicators.

2 Data preparation

First we discuss preparation for analysis of curves $Q(t), D(t)$ and $R(t)$ for $t \in \mathcal{T}$ corresponding to different locations, where the time interval $\mathcal{T}$ depends on a question under investigation. For example $\mathcal{T}$ can be a length of the first wave, or the first d days after some event etc.

In general we have a set of daily observations $x_{ji}, i = 1, \dots, m_j$, $j = 1, \dots, n$, corresponding to either of parameter under investigation, where $j \in J$ indicates location and J is a selected collection of locations. In this paper we investigate five separate parts of the world: Europe, Asia, Americas, Africa, Oceania. From this set of observations we build functional sample $x_j = (x_j(t), t \in [0,1]), j \in J$, by setting a model

$$x_j(t) = 0 \ \text{ if } \ 0 \le t \le 1 - \frac{m_j}{T}, \tag{2}$$

$$x_{j,i} = x_j(1 - (m_j - i)/T) + \varepsilon_{ji}, \quad i = 1, \dots, m_j, \tag{3}$$

where T is a total number of days taken into analysis and $\varepsilon_{jk}, k = 1, \dots, m$ are some errors. Sometimes it is convenient to transform the functions defined on the interval $[0,1]$ to the functions defined on the interval $[T_0, T_1]$. This can be done by a simple time transformation $t \to (t - T_0)/(T_1 - T_0)$.

In order to reconstruct each curve x_j as smooth monotone function we represent, as suggested in [4],

$$x_j(t) = \int_{\tau_j}^{t} \exp\{w_j(u)\} du, \quad t \in [\tau_j, 1],$$

where $\tau_j = 1 - m_j/T$. For the function w_j we use basis expansion

$$w_j(t) = \sum_{j=0}^{K} w_{jk}\psi_k(t), \quad t \in [0,1],$$

where $\psi_k, k \ge 1$ are basis functions and we estimate the regression

$$x_{ji} = \int_{\tau_j}^{t} \exp\Big\{ \sum_{k=1}^{K} w_{jk}\psi_k(u) \Big\} du + \varepsilon_{ji}$$

by minimizing the criterion

$$\sum_{i=1}^{m_j} \Big[x_{ji} - \int_{\tau_j}^{t} \exp\Big\{ \sum_{k=1}^{K} w_{jk}\psi_k(u) \Big\} du \Big]^2 + \lambda \int_0^1 w_j^2(u) du,$$

with respect to $w_{j1}, \dots, w_{jK}$.

Example of smoothed observations of the random process $Q(t)$ for four countries are depicted in Figure 1.

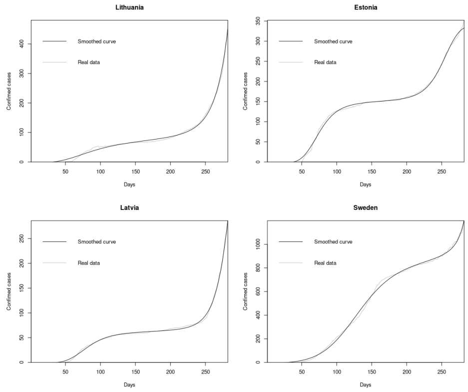


Figure 1: Discrete vs. smoothed quarantined observations in four Europe countries

The use of derivatives of functional sample is important in the development of methodology of statistical inference and in extending the range of graphical exploratory methods. The derivative curves of each of processes $Q(t), D(t)$ and $R(t)$ provide extra information of disease status and can be interpreted as intensities of quarantined, closed and recovered cases respectively. Derivatives of $Q(t)$ for the four countries are depicted in Figure 2.

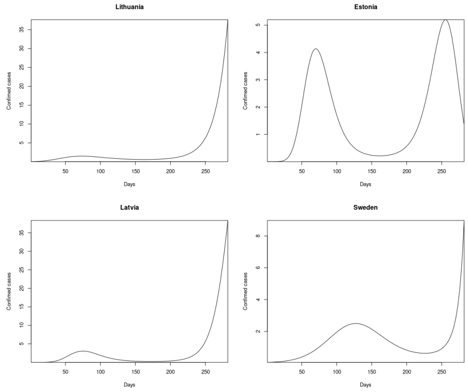


Figure 2: Derivatives of smoothed quarantined cases in four Europe countries

In the following we shall denote by $\widehat{Q}_j, j \in J$, $\widehat{D_j}, j \in J$, and $\widehat{R}_j, j \in J$, functional sample of quarantined, closed and recovered cases respectively. The corresponding derivative samples are $\widehat{Q}'_j, j \in J$, $\widehat{D_j}', j \in J$ and $\widehat{R}'_j, j \in J$. Each of the sample functions are defined on the interval $[0,1]$. The ratio

$$\widehat{\rho}_j(t) := \frac{\widehat{Q}'_j(t)}{\widehat{Q}_j(t)}, \quad t \in [0,1],$$

(throughout we assume $0/0 = 0$) can be used to understand a profile of the reproduction number. Indeed, as the chance of infection increases with the number of infected individuals, assume that the variations in the population of the number of infected $I(t)$ between time t and $t+\tau$ (over relatively small intervals τ) is proportional to the number of infected individuals, i.e.,

$$\frac{dI(t)}{dt} \approx \frac{I(t+\tau) - I(t)}{\tau} = \rho(t) I(t)$$

The function $\rho(t)$ models how the infected population evolves over time. If we assume that $I(t) = \lambda Q(t)$ at least in a small time interval then we have

$$\rho(t) \approx \frac{Q'(t)}{Q(t)} = \frac{I'(t)}{I(t)}.$$

Any functional data display two types of variability: a phase variability and an amplitude variability. In order to make correct comparisons among the features, observed in different countries, we need to separate these two types of variability and look for a new parameterization of each of the countries COVID-19 indicator. This can be achieved by means of a registration procedure that finds warping functions h_j of the abscissa, leading to the new registered functions $\widetilde{x}_j$:

$$\widetilde{x}_j = x_j \circ h_j^{-1} \quad \text{or} \quad \widetilde{x}_j \circ h_j = x_j.$$

Note that the registered features $x_j(s)$ of the curve x_j are obtained by moving the observed features to their new location $h_i(s)$. For example, concerning Q_j such a feature could be the first confirmed case. For Q_j' it could be the extreme points of the curve.

3 Exploratory analysis

Let $x_1, x_2, \ldots, x_n$ be a functional sample each function being defined on the interval $[0,1]$. Exploratory data analysis is an approach to analyzing data sets by summarizing their main characteristics, often with visual methods. Typical graphical techniques include functional sample mean

$$\overline{x}_n(t) := \frac{1}{n} \sum_{k=1}^{n} x_k(t), t \in [0,1];$$

sample variance

$$\sigma_n^2(t) = \frac{1}{n-1} \sum_{k=1}^{n} (x_k(t) - \overline{x}_n(t))^2, \quad t \in [0,1];$$

sample standard deviation function $\sqrt{\sigma_n^2}$, sample covariance function

$$c_n(t,s) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i(t) - \overline{x}_n(t))(x_i(s) - \overline{x}_n(s)), \quad s,t \in [0,1];$$

estimated functional principal components are functions $v_j(t)$ such that the centred functions $x_i(t) - \overline{x}_n(t)$ are represented as

$$x_i(t) - \overline{x}_n(t) \approx \sum_{j=1}^{p} z_{ij} v_j(t), \quad t \in [0,1],$$

with p much smaller than the number of basis functions used to in reconstruction of $x_i(t)$. The coefficients z_{ij} are scores corresponding to principal components.

The functional depth is a concept which measures how deep (or central) is a datum respect to a population. The aim of data depth for functional data is to measure the centrality of a given curve, x_i, within a set of curves, $x_1, \ldots, x_n$. In functional context this provides a method to order sample curves according to decreasing depth values: $x_{[1]}$: the deepest (most central or median) curve, $x_{[n]}$ the most outlying (least representative) curve. The Fraiman-Muniz functional depth, of a curve x_i with respect the set $x_1, \ldots, x_n$ is given by

$$FMD_n(x_i) = \int_0^1 D_n(x_i(t))dt,$$

where $D_n(x_i(t))$ denotes the univariate depth of the data $x_i(t)$ in the sample $x_1(t), \ldots, x_n(t)$.

Outliers in a functional data set can arise for, at least, two reasons: (1) outliers may be curves with gross errors such as measurement, recording, and typing mistakes. These errors should be identified and corrected whenever possible; (2) outliers may be real data curves in the sense that they are not gross errors but are somehow suspicious or surprising as they do not follow the same pattern as that of the rest of curves. We are interested in detecting and examining precisely such surprising curves as, first, they may bias functional estimates and we would like to prevent this, and, second, they may allow to discover which sources produce these outlying curves.

In order to identify outliers in functional data sets, we make use of functional depths. If an outlier is in the data set, the corresponding curve will have a significant low depth. Therefore, a way to detect the presence of functional outliers is to look for curves with lower depths. Consequently, we have the following functional outlier detection procedure for detecting outliers in a given dataset of functional curves $x_1, \ldots, x_n$:

(1) Obtain the functional depths $D_n(x_1), \ldots, D_n(x_n)$.

(2) Let $x_{i_1}, \ldots, x_{i_k}$ be the k curves such that $D_n(x_{i_k}) \leq C$, for a given cutoff C. Then, we assume that $x_{i_1}, \ldots, x_{i_k}$ are outliers and delete them from the sample.

(3) Then, come back to step (1) with the new dataset after deleting the outliers found in step (2). Repeat this until no more outliers are found.

We select C such that, in the absence of outliers

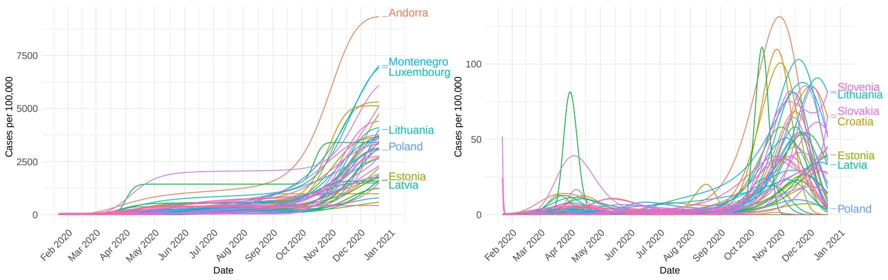
$$P(D_n(x_i) \leq C) = 0.01, i = 1, \ldots, n.$$

Thus, the C taken is the 1'th percentile of the distribution of the functional depth under consideration.

The rest of Section 3 explains the situation of COVID-19 outbreak only in the countries that are in Europe region. Equivalent analysis is carried out for the rest of the regions, but is not described in the further sections.

3.1 Quarantined cases

After the SARS-CoV-2 virus was first discovered in China, it spread rapidly to other regions. First in Europe and then in America and elsewhere. Figure 3 illustrates smoothed curves of quarantined cases $Q_j(t)$ as well as their derivatives in Europe region for the period January 22, 2020 - December 20, 2020. The figures highlight the Baltic region, Poland and the top 3 countries that had the most cases on December 20, 2020. Almost every affected country share same pattern of the epidemic spread which is S-shaped curve for the confirmed cases. The highest total quarantined cases are in Andorra, Montenegro and Luxembourg, followed by Lithuania, Poland, Estonia and Latvia. Number of quarantined cases discovered daily are the highest in Slovenia, Lithuania and Croatia in December 20, 2020.



(a) Smoothed $Q_j(t)$ (b) Derivatives of $Q_j(t)$

Figure 3: Quarantined cases in Europe

Figure 4 represents estimated functional sample mean, 95% confidence interval, the Fraiman-Muniz functional depth and outliers for Europe region. In Europe, one country distinguishes from the situation of pandemic, which is Andorra. Since quarantined cases are observed as a indicator per $100,000$ inhabitants, later country is characterized by a marginal population size and a high rate of infection is observed. Such countries should be analyzed separately due to the latter reasons. The most customary situation according the depth analysis is in Austria.

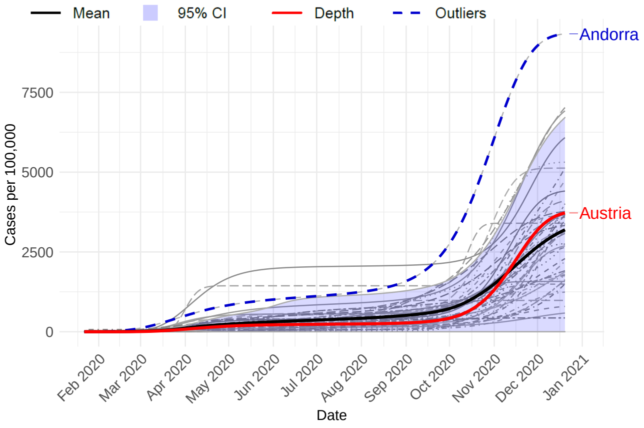
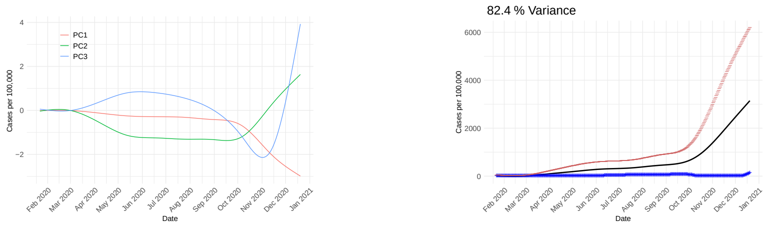
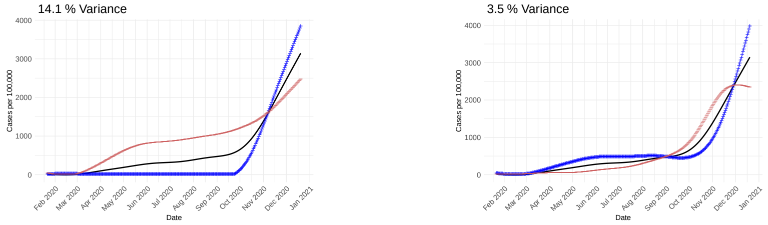


Figure 4: Exploratory analysis of quarantined cases for Europe region

To understand the primary modes of variation in the COVID-19 data, functional principal component analysis is performed. Three components explain almost all variability across the European countries. The first functional component explains 82.4% of variation, seconds – 14.1% and third – 3.5%. Figure 5a displays the effect of the first three principal components to the situation of pandemic of COVID-19. In each case of Figures 5b-5d, the solid curve is the overall mean quarantined cases function, and the dotted and dashed curves show the effects of adding and subtracting a multiple of each principal component curve. This considerably clarifies the effects of the first three components. The first principal component corresponds to a overall situation in Europe, where the second waves of Covid-19 emerge and the first fades before the latter. The second corresponds to an effect of the start of the first and second wave of Covid-19. Whereas the third corresponds to the magnitude of the first, second and even third waves for some countries of Covid-19.



(a) Functional PCs (b) 1*st* functional PC



(c) 2*st* functional PC (d) 3*st* functional PC

Figure 5: The first three principal components of quarantined cases and the mean curves with the effects of adding (+) and subtracting (−) a suitable multiple of each PC curve.

It can be profitable to plot the principal component scores for pairs of harmonics to see how curves cluster and otherwise distribute themselves within the 3-dimensional subspace spanned by the eigenfunctions. Figure 6 reveals some fascinating structure. Two Baltic countries Estonia and Latvia behaves very similarly as some Nordic countries and Germany, while Lithuania is further apart.

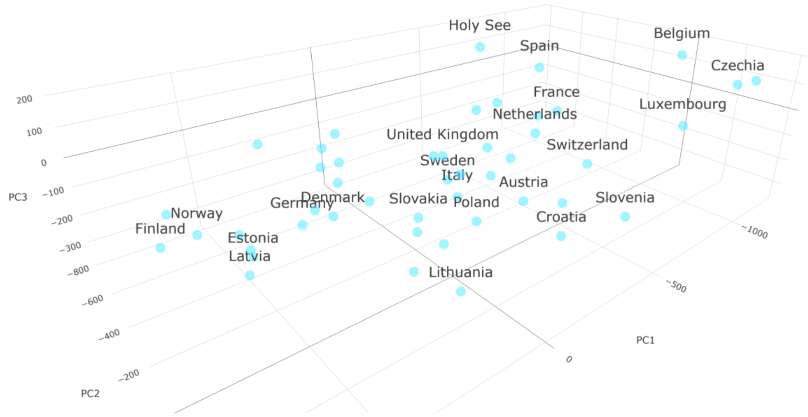
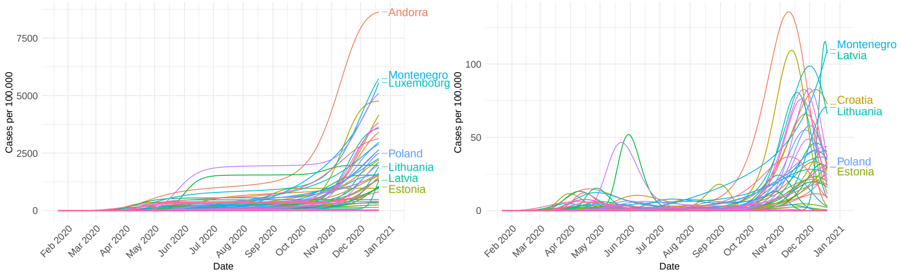


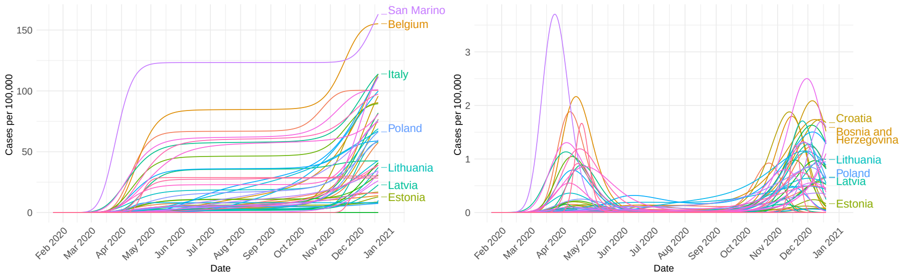
Figure 6: The scores of the Europe countries on the first three principal components of quarantined cases.

3.2 Recovered and closed cases

Equivalent analysis is carried out for the recovered $R(t)$ and closed $D(t)$ cases. Figure 8 illustrates smoothed curves of recovered and closed cases as well as their derivatives in Europe region for the period January 22, 2020 - December 20, 2020. The figures highlight the Baltic region, Poland and the top 3 countries that had the most cases on December 20, 2020. The highest total recovered cases are in Andorra, Montenegro and Luxembourg, followed by Poland, Lithuania, Latvia and Estonia and for the closed cases - San Marino, Belgium, Italy in December 20, 2020. Number of recovered cases discovered daily are the highest in Montenegro, Latvia and Croatia, whilst for closed cases - Croatia, Bosnia and Herzegovina and Lithuania in December 20, 2020.



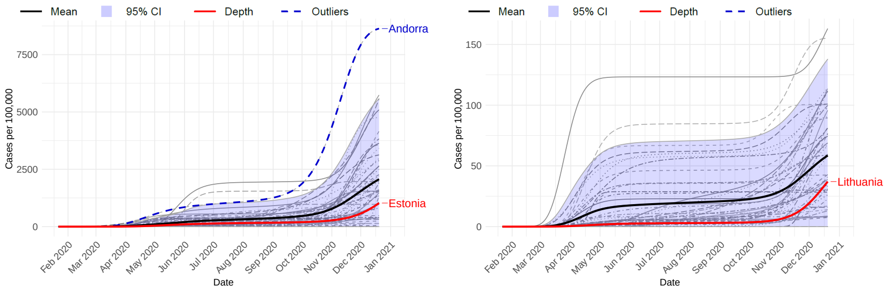
(a) Smoothed $R_j(t)$ (b) Derivatives of $R_j(t)$



(c) Smoothed $D_j(t)$ (d) Derivatives of $D_j(t)$

Figure 7: Recovered $R_j(t)$ and closed $D_j(t)$ cases in Europe

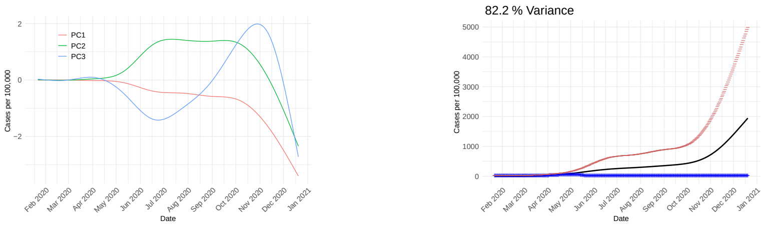
Figure 8 represents estimated functional sample mean, 95% confidence interval, the Fraiman-Muniz functional depth and outliers for both recovered and closed cases in Europe region. In Europe, one outlier is dicovered for recovered cases, which is Andorra. The most typical situation according the depth analysis is in Estonia for the recovered cases and in Lithuania for the closed cases.



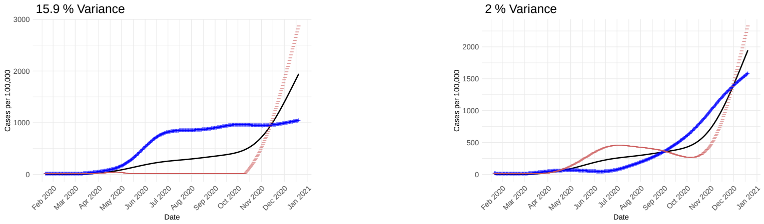
(a) Recovered cases (b) Closed cases

Figure 8: Exploratory analysis of recovered and closed cases for Europe region

The analysis of functional principal component analysis reveals that the first functional component explains 82.2% of variation, seconds – 15.9% and third – 1.9% for recovered cases in Europe. Figure 9 displays the effect of the first three principal components to the situation of pandemic of COVID-19 for recovered cases. In each case of Figures 9b-9d, the solid curve is the overall mean recovered cases function, and the dotted and dashed curves show the effects of adding and subtracting a multiple of each principal component curve. The first principal component corresponds to a overall situation in Europe, where the second waves of Covid-19 emerge and the first fades before the latter. The second corresponds to an effect of the start of the first and second wave of Covid-19 for recovered cases. Whereas the third corresponds to the magnitude of the first, second and even third waves for some countries of Covid-19.

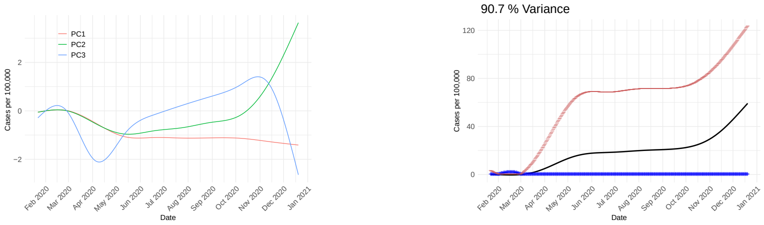


(a) Functional PCs (b) $1st$ functional PC

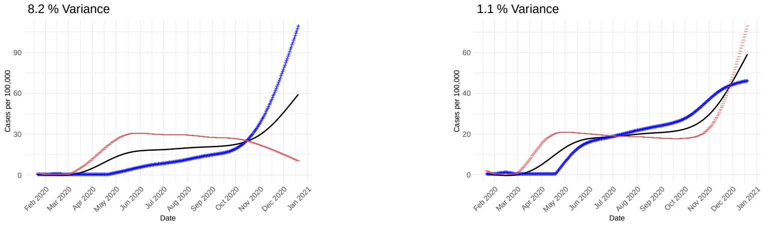


(c) $2nd$ functional PC (d) $3rd$ functional PC

Figure 9: The first three principal components of recovered cases and the mean curves and the effects of adding (+) and subtracting (–) a suitable multiple of each PC curve.



(a) $1st$ functional PC (b) $1st$ functional PC



(c) $2nd$ functional PC (d) $3rd$ functional PC

Figure 10: The first three principal components of closed cases and the mean curves and the effects of adding (+) and subtracting (–) a suitable multiple of each PC curve.

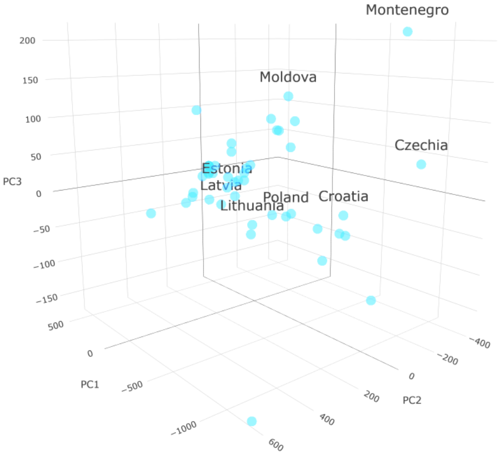


Figure 11: The scores of the Europe countries on the first three principal components of recovered cases.

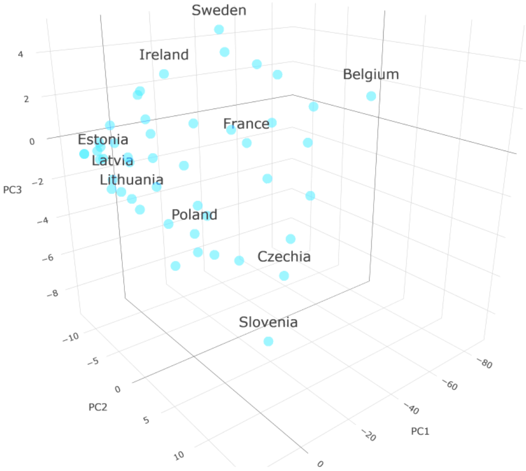


Figure 12: The scores of the Europe countries on the first three principal components of closed cases.

3.3 Mortality and cure rates

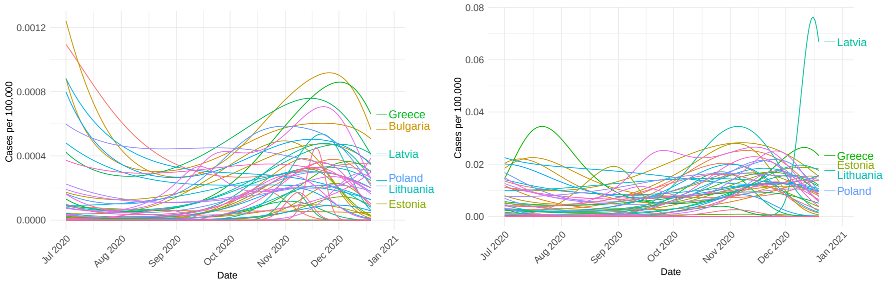
Mortality, respectively cure, rate at time t corresponding to country j is estimated by

$$\widehat{\kappa}_j(t) = \frac{\widehat{D}'_j(t)}{\widehat{Q}_j(t)}, \quad t \in [0,1]$$

respectively

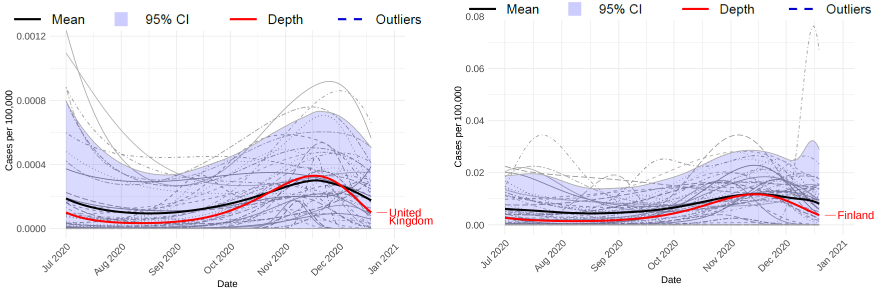
$$\widehat{\lambda_j}(t) = \frac{\widehat{R}'_j(t)}{\widehat{Q}_j(t)}, \quad t \in [0, 1],$$

assuming $0/0 = 0$.



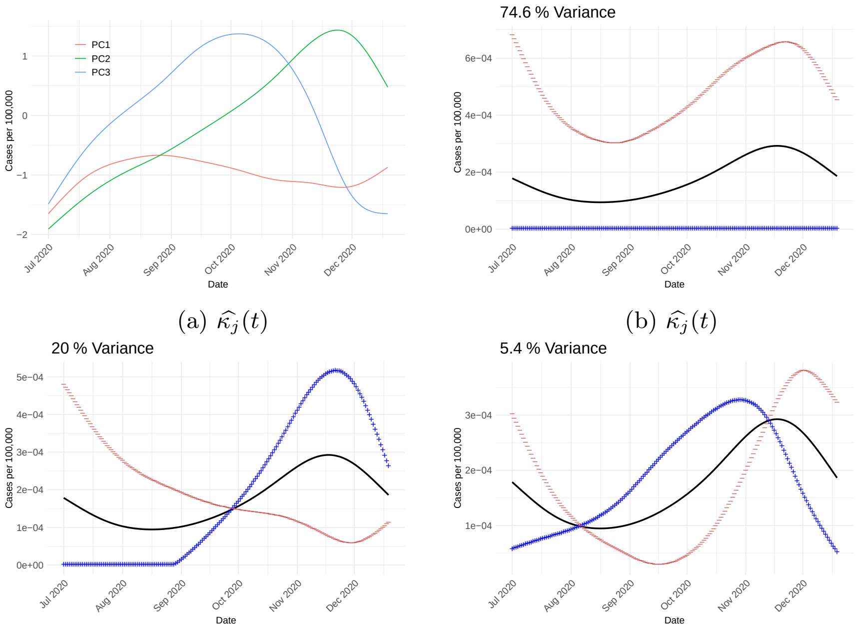
(a) $\widehat{\kappa}_j(t)$ (b) $\widehat{\lambda_j}(t)$

Figure 13: Mortality and cure rates in Europe



(a) $\widehat{\kappa}_j(t)$ (b) $\widehat{\lambda_j}(t)$

Figure 14: Exploratory analysis of mortality and cure rates in Europe



(a) $\widehat{\kappa}_j(t)$ (b) $\widehat{\kappa}_j(t)$

(c) $\widehat{\kappa}_j(t)$ (d) $\widehat{\kappa}_j(t)$

Figure 15: fPCA $\widehat{\kappa}_j(t)$

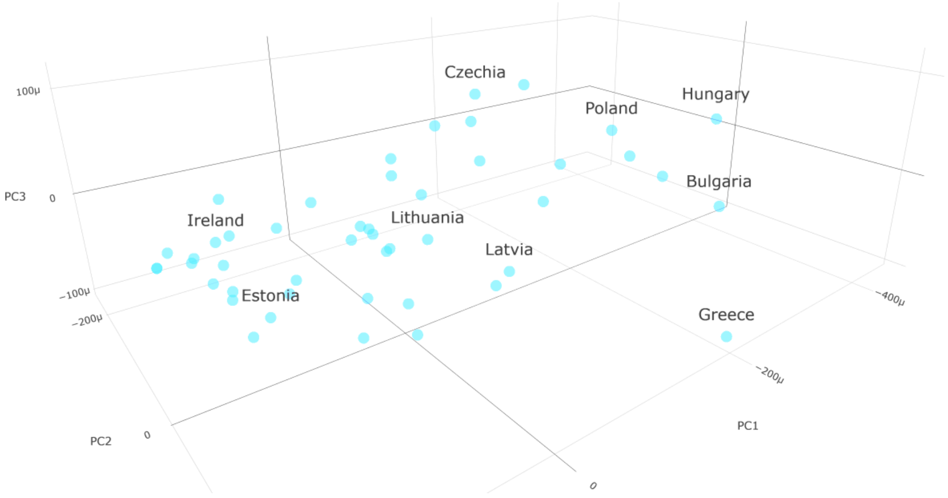
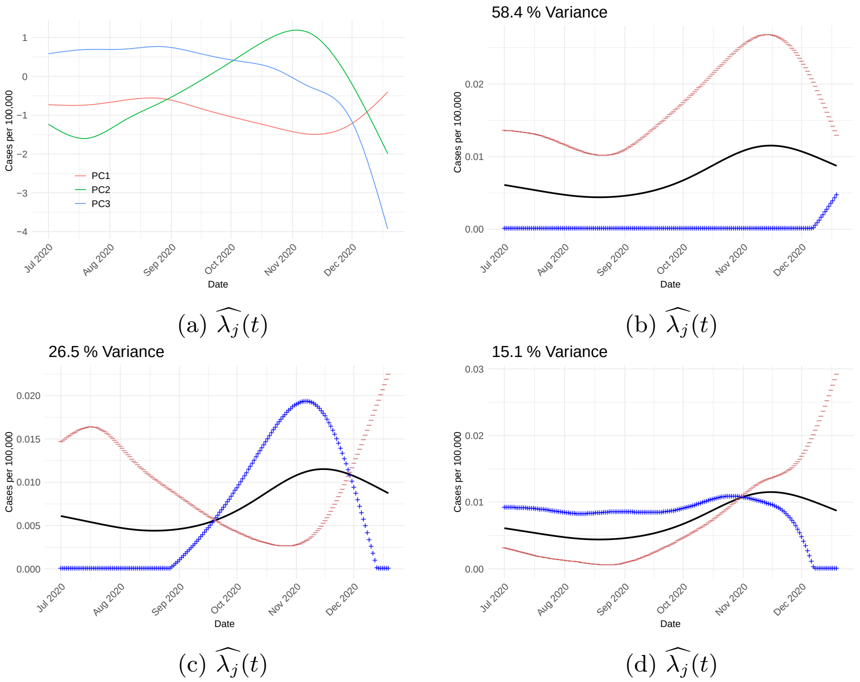


Figure 16: Scores $\widehat{\kappa}_j(t)$



(a) $\widehat{\lambda_j}(t)$ (b) $\widehat{\lambda_j}(t)$

(c) $\widehat{\lambda_j}(t)$ (d) $\widehat{\lambda_j}(t)$

Figure 17: fPCA $\widehat{\lambda_j}(t)$

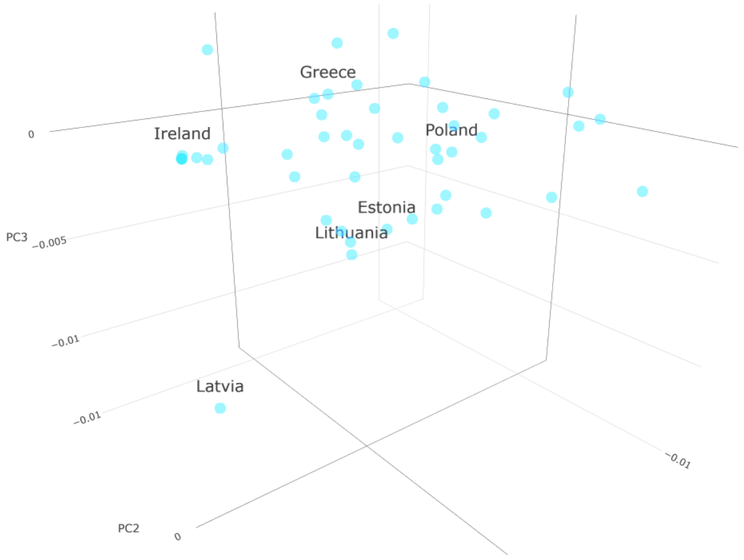


Figure 18: fPCA $\widehat{\lambda_j}(t)$

4 Epidemic waves

Propagated epidemic curves usually have a series of successively larger peaks. The successive waves tend to involve more and more people, until the pool of susceptible people is exhausted or control measures are implemented. The 2019–20 coronavirus (COVID-19) pandemic shows classical pattern of propagated epidemic with more than one wave. Arguably, slowdown in the first wave was achieved because of the strict control measures which were implemented worldwide. However, once the restrictions were relaxed and countries opened their economies, the spread of the virus started to accelerate again, thus we experienced the second wave.

In this section we explore the epidemic waves and how functional data can help to localize each wave using first and second order derivatives. Then, we propose a similarity metric of two waves. This metric can be used for clustering and forecasting. The inclusion of the similarity into the forecasting models can improve accuracy because it carries the behavioral patterns of the population.

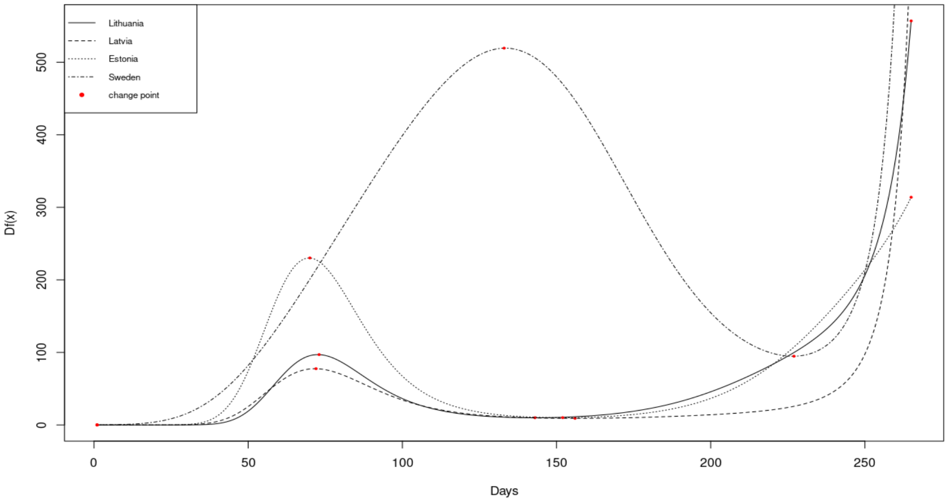


Figure 19: Derivatives of the confirmed cases in four Baltic countries with marked change points in epidemic wave acceleration.

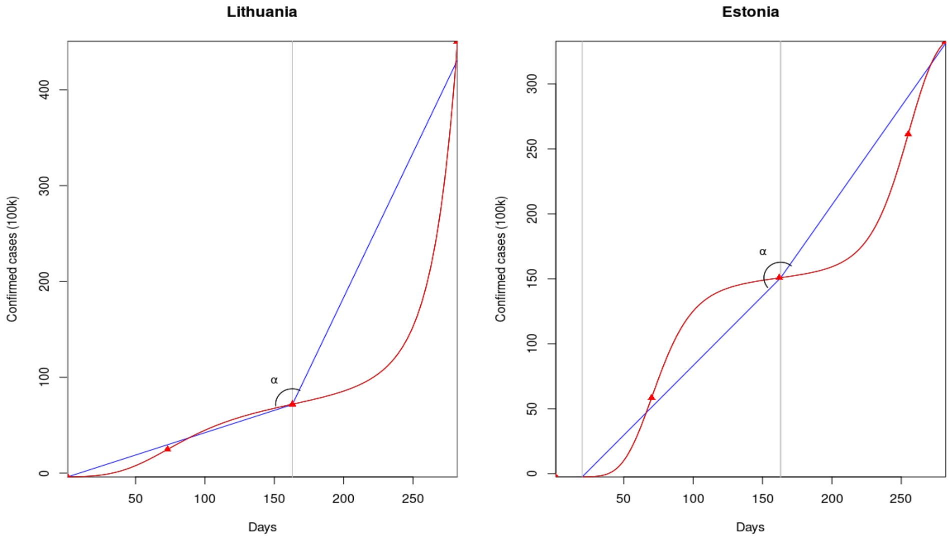


Figure 20: COVID-19 cases confirmed in Lithuania and Estonia with segmented epidemic waves.

There is no strict definition for an epidemic wave. Normally, a wave implies a rising number of affected individuals, following by a peak of new cases and the decline. The decline does not mean that outbreak is over. The future outbreaks of disease are possible. To better understand and compare the epidemic waves first we need to localize the beginning and the end of the wave. We will take advantage of the information in derivatives which is probably the most distinctive feature of functional data. The figure 19 shows the derivatives of the confirmed cases in the Baltic region. Having data in this form helps to find epidemic wave peaks. The peak points are considered if $Q'(t^-) < Q'(t) > Q'(t^+)$ and epidemic wave end point is considered if $Q'(t^-) > Q'(t) < Q'(t^+)$ which is also a start point for the successive wave.

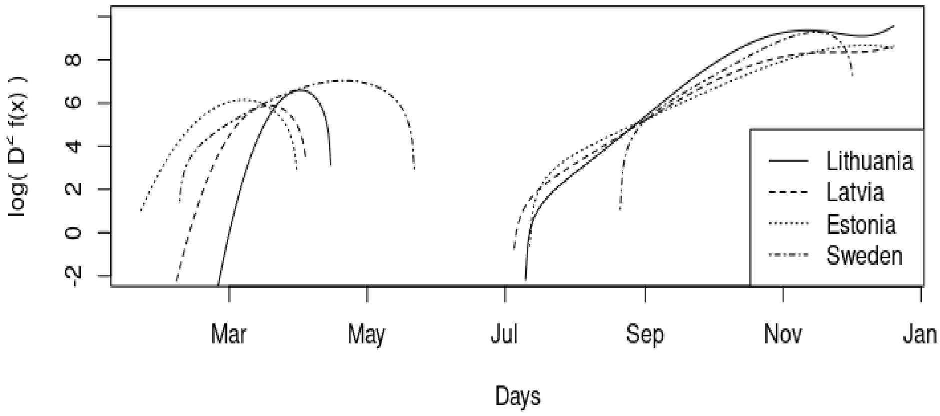


Figure 21: Second order derivatives of four Baltic countries.

The second order derivative can be even more informative. In figure 21 we plot $log(Q''f(t))$ and empty area is where $Q''(t) \leq 0$.

4.1 Epidemic growth rate similarity

Many factors like restrictions or susceptible people pool may influence the epidemic growth rate therefore each segment may have different characteristics. While each wave can be analyzed independently we compare the rate by measuring the angle between segments. The figure 20 shows the confirmed cases with marked segments and angle between vectors. If the angle α is equal or larger than 180° it may indicate that the successive wave is starting to accelerate or reaching it's peak value, while the lower value indicates that the successive wave is still progressing. Normally, when we take more days into account the curve of the angles should form U-shape, starting with value around $\alpha \approx 180°$ when epidemic catches momentum the angle declines and finally, when epidemic reaches it's peak, the value should again stay around.

To reveal similar patterns between the countries we use functional kmeans as a clustering method. The results are plotted in figure 23. There are strong differences between blue cluster with three countries (Bulgaria, Poland, Czechia) and green and red with the rest of the european countries.

4.2 Epidemic wave acceleration

As we have seen from similarity curves and their clusters, there are some behavioral differences and similarities on how the epidemic waves develops over time.

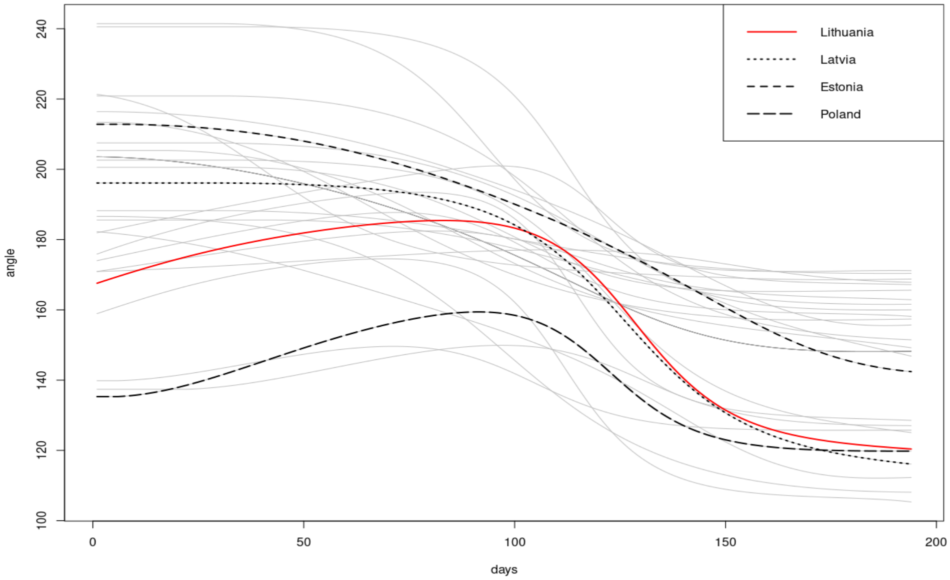


Figure 22: Angle curves of the European countries

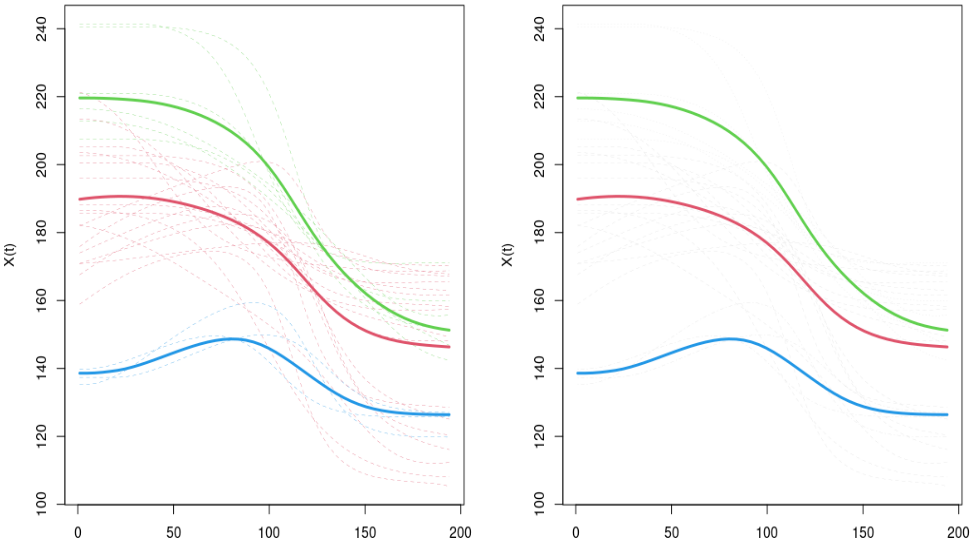


Figure 23: Angle curve clusters

Green	Italy, Estonia, Finland, Germany, Norway, Ireland
Red	Austria, Belgium, Latvia, Lithuania, Luxembourg, Cyprus, Netherlands, Sweden, Denmark, Portugal, Hungary, United Kingdom, Switzerland
Blue	Bulgaria, Czechia, Poland

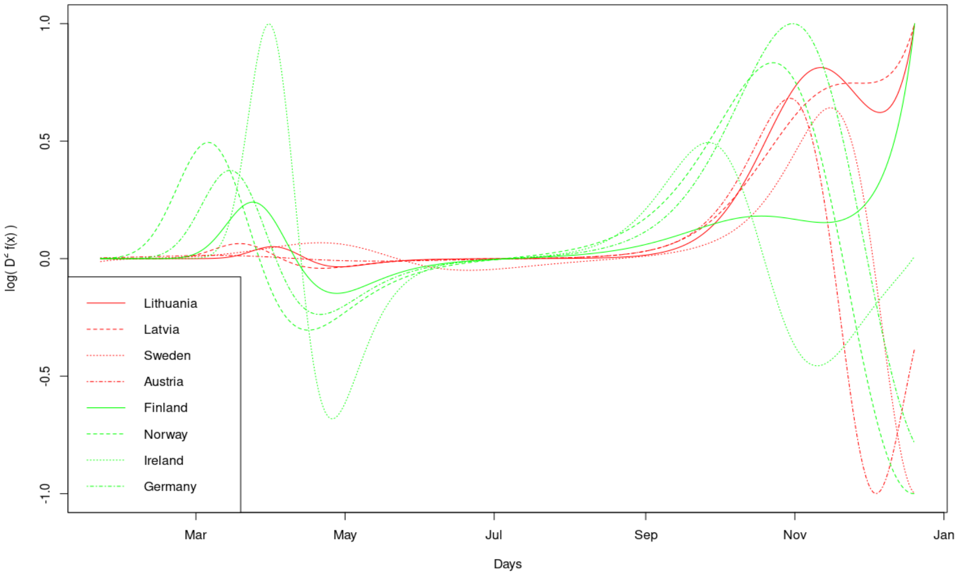


Figure 24: Second order derivatives comparison between different clusters

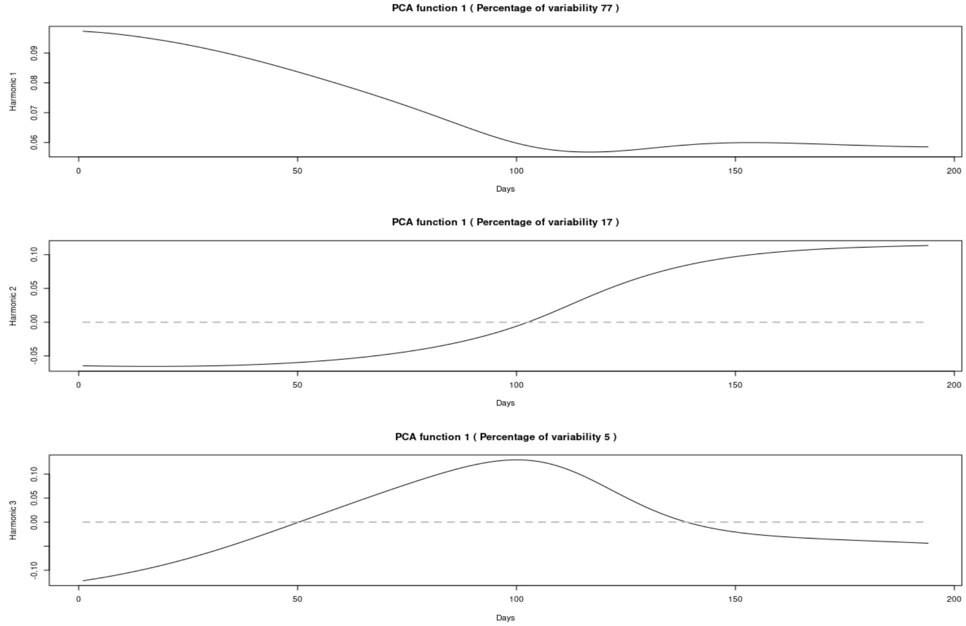


Figure 25: Functional principal components of the angle curves of European countries

Using the power of functional data we look into second order derivative for the explanation of the difference particularly between green and red clusters as seen in the figure 23. The figure 24 shows second order derivatives of confirmed cases for countries belonging to different clusters. We can see, that countries in the green cluster tend to have more symmetric grow speed compared to red cluster. Which explains why starting angle for green cluster are normally larger. There can be different interpretations but one way of explaining this, is that countries in green cluster has very similar behavior during both epidemic waves. This suggests that, compared to other countries where second wave is much stronger, their approach of managing epidemic is more effective. At least during the second wave.

4.3 Similarity principal components

Functional principal components can reveal similarity behavior of the epidemic waves illustrated in figure 25. Three components explains 99% of variability and we can see three primary modes of variations in the similarity behavior.

A more informative way of interpreting the principal component functions or harmonics is to look as perturbations of the mean (see figure 26). The + and − signs show what happens when a principal component is added or subtracted to the mean. The first component which explains majority of the

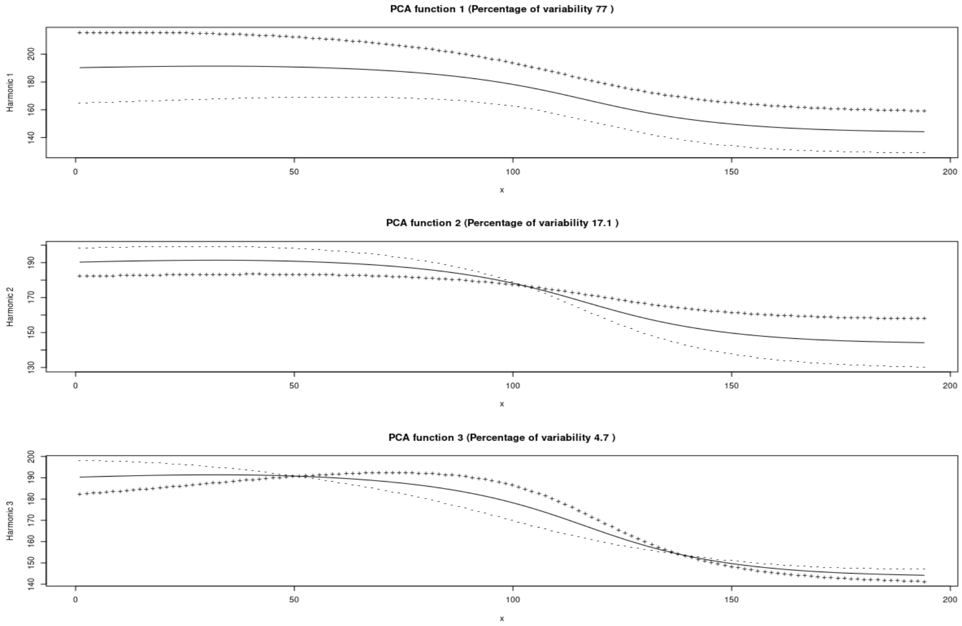


Figure 26: Functional principal components of the angle curves of European countries are shown as perturbations of the mean, which is solid line. The + and − signs show what happens when a principal component is added or subtracted to the mean.

variations explains common pattern, slightly less variation is captured at the end. Which may indicated that the progress looks very similar for all countries at the start and later, when countries implement different strategies to combat with pandemic diverges from the mean.

5 Covid-19 effect on the regions

In order to investigate the position of all countries in the presence of Covid-19, a model is being developed to assess the impact of a pandemic.

5.1 Modeling quarantined, recovered and closed cases

Each country belongs to one of the five regions: Europe, Asia, Americas, Africa, Oceania. We denoted by $Q_j(t)$ the quarantined cases at location j on the tth day of the year. In order to see the differences among the regions, the model follows

$$Q_j(t) = \mu(t) + q_{r(j)}(t) + \varepsilon_j(t) \tag{4}$$

with modeled period from January 11, 2020 to December 20, 2020. The function μ represents the overall pattern of quarantined cases across the sample that represents the world. The function $i \to r(j)$ assigns one of the five regions to the location j. The function q_r describes the effect of the location being in region r, with $r = 2$ for Europe, $r = 3$ for Asia, $r = 4$ for Americas, $r = 5$ for Africa and $r = 6$ for Oceania. To ensure identifiability, we assume that $\sum_{r=2}^{6} q_r(t) = 0$.

Figure 27 depicts estimated coefficients of the model (4). An intercept function illustrates an overall average of the behavior of the virus analyzed in the model, along with coefficient functions that capture deviation from this average for each of five regions. From the results of Figure 27, the least affected by the number of quarantined cases of Covid-19 is Americas region, since the deviation from overall mean is decreasing, which means that fewer people get sick compared to the previous period. Africa is the in the worst position affected by the quarantine cases of Covid-19 since the end of autumn of 2020 as more people tend to be infected by the corona virus. In Asia, Europe and Oceania regions the situation is in average the same compared to the overall mean. Although the number of quarantined cases in Europe from 1 June, 2020 to August 31, 2020 has slightly declined and then came to to the overall average, whereas in Oceania of the same period the situation of infected people had worsened, but improved again after the end of the cold season.

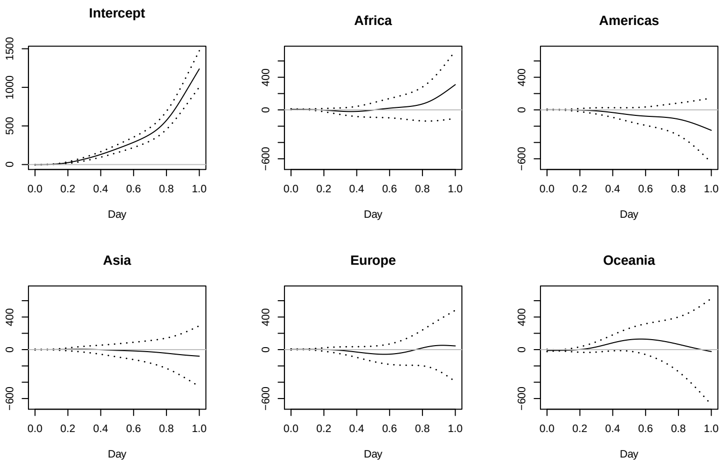


Figure 27: Estimated overall and regions effect for the quarantined cases of COVID-19

Equivalently, models for recovered $R_j(t)$ and closed $D_j(t)$ cases are defined. Figure 28 depicts model results for the recovered cases, whilst Figure 29 illustrates the Covid-19 position in closed cases. The main difference between the two models is that recovery is expected to have as high a function value as possible and, in the case of closed cases, loss to a minimum. Figure 28 reveals that in Africa the probability of recovery increases recently compared to the rest of the world and decreases for the Americas region. But such effect is due to increasing and decreasing number of quarantined cases in the latter regions.

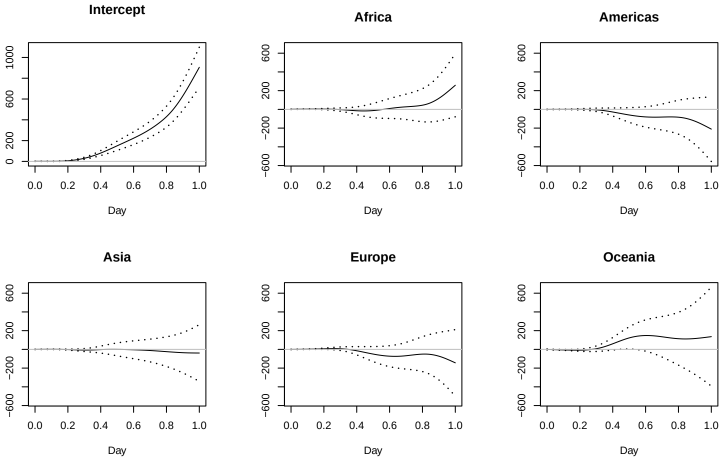


Figure 28: Estimated overall and regions effect for the recovered cases of COVID-19

Figure 29 shows the deteriorating situation in Africa and the improving in Oceania regions. However, again, this depends on an increasing or decreasing number of quarantined cases.

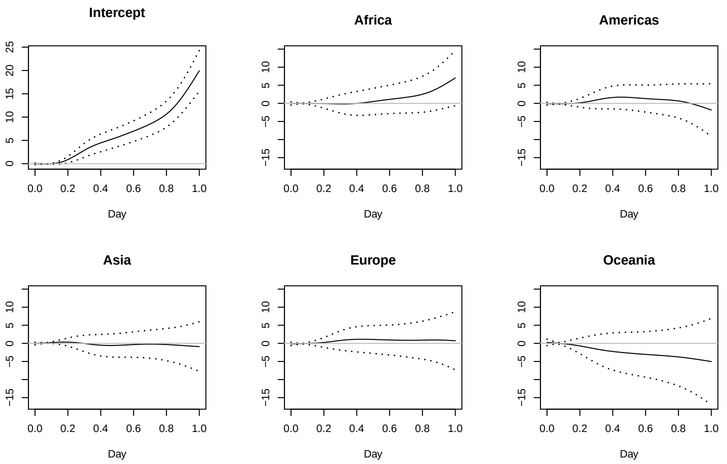


Figure 29: Estimated overall and regions effect for the closed cases of COVID-19

As we see the number of quarantined, recovered and closed cases analyzed separately does not disclose the complete context about the Covid-19 situation in the world. Therefore, mortality and recovery rates are examined, which take into account the number of quarantined cases.

5.2 Modeling mortality and cure rates

The same concept of model (4) is applied to design the mortality $\widehat{\kappa_j}(t)$ and cure $\widehat{\lambda_j}(t)$ rates with modeling period July 1, 2020 - December 20, 2020. Figure 30 discloses estimated coefficients for the mortality rate of Covid-19. In overall, since the beginning of July, 2020 the mortality rate has declined, which might show that people have acquired immunity to the virus or treatment methods have made it possible to control the spread of the disease in the human body, even though no vaccine has been developed. However, in the beginning of July, 2020 the mortality rate was even higher in Africa region, while it was lower in Europe, Americas and Oceania regions. Nevertheless, as of the beginning of August, 2020 mortality rates are similar around the world. Additionally, it is likely that mortality rates should be further reduced once people are vaccinated.

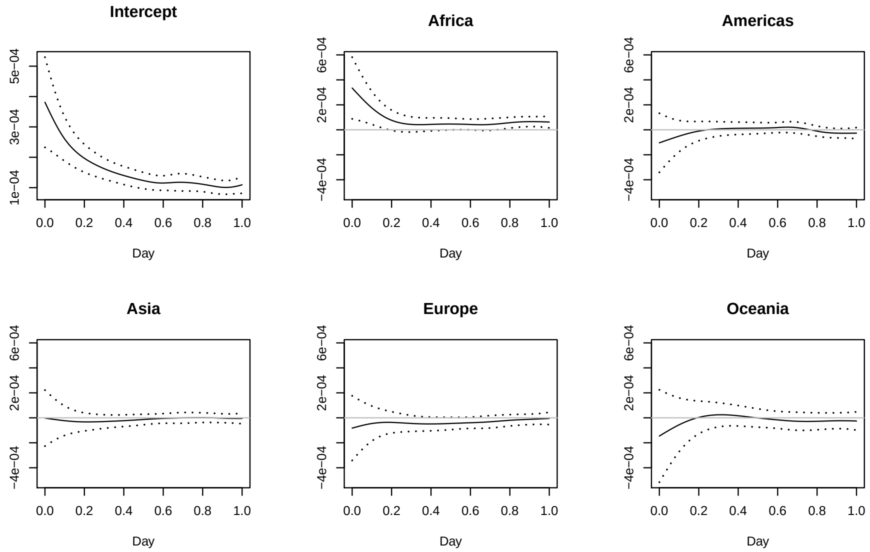


Figure 30: Estimated overall and regions effect for $\widehat{\kappa_j}(t)$ of COVID-19

In Figure 31 In overall, since the beginning of July, 2020 the cure rate has declined. This effect might be explained by the fact that more and more people are infected and there is an insufficient number of beds in treatment facilities, where a certain percentage of infected people are treated at home and do not have emergency care if the situation worsens. Surprisingly, the highest recovery rate is in Africa. As of the end of the autumn of 2020, the situation has deteriorated in the European and American regions, whilst in Asia the cure rate has risen.

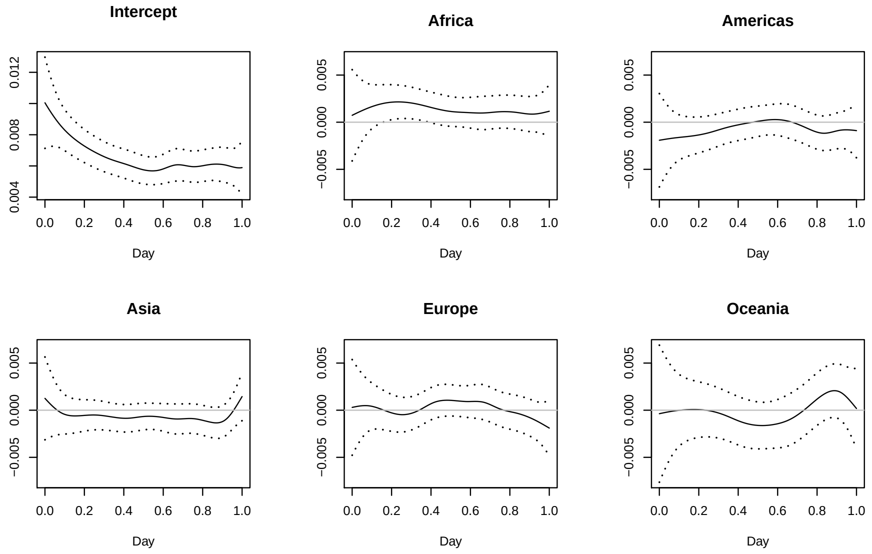


Figure 31: Estimated overall and regions effect for $\widehat{\lambda_j}(t)$ of COVID-19

References

[1] Cuevas, A., Febrero-Bande, M., Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22 (3): 481–496.

[2] Kokoszka, P. Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Chapman Hall/CRC, New York.

[3] Ramsay J.O., Silverman B.W. (1997). *Functional Data Analysis*, Springer Series in Statistics, 310 p.

[4] Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, London.

[5] Ramsay, J. O. & Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer-Verlag, New York.

[6] Chen, K. and Müller, H.-G. (2012). Conditional quantile analysis when covariates are functions, with application to growth data. J. R. Stat. Soc. Ser. B Stat. Methodol. 74 67–89. MR2885840

[7] Tang, R. and Müller, H.-G. (2009). Time-synchronized clustering of gene expression trajectories. Biostatistics 10 32–45.

[8] GROMENKO, O. and KOKOSZKA, P. (2013). Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination. Comput. Statist. Data Anal. 59 82–94. MR3000043

[9] Bircan Erbas, Muhammed Akram, Dorota M Gertig, Dallas English, John L. Hopper, Anne M Kavanagh, and Rob Hyndman, Using Functional Data Analysis Models to Estimate Future Time Trends in Age-Specific Breast Cancer Mortality for the United States and England–Wales, J Epidemiol 2010;20(2):159-165 doi:10.2188/jea.JE20090072.

[10] Daniel J. Levitin, Regina L. Nuzzo, Bradley W. Vines, Davis J. O. Ramsay, Introduction to Functional Data Analysis, Canadian Psychology, 2007, Vol. 48, No. 3, 135-155. DOI: 10.1037/cp2007014

[11] Reza Sameni, Mathematical Modeling of Epidemic Diseases; A Case Study of the COVID-19 Coronavirus. Preprint: https://arxiv.org/abs/2003.11371

[12] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, Mathematical models in population biology and epidemiology. Springer, 2012, vol. 2.

[13] López, L., Rodó, X., A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics, Results in Physics (2020), doi: https://doi.org/10.1016/j.rinp.2020.103746

[14] Y. Chen, J. Cheng, Y. Jiang, K. Liu, A time delay dynamical model for outbreak of 2019-ncov and the parameter identification, arXiv preprint arXiv:2002.00418.

[15] B. Cantó, C. Coll, E. Sánchez, Estimation of parameters in a structured sir model, Advances in Difference Equations 2017 (1) (2017) 33.

[16] S. Ma, Y. Xia, Mathematical understanding of infectious disease dynamics, Vol. 16, World Scientific, 2009.

[17] C. Liu, G. Ding, J. Gong, L. Wang, K. Cheng, D. Zhang, Studies on mathematical models for sars outbreak prediction and warning, Chinese Sci Bull 49 (21) (2004) 2245–2251.

[18] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of covid-19 in china by dynamical modeling, arXiv preprint arXiv:2002.06563.

[19] C. Liu, G. Ding, J. Gong, L. Wang, K. Cheng, D. Zhang, Studies on mathematical models for sars outbreak prediction and warning, Chinese Sci Bull 49 (21) (2004) 2245–2251.

[20] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1