AIMS Mathematics, 7(4): 5347–5385.
DOI: 10.3934/math.2022298
Received: 15 November 2021
Revised: 20 December 2021
Accepted: 26 December 2021
Published: 06 January 2022

http://www.aimspress.com/journal/Math

Research article

Functional data analysis: Application to daily observation of COVID-19 prevalence in France

Kayode Oshinubi, Firas Ibrahim, Mustapha Rachdi and Jacques Demongeot*

Laboratory AGEIS EA 7407, Team Tools for e-Gnosis Medical & Labcom CNRS/UGA/OrangeLabs Telecom4Health, Faculty of Medicine, University Grenoble Alpes (UGA), 38700 La Tronche, France

* **Correspondence:** Email: Jacques.Demongeot@univ-grenoble-alpes.fr.

Abstract: In this paper we use the technique of functional data analysis to model daily hospitalized, deceased, Intensive Care Unit (ICU) cases and return home patient numbers along the COVID-19 outbreak, considered as functional data across different departments in France while our response variables are numbers of vaccinations, deaths, infected, recovered and tests in France. These sets of data were considered before and after vaccination started in France. After smoothing our data set, analysis based on functional principal components method was performed. Then, a clustering using k-means techniques was done to understand the dynamics of the pandemic in different French departments according to their geographical location on France map. We also performed canonical correlations analysis between variables. Finally, we made some predictions to assess the accuracy of the method using functional linear regression models.

Keywords: canonical correlation; clustering; COVID-19; prediction; functional data analysis; functional principal components analysis; functional linear regression
Mathematics Subject Classification: 62R10, 62P10

1. Introduction

1.1. Historical sketch

Daniel Bernoulli described in his 1760' article [1] the three dynamical modes of a contagious disease, namely the endemic, epidemic and eradicated behaviors and he took as an example a contagious disease very common at the time, the smallpox. Eradication corresponds to the disappearance of the disease in the human species. This was the case for smallpox, which disappeared permanently 40 years ago, according to the World Health Organization (WHO) which certified the global eradication of the disease in 1980, the last major European outbreak of smallpox having been observed in 1972 in Yugoslavia and the last naturally occurring case having been diagnosed in October 1977 [2].

In the past, smallpox had been the subject of measures similar to those implemented today for COVID-19:

1) The inoculation was a vaccination scheme, implemented in the Ottoman Empire during the 16th century. The original idea is probably originated in India, because some ancient Sanskrit medical texts described the process of inoculation [3], then the first practice of inoculation is documented in China as the late 10th century, and reached a wide practice by the 16th century, during the Ming dynasty [4]. Reports on the Chinese practice of inoculation were received by the Royal Society in London in 1700 and Lady Mary Wortley Montagu, spouse of the British ambassador in Turkey observed smallpox inoculation during her stay in the Ottoman Empire and promoted it in England upon her return in 1718 [5].

2) In 1796, Edward Jenner, a doctor in Berkeley (England,) discovered that an immunity to smallpox was produced by inoculating a person with material from a cowpox lesion in cow (an animal disease similar to small pox). He called the living material used for this new type of inoculation vaccine (from Latin word vacca, the cow).

3) Before the vaccination, many countries adopted quarantine measures during the epidemic waves of variola.

Daniel Bernoulli proposed his famous mathematical model to describe the smallpox epidemic waves, using the first the logistic model as a phenomenological approach to fit data corresponding to the succession of the new daily cases of the disease. He described the endemic state of the disease after a wave, but did not propose a model for describing the endemic dynamic (now considered as a stationary stochastic process with constant average and variance), nor used a precise method to detect the critical boundary times corresponding to ruptures between epidemic and endemic phases. However, we can consider that with his SI (susceptible-infected) model:

$$dS/dt=-ßSI-\mu S+f+\sigma I,$$

$$dI/dt=ßSI-\nu I-\rho I, \tag{1}$$

with $S(0)=S_0$, $I(0)=I_0$ and where S(t) (resp. I(t)) is the size of susceptible (resp. infected) at time t, ß the transmission rate, μ (resp. ν) the natural (resp. viral) death rate, f the fecundity, σ (resp. ρ) the post-infected non immunized (resp. immunized) rate. If $f=\mu=\nu=\rho=\sigma=0$, the solution of Eq (1) is given by:

$$S(t)=S_0/(1+e^{ß(t+to)}),$$

$$I(t)=S_0e^{ß(t+to)}/(1+e^{ß(t+to)}), \tag{2}$$

with $S+I=S_0$. If $I(0)=1$, then if $t_0=-Log(S_0-1)/ß$, $S(0)=S_0/(1+e^{ßto})\approx S_0-1$. If S_0 is large, then during the time lapse at which $S(t)$ remains close to S_0, if $\mu S_0=f$, then Eq (1) becomes close to the linear system:

$$dS/dt=(-ßS_0+\rho)I,$$

$$dI/dt=(ßS_0-\nu-\rho)I, \tag{3}$$

with the solution:

$$S(t)=(-ßS_0+\rho)e^{(ßSo-\nu-\rho)t}/(ßS_0-\nu-\rho),$$

$$I(t)=e^{(ßSo-\nu-\rho)t}, \tag{4}$$

where $ßS_0$ represents the basic reproduction number of the epidemic disease, usually denoted R_0.

The Bernoulli model has been improved for modelling malaria first by Ross, who added the animal vector [6] and after by McKendrick [7], who considered a latency period between the moment of the mosquito's contamination and the one when it actually becomes infectious. More recently, models taking into account the spatial diffusion of infected and infectious agents are proposed [8–11] as well as some models including not reported and vaccinated patients [12–16].

The present models are in general used to simulate an epidemic wave, as an excursion in the phase plane of an ordinary two-dimensional differential system having some excitable ability, i.e., the existence in his phase portrait of large return trajectories after perturbation of its stable stationary state or of its parameters, these trajectories returning to either an endemic or an eradicated final state. The difference between these two asymptotic behaviors is that in endemic case, the final value of the infected, back to the stationary state is non-zero and, in eradicated case, this value is zero. These behaviors are illustrated by simulations of the Bernoulli model in the possible final states corresponding to epidemic (transients in Figure 1), eradication (end state in Figure 1 top) and endemic (low level end in Figure 1 middle and high level in Figure 1 bottom).

If we consider that a parameter of the Bernoulli model like the transmission rate ß is evolving in time due to various influences as geoclimatic factors (temperature, humidity, elevation, etc.), sociodemographic determinant (density and median age of the target population) and economic variables (GDP, Gini's index, inequality index, etc.), the actual curve of the new infected cases observed during the epidemic outbreak is in fact also driven by the slow dynamics of the seasonally varying or linearly growing parameters.

Then, the observed dynamics of new infected cases results from the epidemic dynamics combined with slow evolution of these external factors on variation surfaces mixing both epidemic variables and parameters dynamics. The Figure 2 shows such a surface corresponding to the variations of the transmission rate ß due to climate changes (transition between winter and summer times) or public health policies (quarantine, vaccination, etc.), and the actual (S, I) trajectory lies on the corresponding surface. The influence of these different factors on the variations of the trajectories (S, I) are difficult to take into account in a differential model of the type Eq (1), which is why the article will now emphasize the statistical approach providing a better understanding of the correlations between the exogenous determinants of the epidemic and the level of new cases of infected.

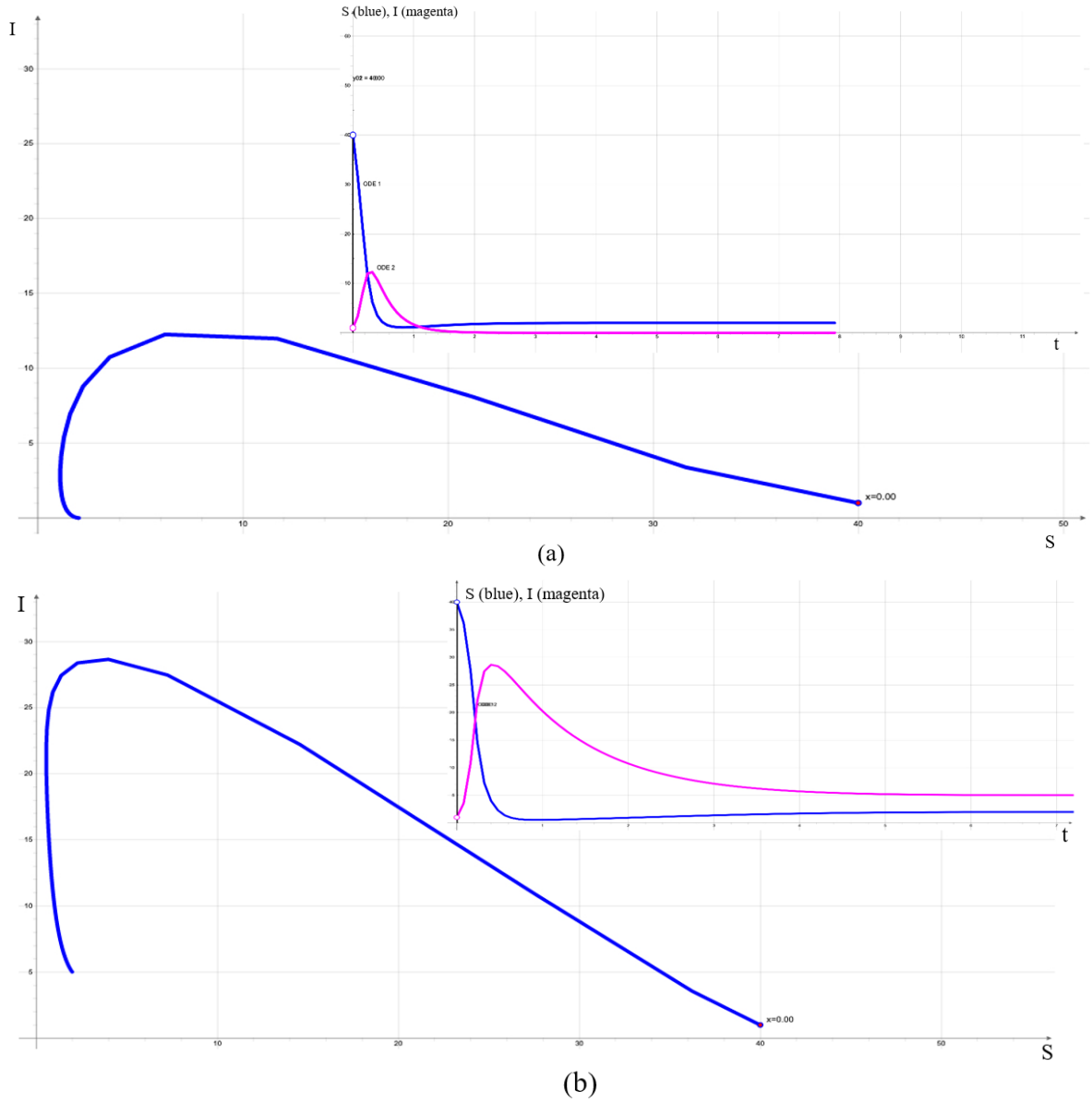


Figure 1. Dynamics of the variables S and I in eradication (top) and endemic behavior (bottom) (after https://elsenaju.eu/ Calculator/ODE-System-2x2.htm#RWD-1). (a) Eradication ß=0.6, μ=2, f=4, ν=4, ρ=0, S(0)=40, I(0)=1; (b) Endemic behavior ß=0.5, μ=0.5, f=6, ν=4, ρ=0, S(0)=40, I(0)=1.

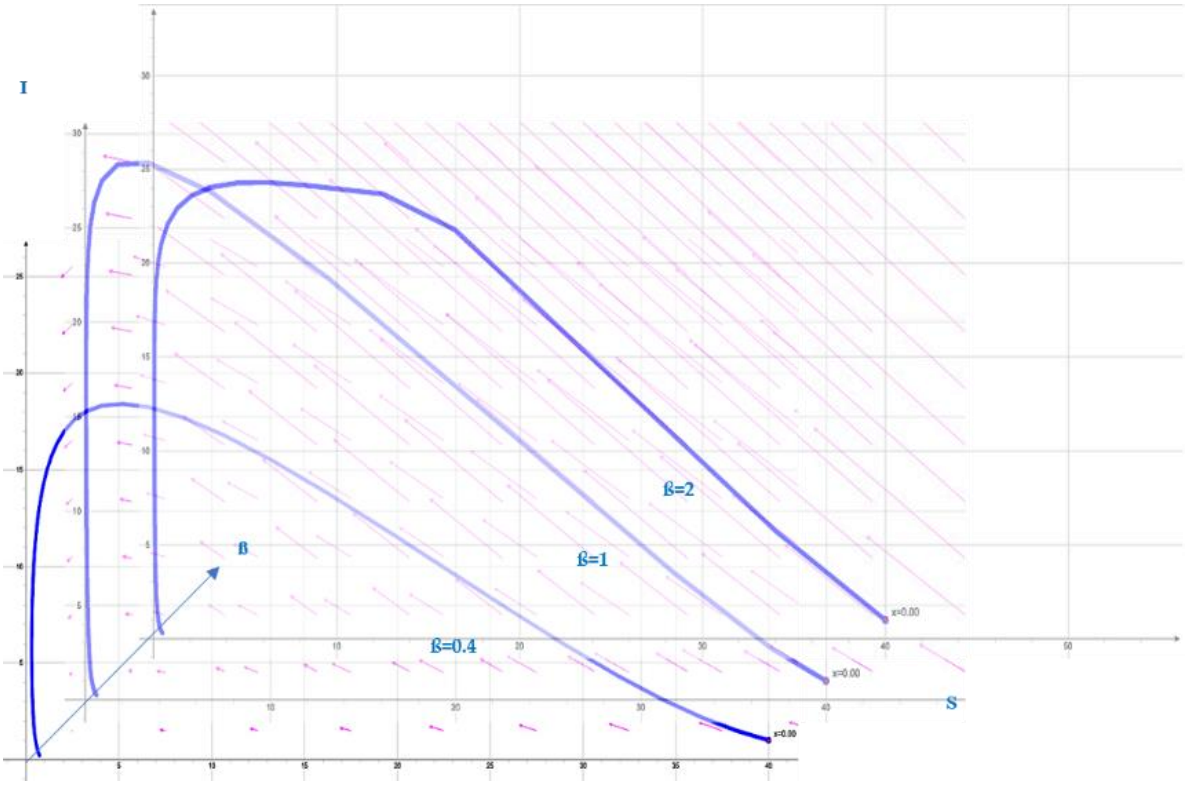


Figure 2. Surface of evolution of 3 covariables, S, I and ß, the transmission rate.

1.2. Background and literature review

The COVID-19 pandemic is still evolving in France as there has been three waves with possibility of a fourth wave due to a more contagious variant (Delta variant) which may lead to another lockdown following three lockdowns alongside with several non-pharmaceutical measures to mitigate the spread of the diseases. France has a total of 7,275,149 cases as at 14/11/2021, 118,137 deaths representing 2% of the 5,280,894 cases which had an outcome (5,162,757 having recovered from the disease), and 1,994,255 currently infected patients with 1,993,206 (**99.9**%) in mild condition and 1,049 (**0.1**%) in serious or critical condition (data from [17–19]). Modeling COVID-19 pandemic across the globe has been approached using different techniques in mathematics and statistics, but the use of functional data analysis (FDA) has been done by few scientists.

Functional data analysis is useful in many fields such as medicine, biology, statistical analysis and econometrics while several books like [20–25] have treated the theoretical aspects and methodology and more recently, researchers have dealt with FDA application to COVID-19 modeling [26–28], trying to propose forecasting new case counts through a framework facilitating the quantification of the effects of demographic covariates and social mobilities on reproduction number and fatality rate through time-varying regression models. Among these recent works using functional data analysis for the modelling of COVID-19 pandemic, we can notice: [26] applies functional data analysis to United States data by using FCPA (Functional Principal Component Analysis) and FCCA (Functional Canonical Correlation Analysis) tools and they finally use functional time series to fit the cumulative confirmed cases in the United States and make forecasts based on the dynamics of FPCA. [27] works on the imputation of missing data of COVID-19 hospitalized and intensive care curves in Spain

regions. They used function-on-function regression technique to estimate missing values and Canonical Correlation Analysis was performed to interpret the relationship between hospital occupancy rate and illness response variables. The shapes of an epidemic curve using functional data analysis to characterize COVID-19 in Italian regions and their association with mobility, positivity, socio-demographic structure and environmental covariates was worked on by [28]. The cited authors have used different methods of functional data analysis like function-on-function regression techniques, clustering methods and smoothing techniques for the functional data considered.

A robust phenomenological approach to France COVID-19 data has been recently investigated by [29,30] and a new method to calculate the cumulative cases in France was proposed which illustrates the epidemic and endemic nature of the virus infection in France. Authors in [31] used methods like principal component analysis, generalized additive model and hierarchical ascendant classification to study the impacts of population age structure, epidemic spread and transmission mitigation policies on COVID-19 morbidity or mortality heterogeneity in France. Authors in [32–34] used ARIMA models with different parameters to forecast the spread of COVID-19 across nine countries in Europe, Asia and American continents and the study deduced that the method is useful for the prediction of the pandemic at different stages and in [35] they employed statistical methods to analyze the shapes of local COVID-19 incidence rate curves and statistically group them into distinct clusters according to their shapes. The result reveals that pandemic curves often differ substantially across regions of a same country, and the explanation may lie in the existence of a temperature gradient or of differences in other geo-climatic, socio-economic or demographic factors [36–44].

The main purpose of this article is to revisit the data on COVID-19 from public databases using methods that are still little used, such as functional data analysis (FDA), in which there is a great deal of theoretical work [19–25], but practical applications are still rare. In particular, the generalization at the FDA of classical finite-dimensional methods, such as estimation, regression and principal component analysis, shows that it is possible to process epidemic data obtained from a large sample (approximately one million data) concerning the incidence, mortality and exogenous or endogenous factors associated with the COVID-19 epidemic. This sampling concerns about ten variables (sometimes missing, in particular those concerning risk factors) and includes longitudinal (about 600 days all over the world but in our case we considered about 469 days for French departments) and cross-sectional data (about 200 countries, some comprising up to a hundred documented regions while for France which is the main focus of this article there are 101 departments), for the monitoring and prediction of a pandemic whose origin and end are still uncertain, but which, by its magnitude and its dramatic consequences (around 5 million deaths) justifies such a descriptive statistical investment. The descriptive study of many factors associated with the epidemic (namely the socio-economic and geoclimatic ones) makes it possible to understand closely linked mechanisms, those exogenous or endogenous to the viral pathogeny of propagation and endogenous ones of pathogenicity [36–44]. The seminal paper by Bernoulli on the SI model is the origin of all future discussions about epidemic modelling, in particular the first by d'Alembert in his Opuscules mathématiques and Lambert until its last recent refinements. It contains the explicit solution of the SI model given in Eq (4), and permits to consider the empiric S(t) and I(t) curves as functions. We can manipulate (after smoothing) as elements in a functional space. The goal of the article is to apply to these elements the FDA descriptive techniques in order to compare these functions among different departments. This approach is complementary of the works estimating the parameters of the functions S and I (namely R_0 and ß).

1.3. Time series and curve fitting

In this section, we present the time series analysis of daily new cases in France, daily hospitalization in three French departments out of 101 in France, and curve fitting for two French departments. The Root Mean Square Error (RMSE) for other curve fittings is shown in Table 1. Figure 3a gives the time series of recent daily cases of COVID-19 in France which shows stationarity with rolling values (window=12) appearing to be varying slightly. Also, when we used the Augmented Dickey-Fuller Test to test for stationarity of the time series by accepting the alternative hypothesis, we got a p-value less than the 5% (p-value=0.02) critical value, so we can say with 95 percent certainty that this is a stationary series. Also, in Figure 3c we plotted three French departments (Nord, Paris and Essonne) with more prevalent COVID-19 hospitalization cases and Figure 3d shows the fitted curve of two of the French departments (Paris and Seine-Maritime) while all departments have root mean square error in the interval $0.51 \leq RMSE \leq 17.38$ with Essonne department having the highest Root Mean Square Error (RMSE) and Lozère department having the lowest RMSE. We present other RMSE values in Table 1. We present also (Figure 3b) a deep learning forecasting result using Gated Recurrent Units (GRU) for France data between the beginning of the pandemic in France till September 3 2021 by training 80% of the data and testing 20%. The predicted cases curve values decline over the whole-time. The RMSE was computed using:

$$\text{RMSE}=\sqrt{\frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}(\text{Y}_\text{i}-\text{y}_\text{i})^2}$$

where for $\text{i}=1,2,\ldots,\text{n}$, $\text{Y}_\text{i}'\text{s}$ are the observed values, n is the number of data points and $\text{y}_\text{i}'\text{s}$ are the predicted values.

The aim of this paper is to model the prevalence of the virus in France by using several functional techniques like FCCA, K-means clustering and FPCA and to finally make some predictions about the evolution of the disease in France. The analysis was done using both Python and R packages. We considered as functional variables numbers of ICU cases, daily deceased, daily return home and hospitalization which are given as X_1-X_4. Our response variables given as Y_1-Y_6 are numbers of recovery, deaths, infected, vaccination, vaccination per 1000 population and number of tests. We used data from [17–19]. The paper is divided as follows: Section 2 describes the various smoothing methods used in the analysis of the shapes of the functional data used, Section 3 presents the functional principal components analysis results and their interpretation to the dynamics of COVID-19 prevalence in French departments, and Section 4 is dedicated to the results of canonical correlation of the variables. Section 5 shows the clustering result using the K-means method and how it appears on a map of France. In Section 6, we made some predictions for some response variables and also performed function-on-function linear regression, and in Section 7, we opened up some perspectives and presented the results of the analysis.

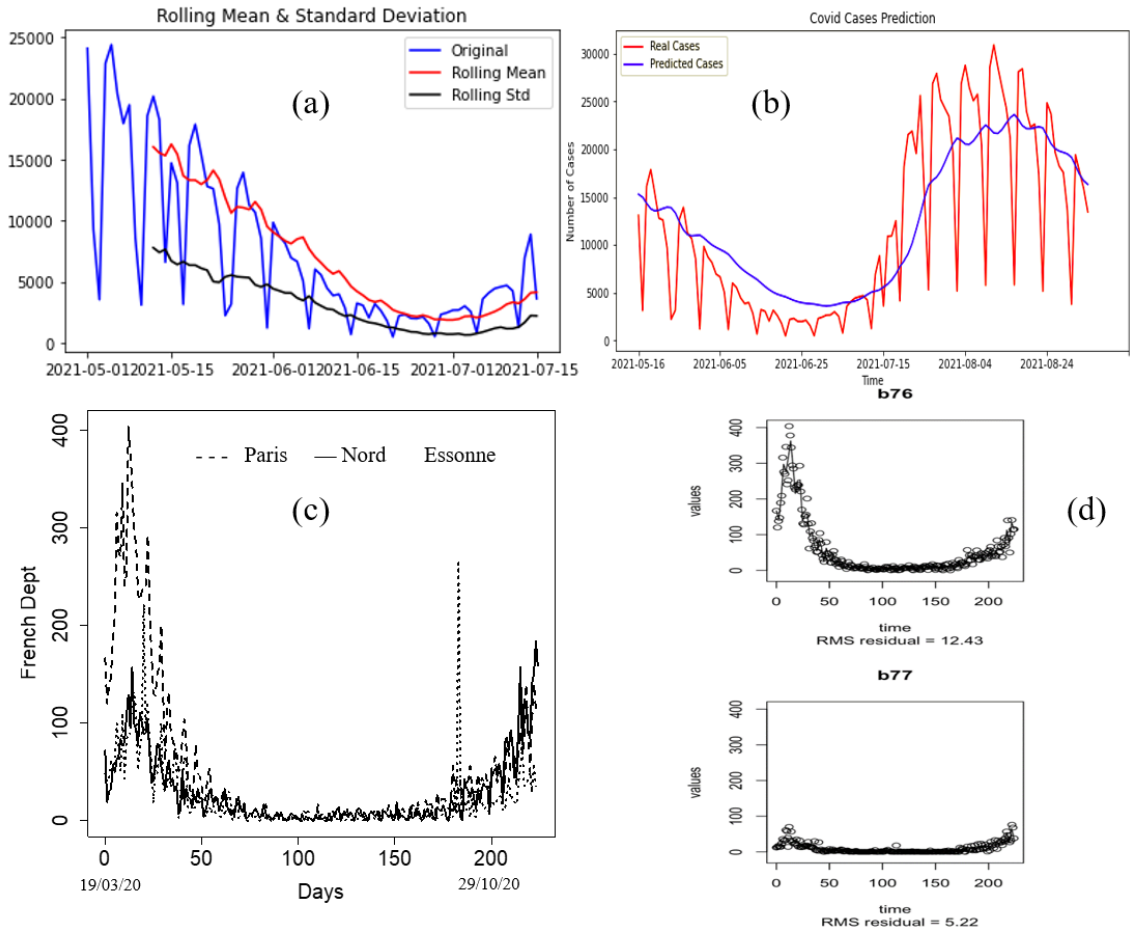


Figure 3. (a) Time series modelling of daily new cases between 01/05/2021–15/07/2021 in France. (b) GRU deep learning forecasting method for daily new cases between 25/02/2020–03/09/2021 in France. (c) Daily hospitalization cases in three French departments: Nord, Paris and Essonne. (d) Fit curve for hospitalization cases in Paris and Seine-Maritime.

Table 1. RMSE confidence interval for all French departments for the fitness curve of the four functional data.

	RMSE before vaccination started	RMSE after vaccination has started
Hospitalized	0.51≤RMSE≤17.38	1.00≤RMSE≤18.00
ICU	0.05≤RMSE≤2.60	0.35≤RMSE≤5.20
Daily return home	0.25≤RMSE≤12.40	1.10≤RMSE≤17.50
Daily deceased	0.04≤RMSE≤4.52	0.32≤RMSE≤4.10

2. Data smoothing

The first step in analyzing functional data is to smooth the curves. In this section we use different smoothing techniques which we shall illustrate and give some basic explanation of the techniques we deployed for smoothing our functional data. We plotted the mean of the data set and the cross-sectional mean, which corresponds to the karcher-mean under the L2 distance. The karcher-mean has an

important role in the warping framework and is used to align the functions in order to improve the matching of features (peaks) across functions for a given set of warping functions $\partial_1, \partial_2, \ldots, \partial_n \in \Delta$ as

$$\overline{\partial_n} = argmin_{\partial \in \mathrm{L2=S}\Delta} \sum_{i=1}^{n} dist(\partial, \partial_i)^2.$$

S is a quotient space, Δ be the differential geometry, $\partial \in \Delta$, ∂_i is the set of warpings functions [24]. We used the *elastic_mean(fd)* and *fd.mean* tool in Python to do the plotting of Figure A1 (supplementary material).

B-spline technique is one of the tools used in smoothing a functional data and this can be done by changing the number of elements (n=2,3,4,…,p) in the basis functions [25]. Sometimes one can use the Fourier basis for the functions to further see the variations in the curves. We give a mathematical expression for the basis representation for the curve of the functional data set of the form:

$$f(t) = \sum_{j=1}^{n} b_j \propto_j (t) \tag{5}$$

where $b_j's$ are basis coefficients and $\propto_j 's$ are the basis functions. For this analysis we choose n=7 which we discovered best suit for the modelling of our data as our number of elements and the tool in Python named *basis.BSpline* was used to perform the plotting of the functional data. The result of this smoothing technique can be seen in Figure 4.

In Figure A2 (supplementary material), we present the correlation coefficient between all the departments in France based on the functional data in consideration, in order to see how well our data is well correlated between the departments and it was observed that there is a high correlation between various departments with except in few cases where we observed low correlation as we can see in the contour plots presented in Figure A2a–h (supplementary material).

We used spline interpolation of order 3 and then smooth the interpolation using the smoothness parameter equal to 1.5 in the cubic spline smoothing. This technique is implemented using *interpolation* and *smoothness_parameter* package in Python. We also use the monotone technique and a piecewise cubic Hermite interpolating polynomial (PCHIP) using a Python package called *monotone*. We present some of the results on Figure A3 (supplementary material).

We also performed Kernel smoothing to show how cross validation score varies over a range of different parameters used in smoothing methods. The essence of this section is to estimate the smoothing parameter *h* that better represents functional data. It has been selected by generalised crossvalidation criteria (GCV). The non-parametric method of smoothing for functional data is based on the smoothing matrix M given:

$$m_{ij} = \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right), \tag{6}$$

$$M(h) = \left(s_j(t_i)\right) = \frac{K\left(\frac{t_i - t_j}{h}\right)}{\sum_{k=1}^{T} K\left(\frac{t_k - t_j}{h}\right)}, \tag{7}$$

where K(.) is the Kernel function. We plotted on Figure A4 (supplementary material) the smoothed curves of the functional data set for three different smoothing methods and show the scores through generalised cross-validation for these different methods. The results show a comparable behavior of these scores by varying the smoothing parameter *h*.

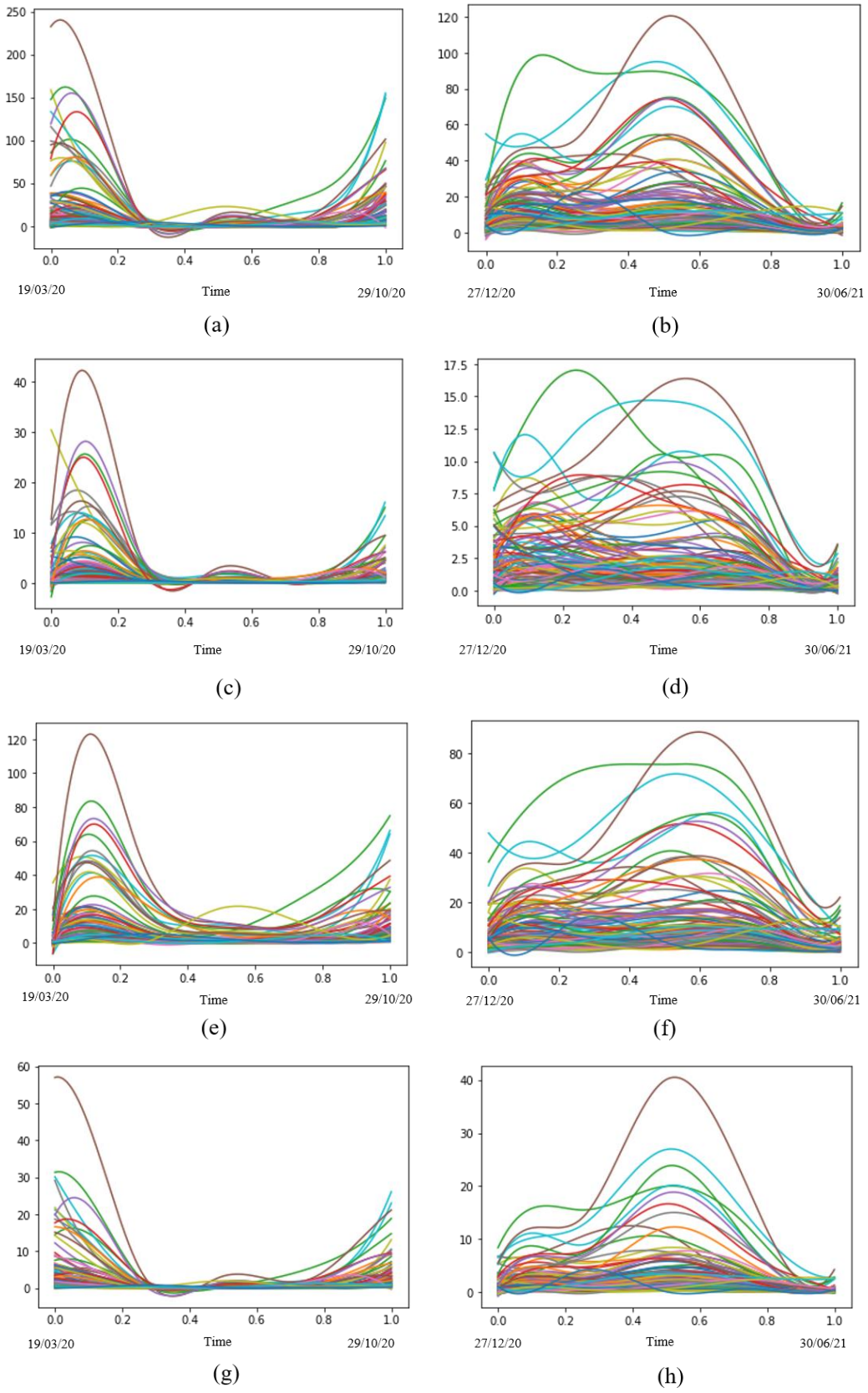


Figure 4. Smoothed curves for the shape of COVID-19 epidemic in all departments in France: (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) daily deceased, (d) daily deceased when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) ICU cases and (h) ICU cases when vaccination has started.

3. Functional principal component analysis (FPCA)

Principal component analysis is a dimension reduction analysis tool in multivariate statistics while functional principal component analysis (FPCA) is a dimension reduction with high correlation in functional data analysis which completes the statistical tools used in the modelling of biomedical data especially epidemiologic ones [45–54]. Let $\{x_i(t)\}_{i=1,m}$ be a given set of functions, in our case m=101 and let α be a weight, FPCA is computed as follows:

1) It finds the principal component weight function $\alpha_1(t)$ for which the principal component score is given by

$$f_{i1} = \int \alpha_1(t)x_i(t)dt, \tag{8}$$

while maximizing $\sum_{i=1,m} f_{i1}^2$ is subjected to

$$\int {\alpha_1}^2(t)dt = \|\alpha_1\|^2 = 1. \tag{9}$$

2) Next, the weight function $\alpha_2(t)$ is computed and the principal component score maximizes $\sum_{i=1,m} f_{i2}^2$, and is subject to the constraint $\|\alpha_2\|^2 = 1$ and to the additional constraint

$$\int \alpha_2(t)\alpha_1(t)dt = 0. \tag{10}$$

3) Then, the process is repeated for as many iterations.

In our analysis, we used a tool called *pca.fd* for the principal component analysis. We present in this section the 4 PCs values plot throughout the days considered and the principal component scores plot for all the different departments providing functional data being before vaccination started and during vaccination.

3.1. Functional PCs

3.1.1. Hospitalization cases

In Figure 5a we observed that PC 1 peaked in the early days of the pandemic between February and March 2020 and then there was a decline after about 50 days becoming stationary till day 150 possibly due to mitigation measures promulgated during this period. The same phenomenon has been observed for PC 2. In Figure 5a, PC 4 shows a sinusoidal shape, peaked at day 100 which is around June 2020 with least values at day 30 and day 180 which are respectively in March and August 2020. Figure 5b shows the same sinusoidal shape for PC 4 and same shape for PC 3 but with a drift in the observation with a difference between the dynamics of hospitalization cases before and after vaccination has started in France. PC 1 in Figure 5b shows a decline across the infective period which may be due to the aggressive vaccination campaign in the country.

3.1.2. ICU cases

In Figure 5c we observed that from day 50 (around April 2020) till day 150 (around July 2020), the PC 1 value which is the major PC is stable throughout this period of various confinement measures in France and all PCs tend to show increasing behavior after the confinement measures have been relaxed and in Figure 5d, PC 1 has strictly positive values while PCs 2–4 show negative values between February to June 2021.

3.1.3. Daily return home

In Figure 5e PC 1 peaked with a positive value at the beginning of the pandemic in France which validates the percentage of recovery as presented in the introduction section while PC 1 in Figure 5f shows a positive decline across the days considered, with a disparity between the period of vaccination and without vaccination.

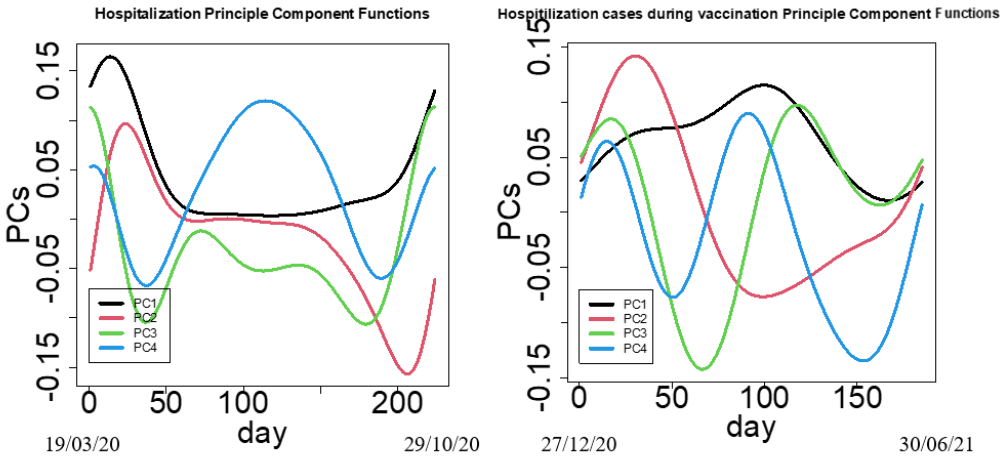
3.1.4. Daily deceased

On the y-axis of Figure 5g,h, we observe that this is the only result with low values for the PCs because the deaths due to COVID-19 in France remain at a low level, while all PCs show almost the same pattern as that observed in previously for the other variables.

In Table 2, we present the PCs variance proportion and we observe that PC 1 is the most important principal component.

Table 2. PCA variance proportion for 4 PCs.

	Before vaccination started				After vaccination has started			
	PC 1	PC 2	PC 3	PC 4	PC 1	PC 2	PC 3	PC 4
Hospitalized	0.945	0.039	0.008	0.005	0.938	0.041	0.012	0.004
ICU	0.960	0.028	0.009	0.001	0.962	0.023	0.008	0.004
Daily home	0.925	0.045	0.015	0.007	0.953	0.030	0.009	0.004
Daily deceased	0.965	0.017	0.013	0.003	0.914	0.055	0.016	0.010



(a) (b)

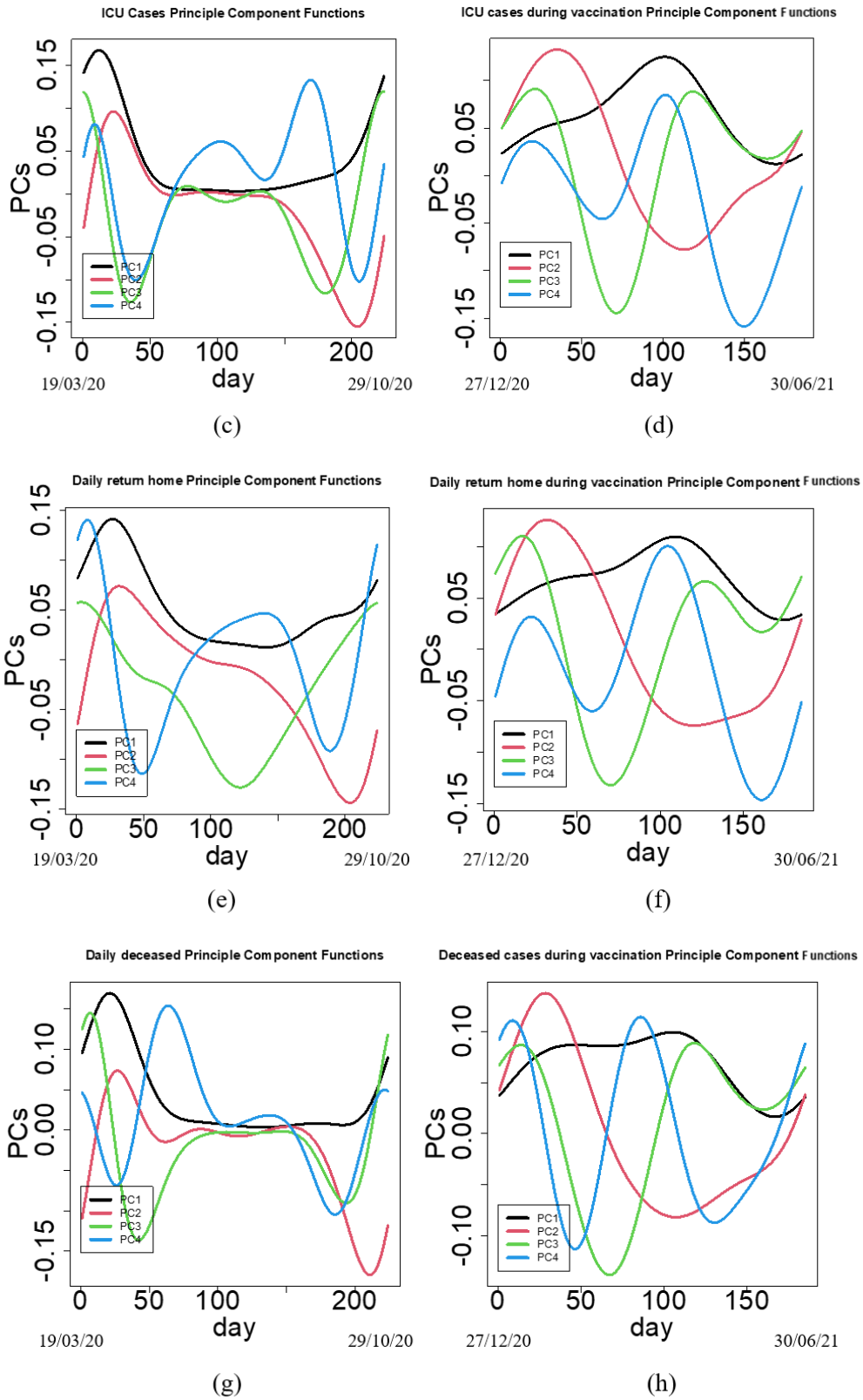
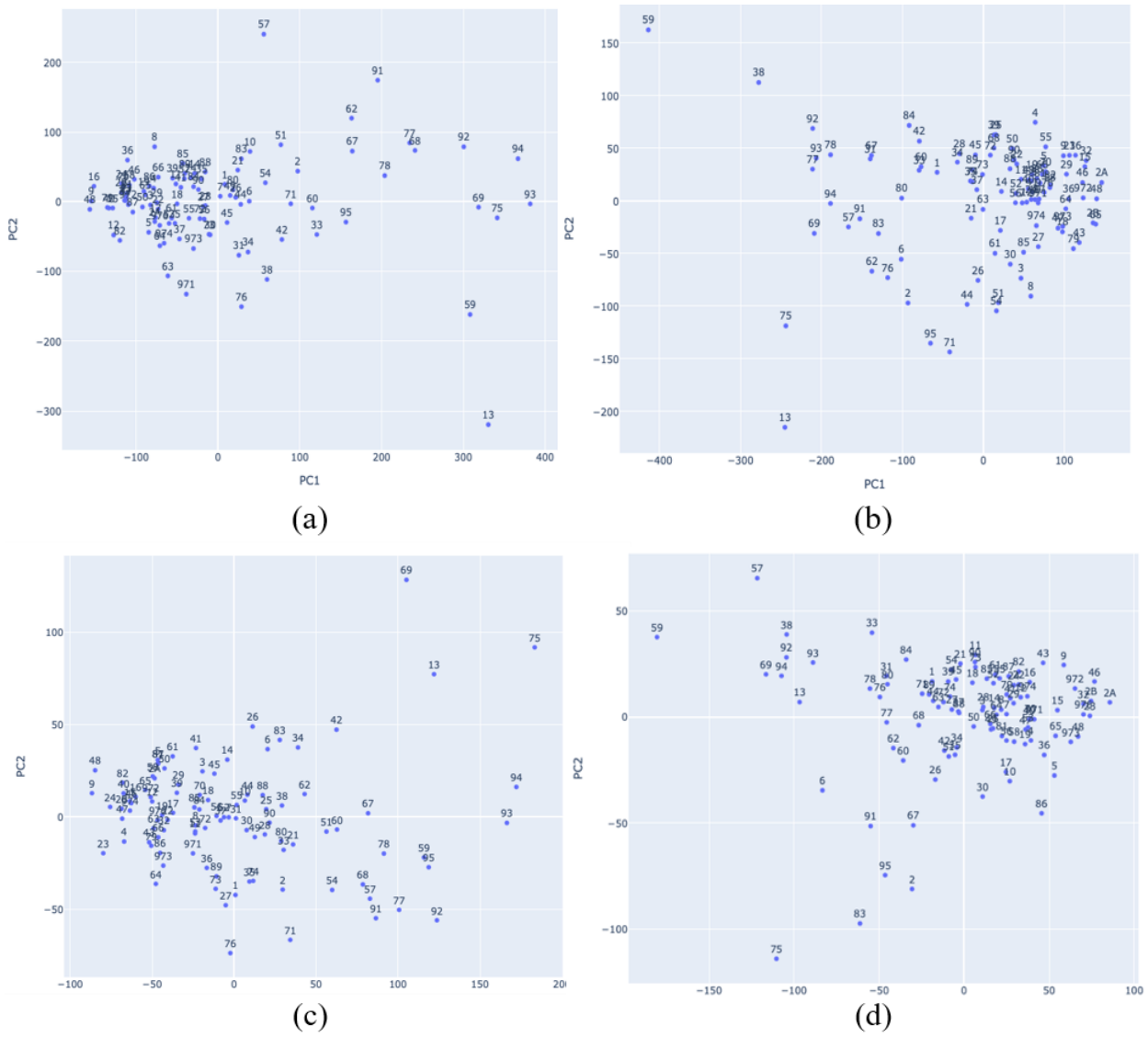


Figure 5. Functional PCs for different functional data before the start of vaccination (19/03/2020–29/10/2020) and when vaccination has started (27/12/2020 to 30/06/2021): (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

3.2. Functional principal component scores and clusters

We will now concentrate our efforts on the departments where the pandemic is most prevalent, as well as on PCs 1 and 2, while ignoring the other PCs. We recall that French departments have code numbers, for example, Nord is code number 59, and we will use this number in this section as well as in the visualization of results shown in Figures 6 and 7. We refer readers to [55] for a complete list of all French department code numbers. In Figure 6a, the Paris department (code number 75) and Nord department (code number 59) have a positive score in PC 1 and negative score in PC 2 while the Essonne department is positive in both PCs. In Figure 6b, the Paris department and Essonne department (code number 91) are negative in both PCs while the Nord department is positive in PC 2 with the highest score and negative in PC 1. In Figure 6c, Nord and Essonne departments are negative in PC 2 but positive in PC 1 while the Paris department is positive in both PCs. The Paris department and Essonne department are negative in both PCs in Figure 6d while the Nord department is positive in PC 2 and negative in PC 1. In Figure 6e, Paris and Nord departments have positive scores in both PCs while the Essonne department is negative in PC 2 and positive in PC 1. Nord department has the highest positive score in PC 1 for Figure 6f and negative for PC 1, Paris department is positive in PC 2 and negative in PC while Essonne department is negative in both PCs. The Paris department has the highest positive score in PC 1 and negative in PC 2 in Figure 6g, Nord department is positive in both PCs while Essonne department is negative in PC 2 but positive in PC 1. Finally, in Figure 6h while Nord department is positive and highest in PC 2, Paris department is the lowest with negative score in PC 2. Both departments are negative in PC 1. Essonne department is positive in PC 2, but negative in PC 1. The above description shows that there is a difference between the vaccination period in France and the period when measures like lockdown, social distancing etc. were only used to control the spread of the virus despite the fact that it has been proven medically that people can be vaccinated and still be infected.

The diagrams in Figure 6a–h show the same shift toward positive PC 1 values. These shifts demonstrate the effect of various mitigation measures differences in the departments and based on population and migration in this area, particularly the five departments outside France's metropole where the rules in France are not strictly enforced. It also demonstrates that PC 1 is the most important PC, from which the majority of the analysis information can be obtained.


PC1
PC2
(a)
(b)
(c)
(d)

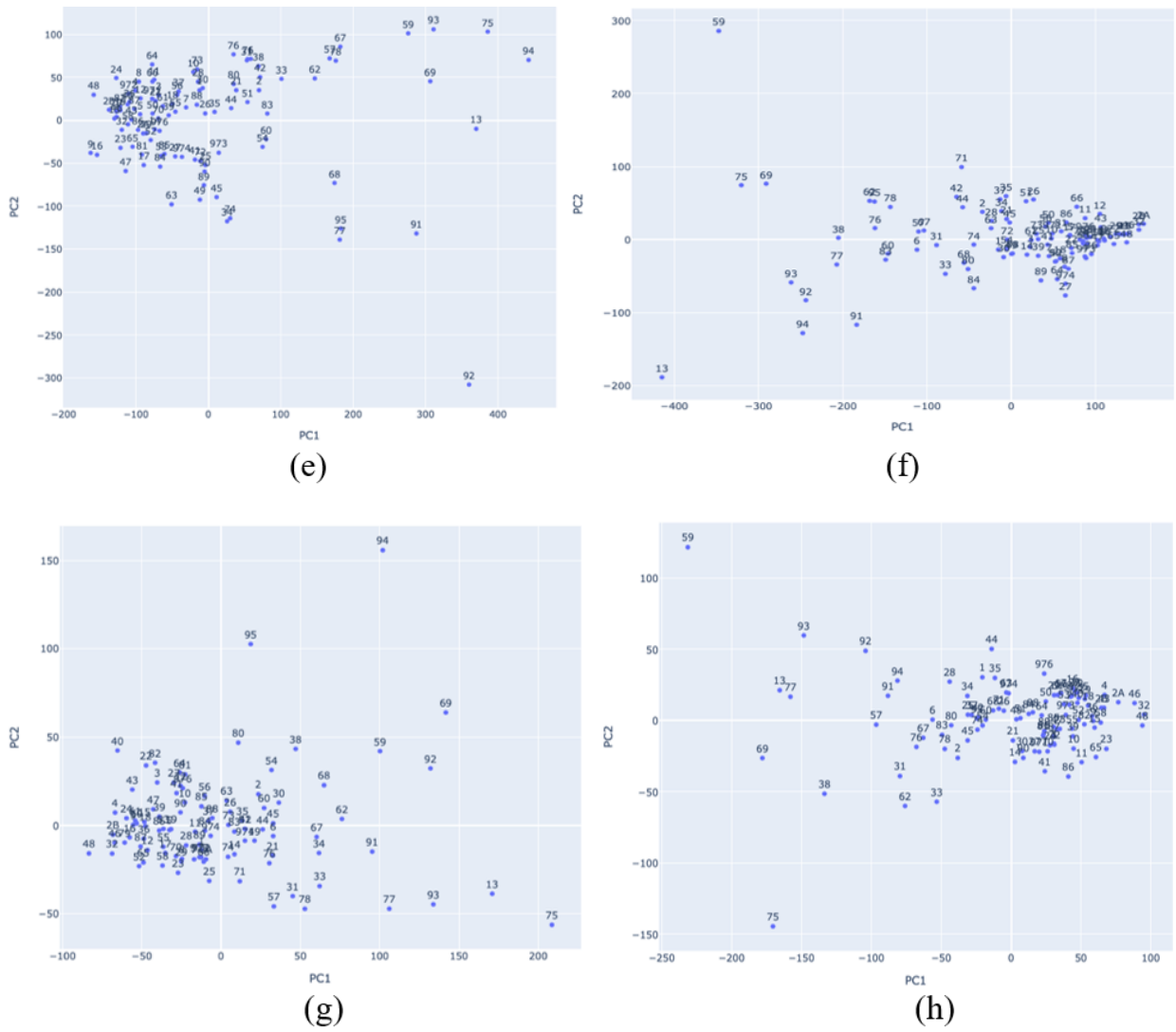


Figure 6. FPCA scores for different functional data before the start of vaccination (19/03/2020–29/10/2020) and when vaccination has started (27/12/2020–30/06/2021). (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started. Note that the numbering of points on the diagram are codes for each French department.

4. Canonical correlation analysis (CCA)

Canonical correlation is an aspect of multivariate statistical analysis method that is used to simultaneously correlate several metric dependent variables and several metric independent variables measured on or observed with similar experimental units. PCA is often used for dimensionality reduction of a particular data set through linear combinations of the initial variables which maximizes the amount of variance explained by these linear combinations while CCA finds linear combinations within a data set with the goal of maximizing the correlation between these linear combinations [23].

Mathematically, it can also be expressed as two groups of n-dimensional variables

$$X=[x_1, x_2, x_3, \dots x_P]$$

and

$$Y=[y_1, y_2, y_3, \dots y_q],$$

where

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ \cdots \\ x_{in} \end{bmatrix}, \; y_j = \begin{bmatrix} y_{j1} \\ y_{j2} \\ y_{j3} \\ \cdots \\ y_{jq} \end{bmatrix}.$$

The purpose of canonical correlation analysis is to find coefficient vectors

$$\boldsymbol{a_1} = (a_{11}, a_{21}, \dots , \; a_{p1})^T \text{ and } \boldsymbol{b_1} = (b_{11}, b_{21}, \dots , \; b_{q1})^T$$

in order to maximize the correlation $\beta = corr(X\boldsymbol{a_1}, \; Y\boldsymbol{b_1})$, while $U_1 = X\boldsymbol{a_1}$ and $V_1 = Y\boldsymbol{b_1}$, linear combinations of X and Y components respectively, constitute the first pair of canonical covariates. Then, the second pair of canonical variates can be found in the same way subject to the constraint that they are uncorrelated with the first pair of variables. By repeating this procedure, the $r = min\{p, q\}$ pairs of the canonical variates can be found and we will finally get the matrix A=$[\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_3}, \dots \boldsymbol{a_r}]$ and the matrix B=$[\boldsymbol{b_1}, \boldsymbol{b_2}, \boldsymbol{b_3}, \dots \boldsymbol{b_r}]$ to transfer X and Y to canonical variates U and V following the below expression:

$$U_{n\times r} = X_{n\times p}A_{p\times r}, V_{n\times r} = Y_{n\times p} \tag{11}$$

If X and Y are both centered, we can concatenate them and calculate the covariance matrix given as:

$$C = Cov([XY]) = \frac{1}{n-1}[XY]^T[XY] = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix}, \tag{12}$$

where C_{xx} and C_{yy} are within-set covariance matrices, and $C_{xy} = \left[C_{yx}\right]^T$ are between-set covariance matrices. The first canonical variates $\boldsymbol{a_1}$ and $\boldsymbol{b_1}$ maximize the equation below:

$$\beta_1 = \frac{\boldsymbol{a_1}^T C_{xy} \boldsymbol{b_1}}{\sqrt{\boldsymbol{a_1}^T C_{xx} \boldsymbol{a_1}} \sqrt{\boldsymbol{b_1}^T C_{yy} \boldsymbol{b_1}}}. \tag{13}$$

The subsequent pairs of canonical variates $\boldsymbol{a_i}$ and $\boldsymbol{b_i}$ (i≥2) maximize:

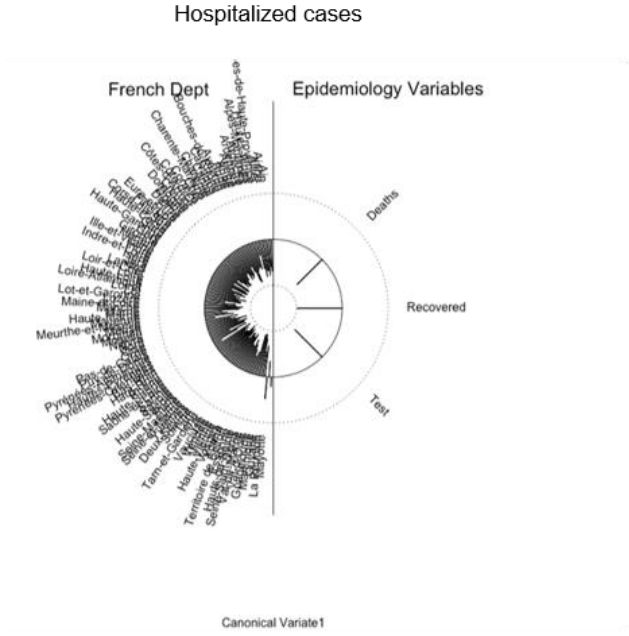
$$\beta_i = \frac{\boldsymbol{a_i}^T C_{xy} \boldsymbol{b_i}}{\sqrt{\boldsymbol{a_i}^T C_{xx} \boldsymbol{a_i}} \sqrt{\boldsymbol{b_i}^T C_{yy} \boldsymbol{b_i}}}, \tag{14}$$

subject to the constraint:

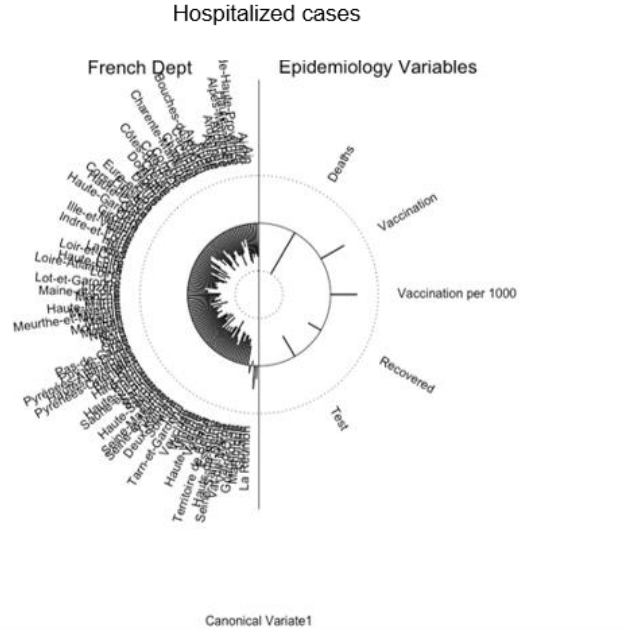
$$\boldsymbol{a_i}^T C_{xx} \boldsymbol{a_j} = 0 \forall j < i,$$

$$\boldsymbol{b_i}^T C_{yy} \boldsymbol{b_j} = 0 \forall \; j < i. \tag{15}$$

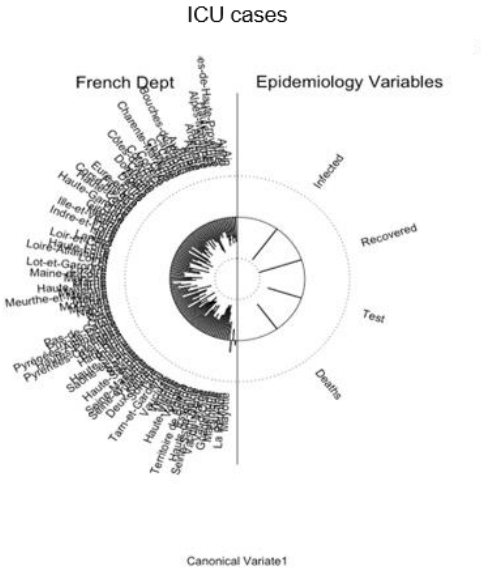
The analysis was performed using a package *CCA*. We present the visualization results on Figures 7 and 8 and also present the correlation scores in tabular form (see Table 3). We used the variables as presented in Table 3. X are the variables are listed in the first row of Table 3 i.e., total number of hospitalizations, daily return home, deceased and ICU cases for all departments and Y variables are the response variables described earlier as presented in the first column of Table 3.



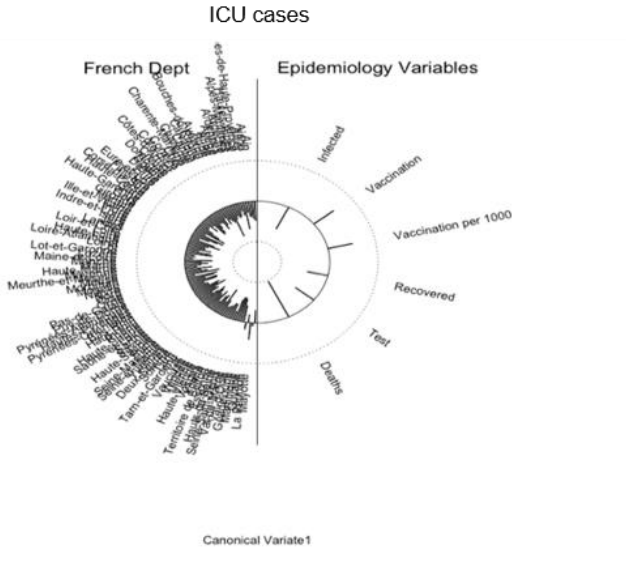
(a)



(b)



(c)



(d)

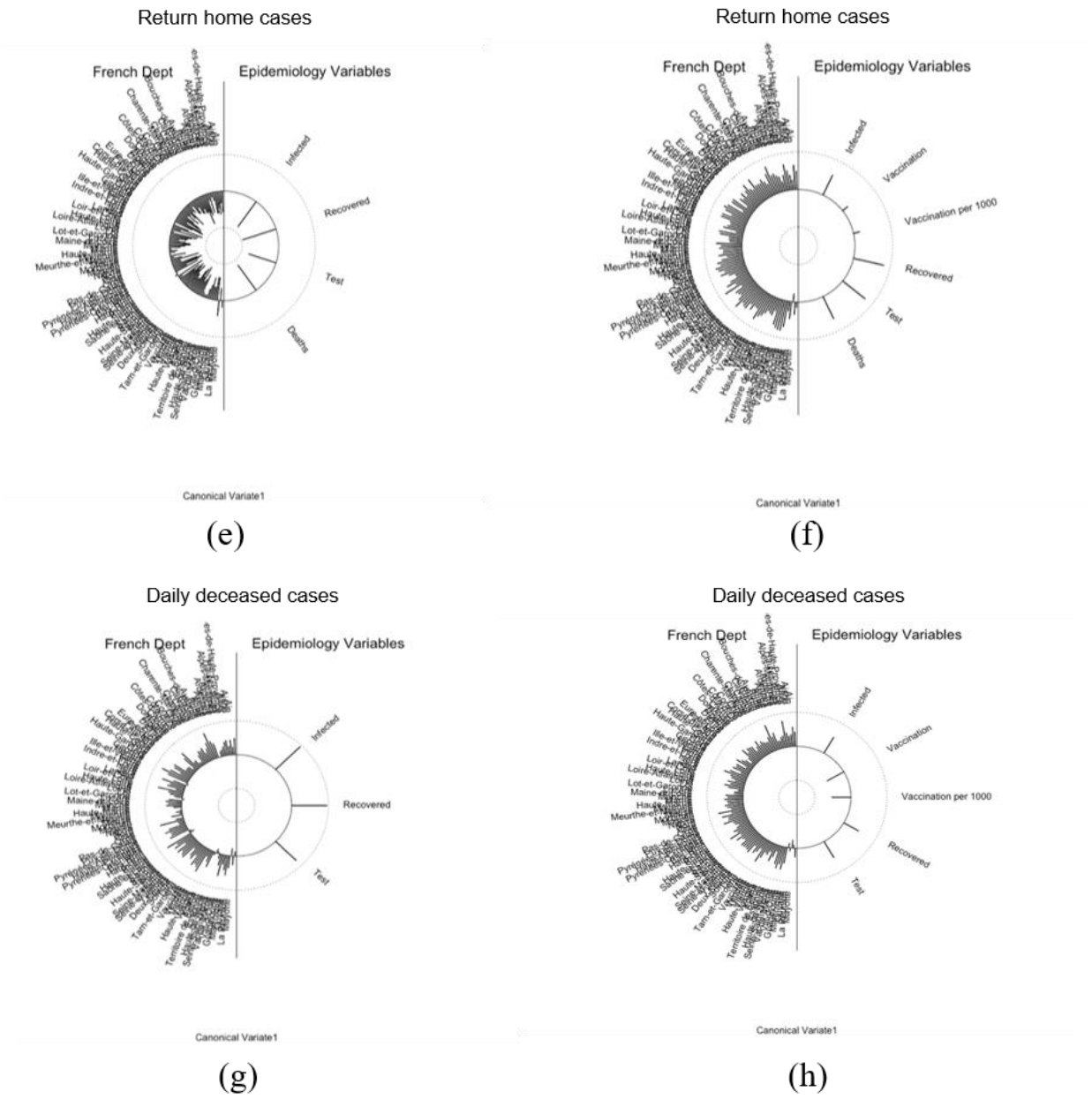
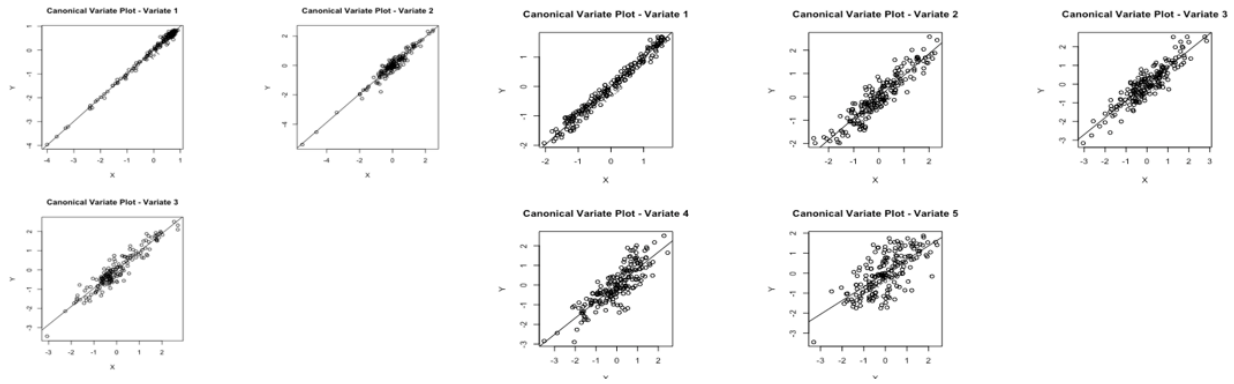
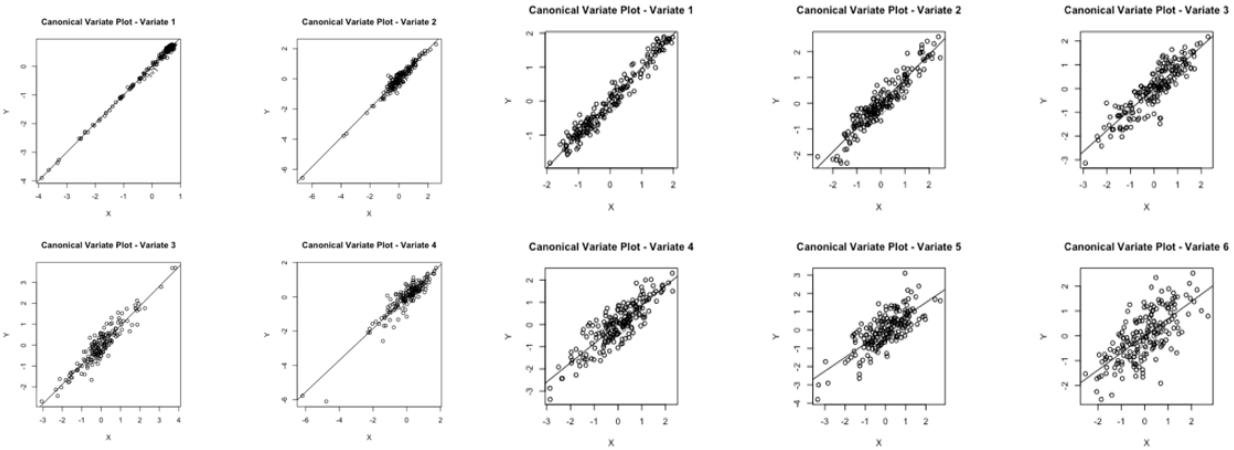


Figure 7. Helio plot for the correlation of French departments for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination started, (e) daily return home, (f) daily return home when vaccination started, (g) daily deceased and (h) daily deceased when vaccination started.

The helio plot in Figure 7 depicts the relationships between the different departments in France, as well as the epidemiology variables and control measures (vaccination). Figure 7a,c,e showed a negative correlation between the epidemiology variables and the hospitalization, ICU, and daily return home cases across departments, whereas Figure 7g showed a positive correlation between the epidemiology variables and the deceased cases across departments. Figure 7b,d,f confirm the effect of vaccination on the number of hospitalizations, ICU, and daily return home during vaccination, demonstrating a positive correlation and the effect of this control measure, whereas Figure 7h shows a negative correlation, indicating a negative relationship between the deceased and the vaccination

introduced. The results presented in Figure 8 show the linear relations in the scatter plot as most of the variables show 95% significance level and from Table 3 there is high correlations between the variables considered. The Figure A5 (supplementary material) presents the redundancy between the canonical variates. We tested the canonical correlation, and the null hypothesis is when the canonical correlation is equal to zero. Figure 8a shows hospitalized cases with p-value<0.05 for all canonical variate, Figure 8b shows hospitalized when vaccination has started with p-value<0.05 except the last canonical variate with p-value=0.88, Figure 8c shows ICU cases with p-value<0.05 for all canonical variate, Figure 8d shows ICU cases when vaccination has started with p-value<0.05 except the last two Canonical variate with p-value=0.68 and p-value=0.87 respectively, Figure 8e shows daily return home with p-value<0.05 for all canonical variate, Figure 8f shows daily return home when vaccination has started with p-value<0.05 except the last two canonical variate with p-value=0.14 and p-value=0.34 respectively, Figure 8g shows daily deceased with p-value<0.05 except the last canonical variate with p-value=0.08 and Figure 8h shows daily deceased when vaccination has started with p-value<0.05 except the last two canonical variate with p-value=0.08 and p-value=0.46 respectively.





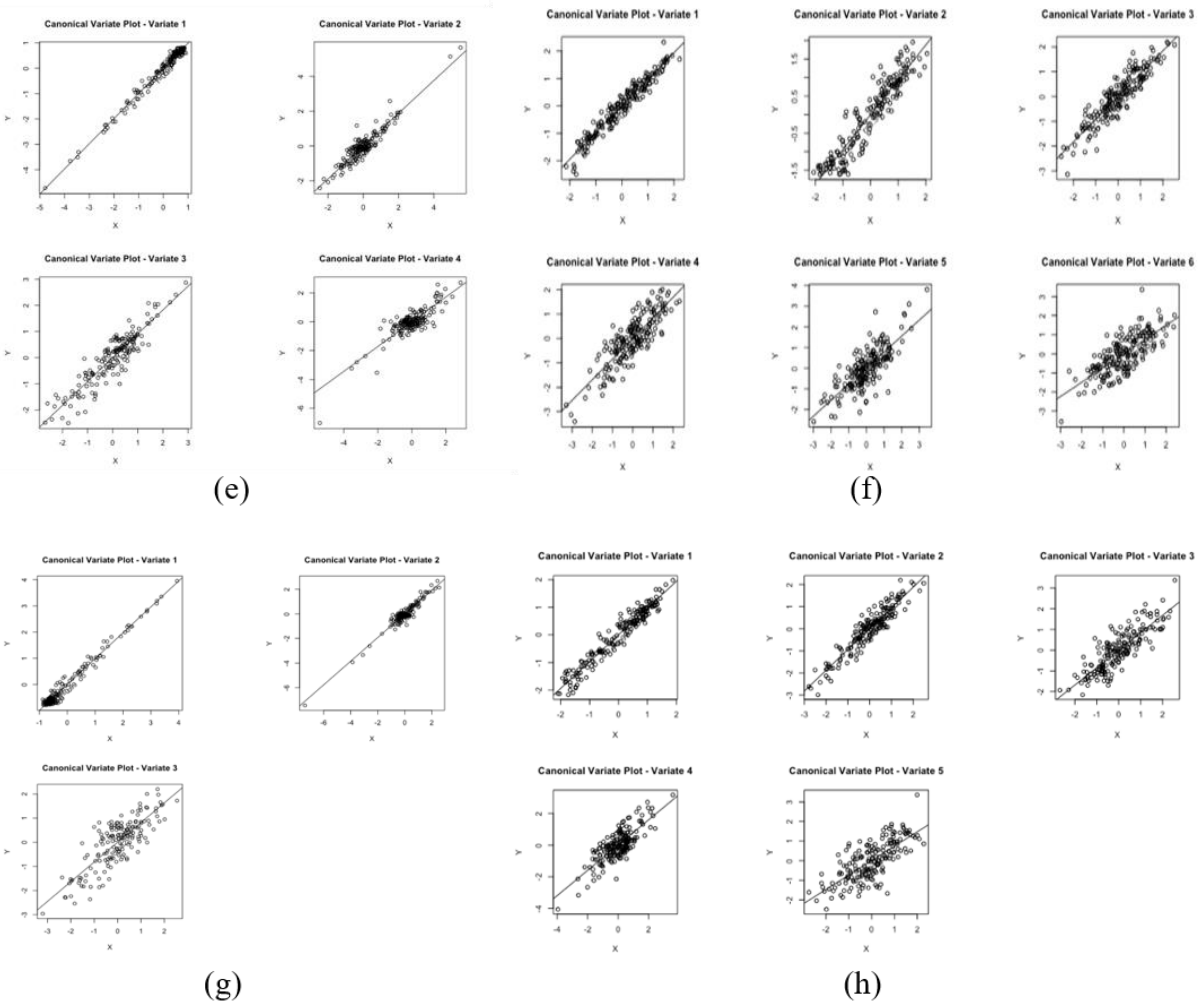


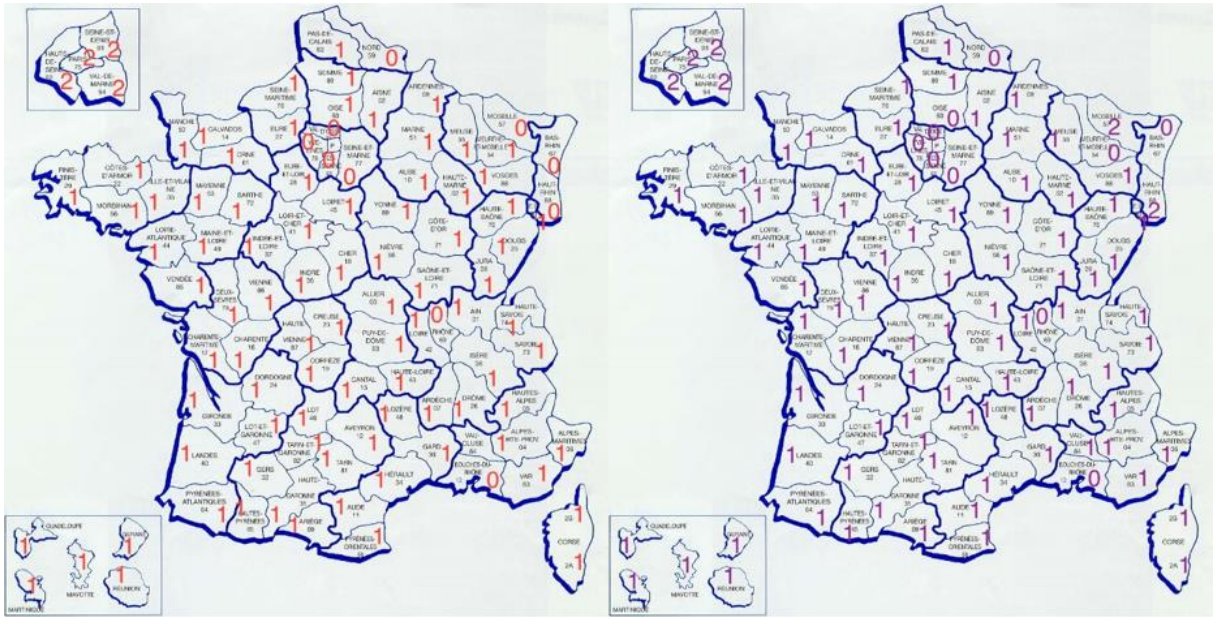
Figure 8. Canonical correlation visualization of (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

Table 3. Canonical correlation scores.

	Before vaccination started				After vaccination has started			
	Hospital	ICU	Daily return	Daily death	Hospital	ICU	Daily return	Daily death
Deaths	0.996	0.926	0.859	-	0.989	0.689	0.745	-
Recovered	0.970	0.973	0.941	0.961	0.838	0.865	0.846	0.797
Test	0.950	0.937	0.911	0.816	0.685	0.750	0.776	0.736
Vaccination	-	-	-	-	0.924	0.942	0.939	0.936
Infected	-	0.998	0.992	0.987	-	0.980	0.971	0.970
Vaccination	-	-	-	-	0.901	0.885	0.917	0.841

5. Clustering method

The clustering of functional data is one method that statisticians are always interested in and in this section we used the K-means and fuzzy K-means techniques whose algorithm is already in Python *skfd.ml.clustering and FuzzyCMeans.* These methods will enable us to visualize how various departments are clustered based on our functional data and to give it the best interpretation based on their geographical location. The basic function used for the K-means clustering is a B-spline and results of our clusters are presented below. We present the result in the cluster form and also on the map of France with indication of the membership to the 3 (i.e., K=3) clusters (0, 1 or 2) to get a clearer view of the result. Even if we increase the value of K, it won't improve the results as K is optimal. We only presented the result for two cases (daily hospitalized and daily deceased) for the period before vaccination begins in France and two cases (daily return and ICU cases) for the period when vaccination has started in France. In Figure 9a–d we present the clusters (0, 1 or 2) that each French department belongs to. The result clustered French departments outside metropole or the French hexagon to the same clusters which of course are not binded with mitigation measures and rules used in departments within France [55]. Also, departments close to Paris are in the same cluster which is the same with departments having the same trend of the pandemic prevalence as presented in Figure 9a–d. In Figure 9e,g we observed same pattern for the data points in the clustering which means that the same way we have more hospitalization cases before vaccination, we also have the more people returning home during vaccination which affirm the fact that the vaccination campaign in France has helped to mitigate the spread of the disease. Figure 9f,h is a bit tricky because of the pattern they followed but it is not surprising that we have more deceased in most departments before vaccination period and less ICU cases during the vaccination period. These patterns of having more cluster points attaining the highest peak in PC 1 is distinct in the results presented.



(a) (b)

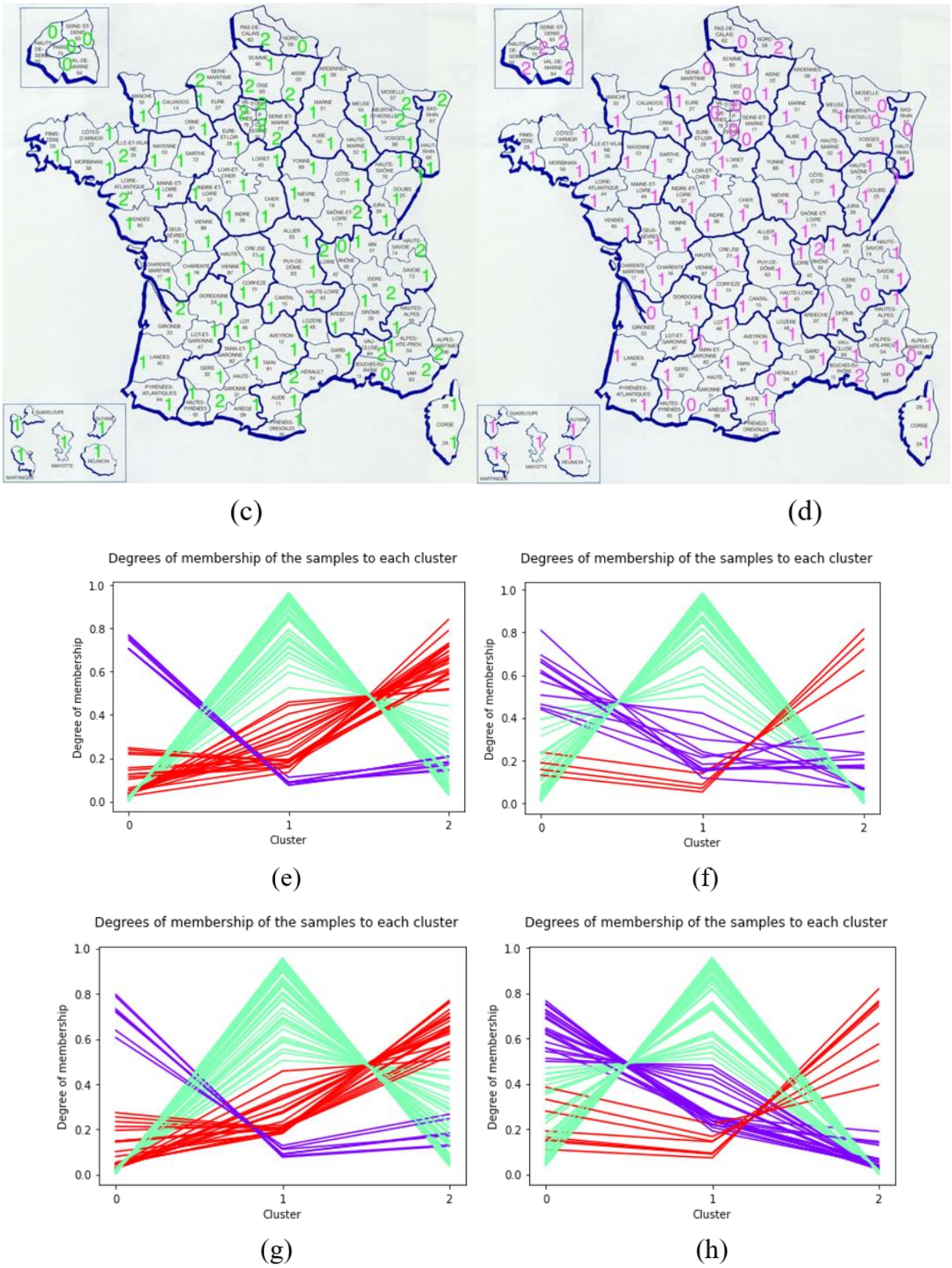


Figure 9. Clustering of all departments in France using K-means: (a) hospitalized, (b) daily deceased, (c) daily return home when vaccination has started and (d) ICU cases when vaccination has started, and fuzzy K-means, (e) clusters for hospitalized, (f) clusters for daily deceased, (g) clusters for daily return home when vaccination has started and (h) clusters for ICU when vaccination has started.

6. Prediction

6.1. Function to scalar linear model

In this section we used functional linear regression model to predict two of our response variables. Let

$$\mathrm{Y}=<\theta,\mathrm{X}>+\epsilon, \tag{16}$$

where θ is the unknown function of the model, X is a functional covariate belonging to some functional space $\mathbb{H}$ which is endowed with an inner product $<.,.>$, Y is the response variable and ϵ is the random error term. Usually, $\mathbb{H}$ is the space $L^2([a,b])$ of square integrable functions on some real compact interval $[a,b]$ and

$$<f,g>=\int_a^b f(t)g(t)dt, \tag{17}$$

is the corresponding inner product, where the functions f, g$\in L^2([a,b])$. Then, we consider C=[0,1], so the Eq (12) can be written as:

$$Y=\int_0^1 \theta(t)X(t)dt\ +\ \epsilon \tag{18}$$

where θ is a square integrable function which is defined on C and ϵ is a random variable such that $\mathbb{E}(\epsilon)=0$ and $\mathbb{E}(X(t)\epsilon)=0$. The Eq (18) can be rewritten as:

$$Y=\Psi(X)+\epsilon, \tag{19}$$

where Ψ represents the integral.

We treated the functional data X (hospitalization) as a curve whose prediction is linked to a scalar Y (number of deaths and tests between 19/03/20–13/09/20 which is the time interval representing $[1,177]$) response variable. The time interval for in which the response was predicted is from 14/09/20–29/10/20 which represent $[178,225]$. The data considered are data before vaccination started in France and we trained 80% of the data and 20% was tested. The visualization of the results is presented in Figure 10 and the tabular form of the numerical results can be found in Table A1 (supplementary material). The prediction affirms the fact that the relaxation in the mitigation measures during the studied period increases the number of deaths and tests in France, the predicted results of deaths being systematically higher than the observed values as seen in Figure 10a and Table A1.

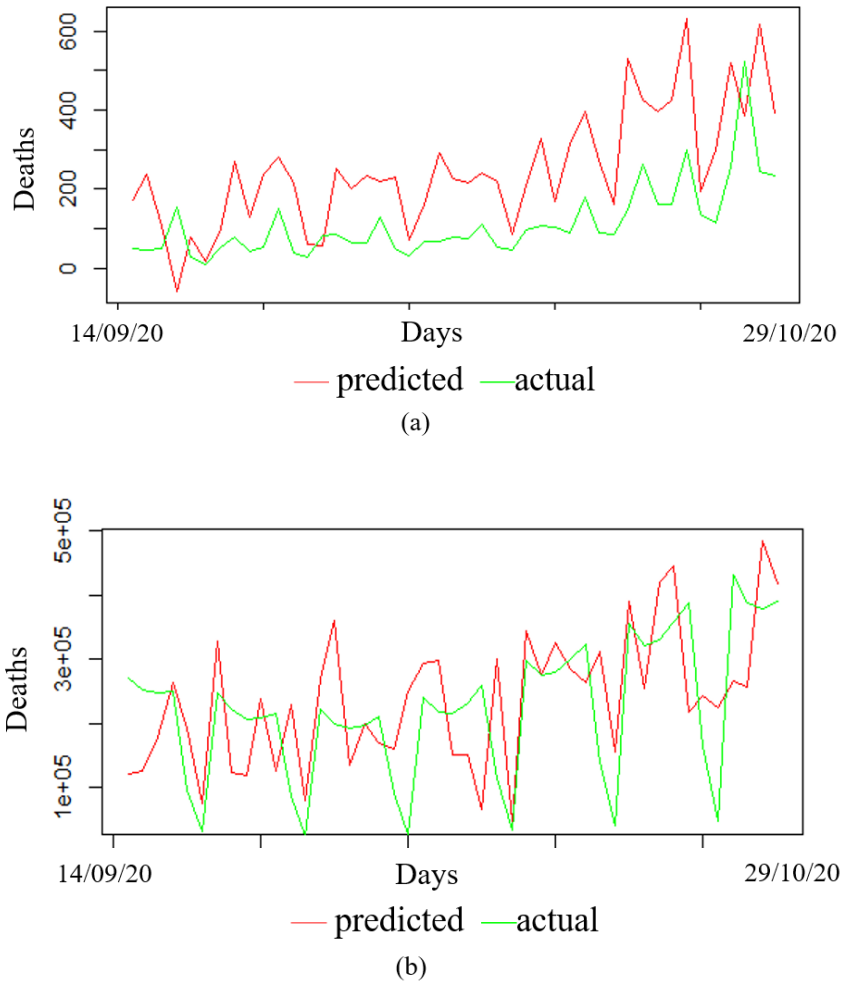


Figure 10. (a) Functional linear regression model prediction for number of deaths in France as response variable before vaccination begins and (b) Functional linear regression model prediction for number of tests in France as response variable before vaccination begins.

6.2. Function-on-function linear model

We consider functional input and output regression model where we treated y(t) as response variable and x(s)' s as predictors at each time t, i.e., $x(t) \rightarrow y(t)$. The functional linear model with an intercept is of the form:

$$y(t) = \beta_o(t) + \int \beta(s,t)x(s)ds + \epsilon(t). \tag{20}$$

We used this method to perform a function-on-function linear regression on our set of functional data by using the functional data curves of 101 days (predictors) to predict another set of curves of 101 days (response) while also estimating the slope $\beta(s,t)$, whose results in considered cases are presented in 3D diagrams of Figure 11.

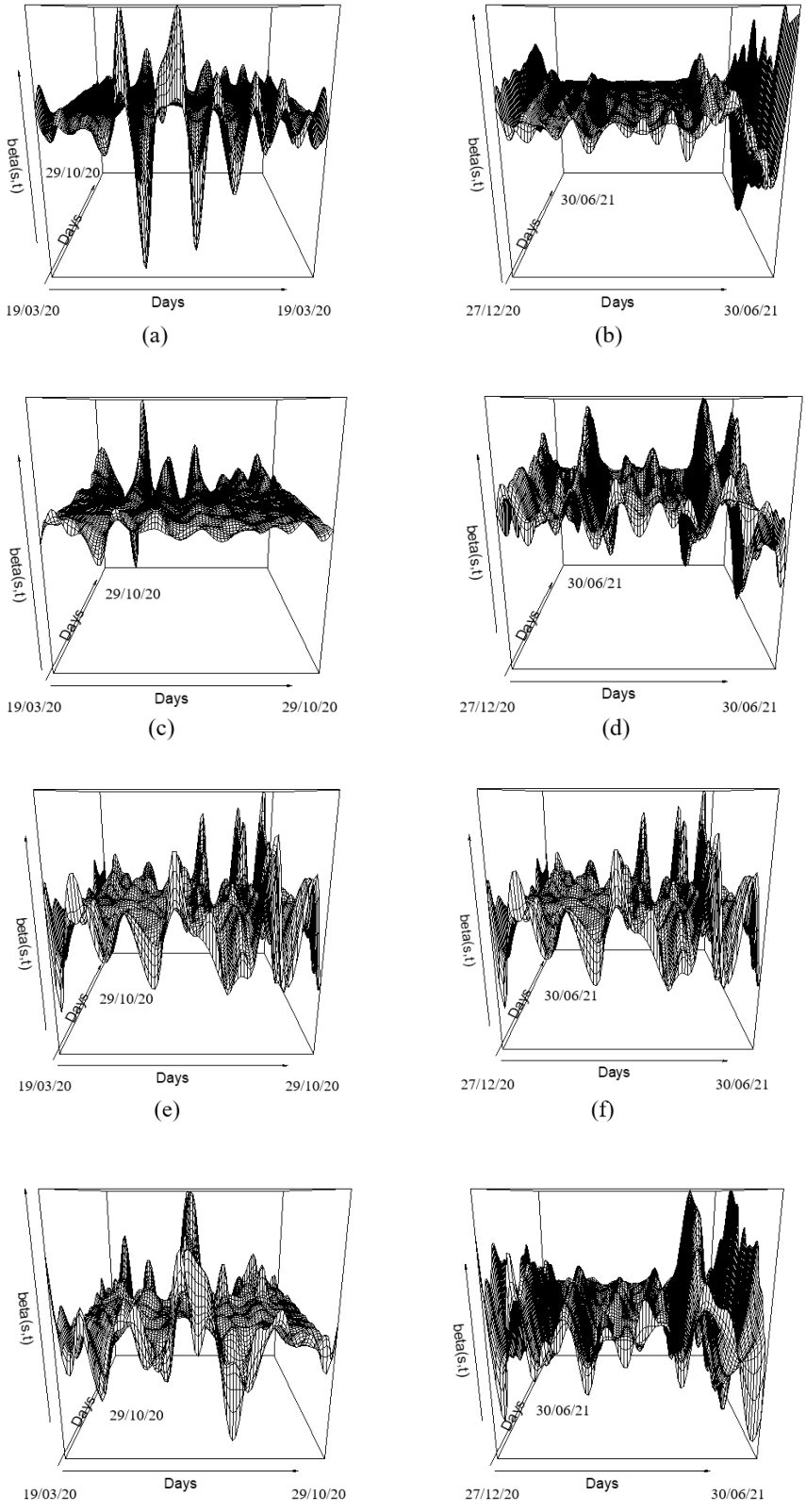


Figure 11. The 3D visualization of function-on-function regression for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.

Figure 11a shows hospitalized cases with the slope on the interval $-2.799063 \leq \beta(s,t) \leq 1.980147$, Figure 11b shows hospitalized when vaccination has started with the slope on the interval $-1.501887 \leq \beta(s,t) \leq 1.076421$, Figure 11c shows ICU cases with the slope on the interval $-1.0733846 \leq \beta(s,t) \leq 0.841100$, Figure 11d shows ICU cases when vaccination has started with the slope on the interval $-0.5646148 \leq \beta(s,t) \leq 0.3661280$, Figure 11e shows daily return home with the slope on the interval $-0.6755000 \leq \beta(s,t) \leq 0.7030529$, Figure 11f shows daily return home when vaccination has started with the slope on the interval $-0.4333295 \leq \beta(s,t) \leq 0.4300995$, Figure 11g shows daily deceased with the slope on the interval $-0.3277864 \leq \beta(s,t) \leq 0.4002531$ and Figure 11h shows daily deceased when vaccination has started with the slope on the interval $-0.3284866 \leq \beta(s,t) \leq 0.3641679$. We observed that in all these Figures in this section, the 3D surfaces yield results whose shapes look roughly similar to the slope curve, functional predictors curve and functional response curve.

7. Perspectives and conclusions

We studied in this article the best way to summarize temporal information relating to the variations of variables linked to the epidemic dynamics of COVID-19, such as hospitalized cases before and after vaccination has started, medical intensive care unit (MICU) cases before and after vaccination, daily return home cases before and after vaccination, and daily deceased before and after vaccination. Using the functional principal component analysis, it was shown that the first functional principal component well summarized the U or W shape observed for the data related to the first three principal components. This discovery confirms the importance of this first component for the explanation and the qualitative prediction from the observed data. The influence of the vaccination is visible, because the U or W shape is attenuated after vaccination, and does not come close to the shapes observed for seasonal influenza [16,44]. The subsequent functional principal components have poor predictive power, but the second component clearly shows the reducing influence of vaccination on all epidemic variables. A further, more in-depth study could undoubtedly show the predictive nature of this second component on the future success of a vaccination policy, by comparing different countries with different vaccination rates and by quantifying the phase of descent of the curves of the second component (for example by its slope at the second inflection and by the value of its minimum). We have given the percentage of variation between actual and predicted values, which quantifies the gain obtained by the mitigation measures during the studied period on the number of deaths in France, systematically higher than the observed values as seen in Figure 8 and Table A1.

Acknowledgments

The authors wish to acknowledge the Petroleum Technology Development Fund (PTDF) Nigeria doctoral fellowship in collaboration with Campus France Africa Unit.

Conflict of interest

Authors declare no conflict of interest.

References

1. D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir, *Histoire de l'Acad., Roy. Sci. (Paris) avec Mem,* 1760, 1–45.
2. D. A. Henderson, The eradication of smallpox-An overview of the past, present, and future, *Vaccine*, **29** (2011), D7–D9. https://doi.org/10.1016/j.vaccine.2011.06.080
3. D. Wujastyk, Medicine in India, In: J. van Alphen, A. Aris, F. Meyer, M. de Fraeye, *Oriental medicine: An illustrated guide to the Asian arts of healing*, London: Serindia Publications, 1995, 19–38.
4. A. M. Silverstein, *A history of immunology*, 2 Eds., London: Academic Press, 2009, 293.
5. L. S. Benjamin, L Melville, *Lady Mary Wortley Montagu, her life and letters (1689–1762)*, Hutchinson, London, 1925.
6. R. Ross, An application of the theory of probabilities to the study of a priori pathometry-part I, *Proc. R. Soc. Ser. A*, **92** (1916), 204–230. https://doi.org/10.1098/rspa.1916.0007
7. A. G. McKendrick, Applications of mathematics to medical problems, *Proc. Edinburgh Math. Soc.*, **44** (1925), 98–130. https://doi.org/10.1017/S0013091500034428
8. J. Gaudart, O. Touré, N. Dessay, A. L. Dicko, S. Ranque, L. Forest, et al., Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, *Mali. Malaria J.*, **8** (2009), 61. https://doi.org/10.1186/1475-2875-8-61
9. J. Gaudart, M. Ghassani, J. Mintsa, M. Rachdi, J. Waku, J. Demongeot, Demography and diffusion in epidemics: Malaria and black death spread, *Acta Biotheor.*, **58** (2010), 277–305. https://doi.org/10.1007/s10441-010-9103-z
10. J. Demongeot, J. Gaudart, A. Lontos, E. Promayon, J. Mintsa, M. Rachdi, Zero-diffusion domains in reaction-diffusion morphogenetic and epidemiologic processes, *Int. J. Bifurcation Chaos*, **22** (2012), 1250028. https://doi.org/10.1142/S0218127412500289
11. J. Demongeot, J. Gaudart, J. Mintsa, M. Rachdi, Demography in epidemics modelling, *Commun. Pure Appl. Anal.*, **11** (2012), 61–82. http://dx.doi.org/10.3934/cpaa.2012.11.61
12. Z. Liu, P. Magal, O. Seydi, G. Webb, Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and importance of major public health interventions, *Biology*, **9** (2020), 50. https://doi.org/10.3390/biology9030050
13. J. Demongeot, Q. Griette, P. Magal, SI epidemic model applied to COVID-19 data in mainland China, *Royal Soc. Open Sci.*, **7** (2020), 201878. https://doi.org/10.1098/rsos.201878
14. Z. Liu, P. Magal, O. Seydi, G. Webb, Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data, *Math. Biosci. Eng.*, **17** (2020), 3040–3051. https://doi.org/10.3934/mbe.2020172
15. P. Magal, O. Seydi, G. Webb, Y. Wu, A model of vaccination for Dengue in the Philippines 2016–2018, *Front. Appl. Math. Stat.*, **7** (2021), 760259. https://doi.org/10.3389/fams.2021.760259
16. K. Oshinubi, M. Rachdi, J. Demongeot, Modelling of COVID-19 pandemic vis-à-vis some socioeconomic factors, *Front. Appl. Math. Stat.*, **7** (2021), 786983.
17. COVID-19 coronavirus pandemic, 2021. Available from: https://www.worldometers.info/coronavirus.
18. Données hospitalières relatives à l'épidémie de COVID-19, 2021. Available from: https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-Covid-19.
19. Live COVID-19 vaccination tracker. Available from: https://covidvax.live/location/fra.

20. F. Ferraty, P. Vieu, *Nonparametric functional data analysis*, New York: Springer, 2006. https://doi.org/10.1007/0-387-36620-2
21. J. D. Tucker, *Functional component analysis and regression using elastic methods*, PhD. Thesis, Florida State University, 2014.
22. J. D. Tucker, W. Wu, A. Srivastava, Generative models for functional data using phase and amplitude separation, *Comput. Stat. Data Anal.*, **61** (2013), 50–66. https://doi.org/10.1016/j.csda.2012.12.001
23. J. O. Ramsay, B. W. Silverman, *Applied functional data analysis: Methods and case studies*, New York: Springer, 2002. https://doi.org/10.1007/b98886
24. A. Srivastava, E. P. Klassen, Functional data and elastic registration, In: *Functional and shape data analysis*, New York: Springer, 2016, 73–123. https://doi.org/10.1007/978-1-4939-4020-2_4
25. J. O. Ramsay, G. Hooker, S. Graves, *Functional data analysis with R and MATLAB*, New York: Springer, 2009. https://doi.org/10.1007/978-0-387-98185-7
26. C. Tang, T. Wang, P. Zhang, Functional data analysis: An application to COVID-19 data in the United States, *arXiv*. Available from: https://arxiv.org/abs/2009.08363.
27. C. Acal, M. Escabias, A. M. Aguilera, M. J. Valderrama, COVID-19 data imputation by multiple function-on-function principal component regression, *Mathematics*, **9** (2021), 1237. https://doi.org/10.3390/math9111237
28. T. Boschi, J. Di Iorio, L. Testa, M. A. Cremona, F. Chiaromonte, Functional data analysis characterizes the shapes of the first COVID-19 epidemic wave in Italy, *Sci. Rep.*, **11** (2021), 17054. https://doi.org/10.1038/s41598-021-95866-y
29. Q. Griette, J. Demongeot, P. Magal, A robust phenomenological approach to investigate COVID-19 data for France, *Math. Appl. Sci. Eng.*, **2** (2021), 149–218. https://doi.org/10.5206/mase/14031
30. Q. Griette, J. Demongeot, P. Magal, What can we learn from COVID-19 data by using epidemic models with unidentied infectious cases, *Math. Biosci. Eng.*, **2** (2021), 149–160. http://dx.doi.org/10.2139/ssrn.3868852
31. J. Gaudart, J. Landier, L. Huiart, E. Legendre, L. Lehot, M. K. Bendiane, et al., Factors associated with spatial heterogeneity of Covid-19 in France: A nationwide ecological study, *Lancet Public Health*, **6** (2021), 222–231. https://doi.org/10.1016/s2468-2667(21)00006-2
32. O. D. Ilie, R. O. Cojocariu, A. Ciobica, S. I. Timofte, I. Mavroudis, B. Doroftei, Forecasting the spreading of COVID-19 across nine countries from Europe, Asia, and the American continents using the ARIMA models, *Microorganisms*, **8** (2020), 1158. https://doi.org/10.3390/microorganisms8081158
33. J. Stojanovic, V. G. Boucher, J. Boyle, J. Enticott, K. L. Lavoie, S. L. Bacon, COVID-19 is not the flu: Four graphs from four countries, *Front. Public Health*, 2021, 628479. https://doi.org/10.3389/fpubh.2021.628479
34. C. Carroll, S. Bhattacharjee, Y. Chen, P. Dubey, J. Fan, A. Gajardo, et al., Time dynamics of COVID-19, *Sci. Rep.*, **10** (2020), 21040. https://doi.org/10.1038/s41598-020-77709-4
35. A. Srivastava, G. Chowell, Modeling study: Characterizing the spatial heterogeneity of the COVID-19 pandemic through shape analysis of epidemic curves, *Res. Square*, 2021, 1–27. https://doi.org/10.21203/rs.3.rs-223226/v1
36. J. Demongeot, Y. Flet-Berliac, H. Seligmann, Temperature decreases spread parameters of the new COVID-19 cases dynamics, *Biology*, **9** (2020), 94. https://doi.org/10.3390/biology9050094
37. H. Seligmann, S. Iggui, M. Rachdi, N. Vuillerme, J. Demongeot, Inverted covariate effects for mutated 2nd vs 1st wave COVID-19: High temperature spread biased for young, *Biology*, **9** (2020), 226. https://doi.org/10.1101/2020.07.12.20151878

38. S. Soubeyrand, J. Demongeot, L. Roques, Towards unified and real-time analyses of outbreaks at country-level during pandemics, *One Health*, **11** (2020), 100187. https://doi.org/10.1016/j.onehlt.2020.100187
39. J. Demongeot, H. Seligmann, SARS-CoV-2 and miRNA-like inhibition power, *Med. Hypotheses*, **144** (2020), 110245. https://doi.org/10.1016/j.mehy.2020.110245
40. H. Seligmann, N. Vuillerme, J. Demongeot, Unpredictable, counter-intuitive geoclimatic and demographic correlations of COVID-19 spread rates, *Biology*, **10** (2021), 623. https://doi.org/10.3390/biology10070623
41. K. Oshinubi, F. Al-Awadhi, M. Rachdi, J. Demongeot, Data analysis and forecasting of COVID-19 pandemic in Kuwait, *MedRxiv*, 2021, 1–17. https://doi.org/10.1101/2021.07.24.21261059
42. J. Demongeot, K. Oshinubi, M. Rachdi, L. Hobbad, M. Alahiane, S. Iggui, et al., The application of ARIMA model to analyze COVID-19 incidence pattern in several countries, *J. Math. Comput. Sci.*, **12** (2022), 1–23. https://doi.org/10.28919/jmcs/6541
43. K. Oshinubi, M. Rachdi, J. Demongeot, Analysis of reproduction number R_0 of COVID-19 using current health expenditure as gross domestic product percentage (CHE/GDP) across countries, *Healthcare*, **9** (2021), 1247. https://doi.org/10.3390/healthcare9101247
44. J. Demongeot, K. Oshinubi, M. Rachdi, H. Seligmann, F. Thuderoz, J. Waku, Estimation of daily reproduction rates in COVID-19 outbreak, *MedRxiv*, **9** (2021), 109. https://doi.org/10.1101/2020.12.30.20249010
45. J. Demongeot, A. Laksaci, F. Madani, M. Rachdi, Functional data: Local linear estimation of the conditional density and its application, *Statistics*, **47** (2013), 26–44. https://doi.org/10.1080/02331888.2011.568117
46. M. Rachdi, A. Laksaci, J. Demongeot, A. Abdali, F. Madani, Theoretical and practical aspects on the quadratic error in the local linear estimation of the conditional density for functional data, *Comput. Stat. Data Anal.*, **73** (2014), 53–68. https://doi.org/10.1016/j.csda.2013.11.011
47. J. Demongeot, A. Laksaci, M. Rachdi, S. Rahmani, On the local linear modelization of the conditional distribution for functional data, *Sankhya A*, **76** (2014), 328–355. https://doi.org/10.1007/s13171-013-0050-z
48. J. Demongeot, A. Hamie, A. Laksaci, M. Rachdi, Relative-error prediction in nonparametric functional statistics: Theory and practice, *J. Multivar. Anal.*, **146** (2016), 261–268. https://doi.org/10.1016/j.jmva.2015.09.019
49. J. Demongeot, A. Laksaci, A. Naceri, M. Rachdi, Local linear regression modelization when all variables are curves, *Stat. Probab. Lett.*, **121** (2017), 37–44. https://doi.org/10.1016/j.spl.2016.09.021
50. A. Belkis, J. Demongeot, A. Laksaci, M. Rachdi, Functional data analysis: Estimation of the relative error in functional regression under random left-truncation, *J. Nonparametr. Stat.*, **30** (2018), 472–490. https://doi.org/10.1080/10485252.2018.1438609
51. A. Henien, L. Ait-Hennani, J. Demongeot, A. Laksaci, M. Rachdi, Heteroscedasticity test when the covariables are functionals, *C. R. Math.*, **356** (2018), 571–574. https://doi.org/10.1016/j.crma.2018.02.010
52. J. Demongeot, O. Hansen, H. Hessami, A. S. Jannot, J. Mintsa, M. Rachdi, et al., Random modelling of contagious diseases, *Acta Biotheor.*, **61** (2013), 141–172. https://doi.org/10.1007/s10441-013-9176-6
53. C. J. Rhodes, L. Demetrius, Evolutionary entropy determines invasion success in emergent epidemics, *PLoS One*, **5** (2010), e12951. https://doi.org/10.1371/journal.pone.0012951

54. S. Triambak, D. P. Mahapatra, A random walk Monte Carlo simulation study of COVID-19-like infection spread, *Physica A: Stat. Mech. Appl.*, **574** (2021), 126014. https://doi.org/10.1016/j.physa.2021.126014
55. Wikipedia, Available online: https://www.wikipedia.org/wiki/Departments_of_France.

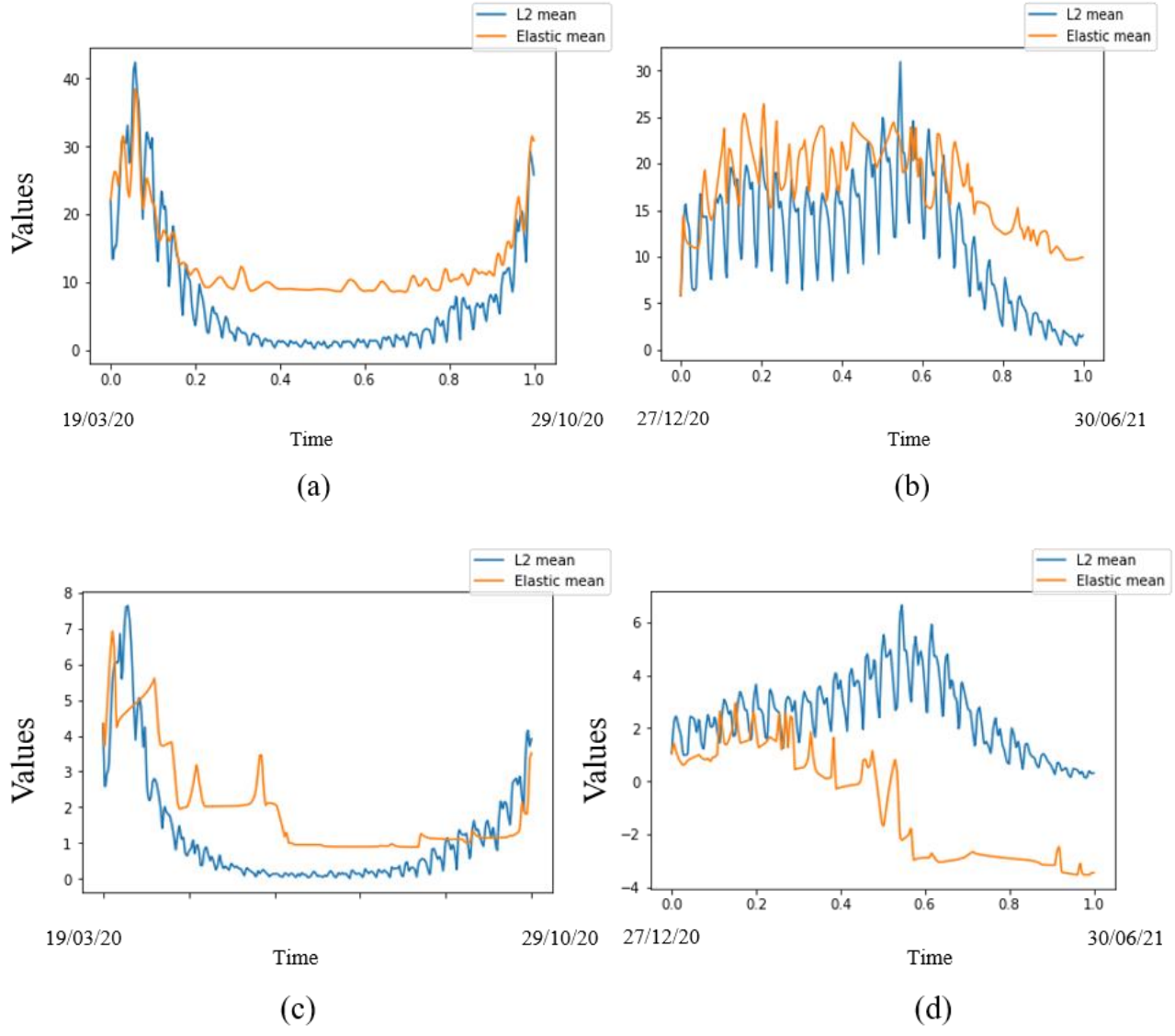
Supplementary

Table A1. Comparison between the predicted and actual values.

Day	Deaths		Tests	
	Actual/%gain	Predicted	Actual	Predicted
1	49/43%	86	269886	120499
2	46/66%	136	251301	126261
3	50/42%	86	248354	178764
4	153/-32%	116	248910	264431
5	26/68%	82	96177	192823
6	11/72%	39	30345	75319
7	53/45%	97	247760	328075
8	78/42%	136	222942	124271
9	43/65%	124	206626	118126
10	52/70%	171	207651	237759
11	150/14%	175	214336	124387
12	39/69%	126	86361	229623
13	27/36%	42	23804	78612
14	81/-12%	72	223293	268357
15	85/51%	172	199948	360891
16	63/62%	168	191917	135525
17	63/74%	246	196259	199318
18	130/11%	146	210495	168640
19	49/71%	167	90639	160399
20	32/56%	59	25699	249580
21	69/62%	181	240612	294387
22	66/64%	185	217585	298154
23	80/57%	188	214258	150142
24	76/61%	194	231306	150122
25	109/50%	219	259073	64680
26	54/74%	209	114369	301318
27	46/41%	78	32368	45122
28	95/52%	198	299121	343681
29	108/63%	292	276013	274313
30	104/37%	165	279376	325882
31	88/68%	275	301465	284502
32	178/44%	320	322468	262879
33	89/63%	238	140298	312368

Continued on next page

Day	Deaths		Tests	
	Actual/%gain	Predicted	Actual	Predicted
34	85/50%	169	40313	154521
35	146/67%	447	355160	390516
36	262/30%	376	321373	254298
37	163/60%	410	330328	419636
38	162/51%	329	357368	445595
39	298/38%	484	388884	217528
40	137/33%	206	165764	242920
41	116/64%	318	47485	223540
42	257/40%	430	430644	264886
43	523/-63%	320	387569	256737
44	244/55%	548	379590	484870
45	235/32%	348	390099	417031



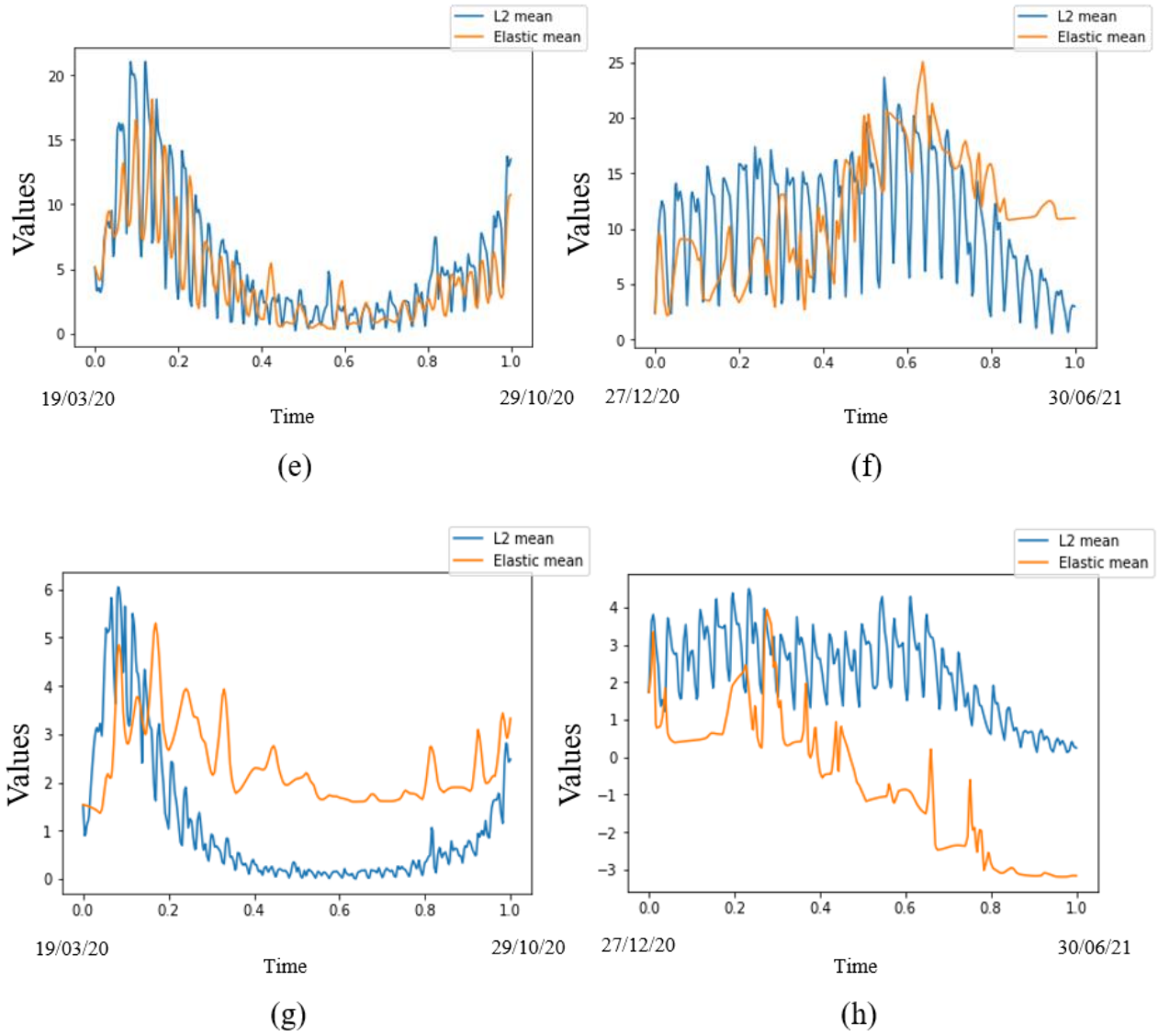
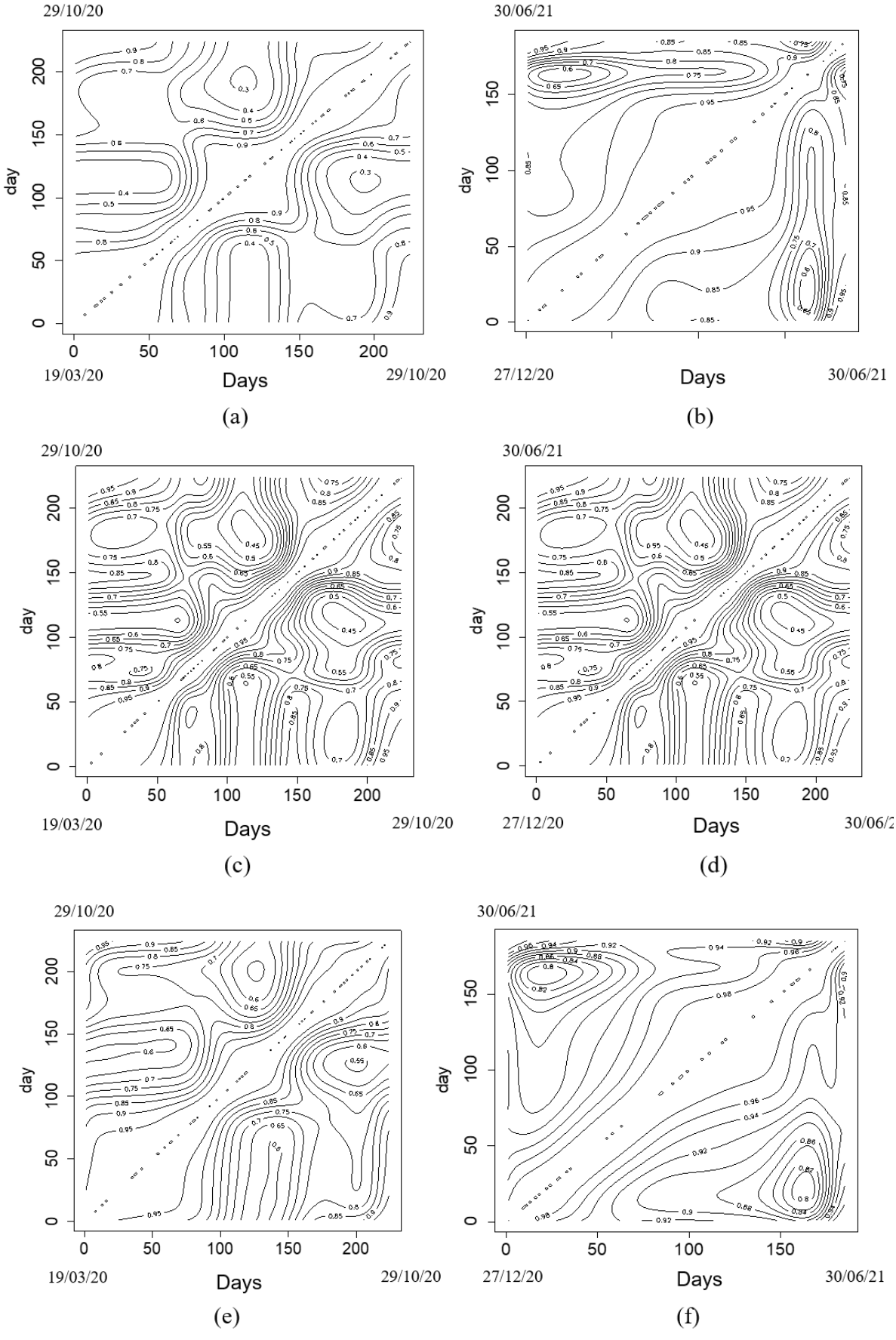


Figure A1. L2 (blue) and elastic (yellow) means of the functional data: (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started. We observed that the elastic mean better captures the geometry of the curves compared to the standard L2 mean for some of the functional data set we considered, since it is not affected by the deformations of the curves. This phenomenon can be seen in (a)–(c) and (e)–(g), but (d) and (h) show a bad shape for elastic mean.


29/10/20
30/06/21
day
Days
19/03/20
29/10/20
27/12/20
30/06/21
(a)
(b)
(c)
(d)
(e)
(f)

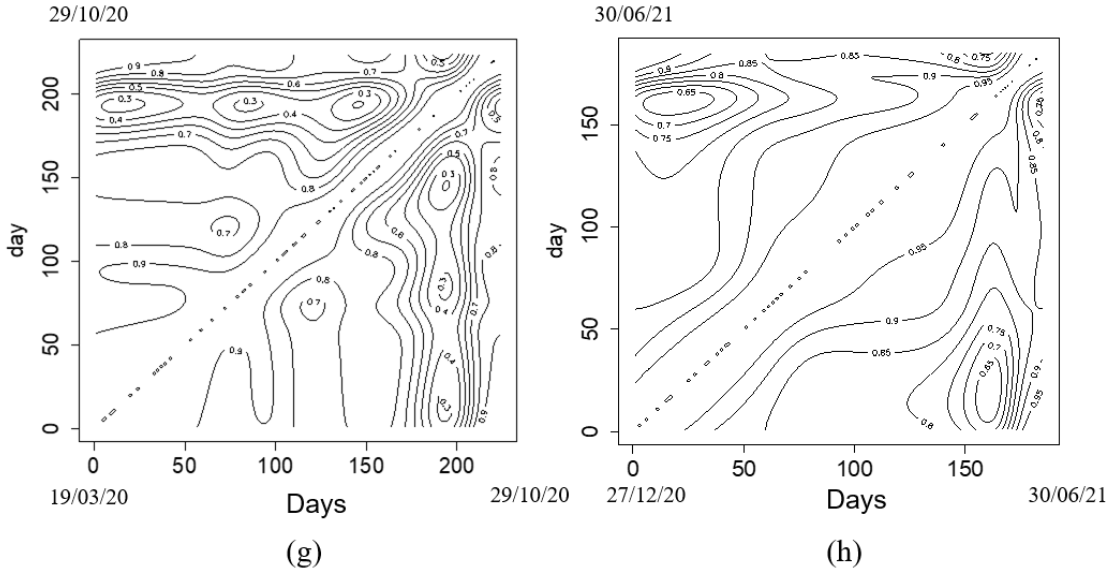
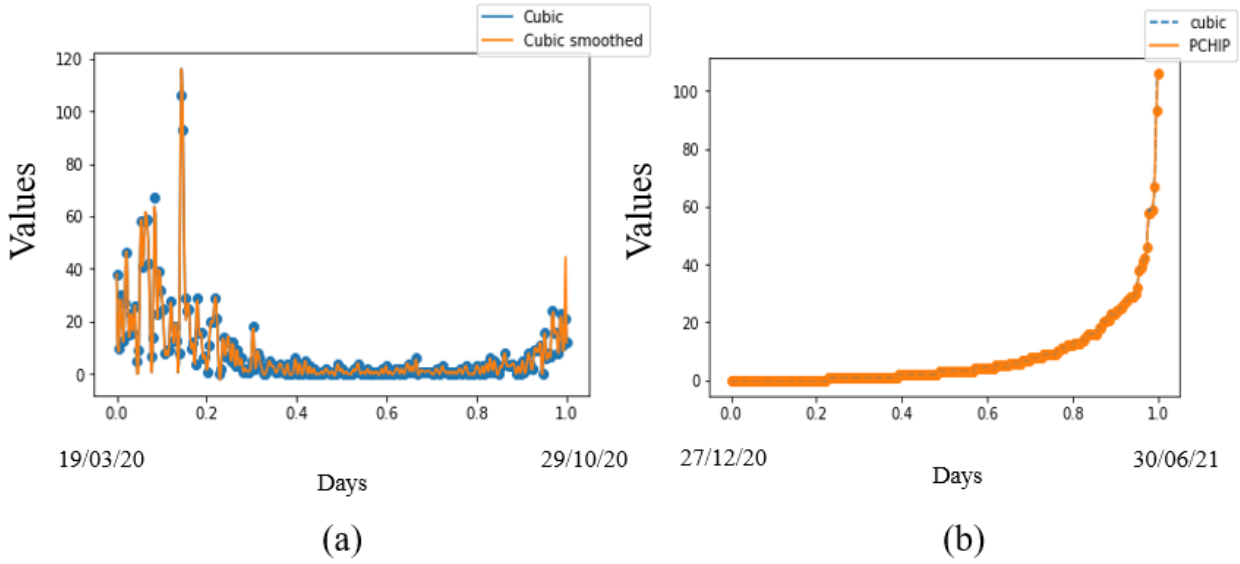


Figure A2. Correlation coefficients between all French departments contour plot. (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.



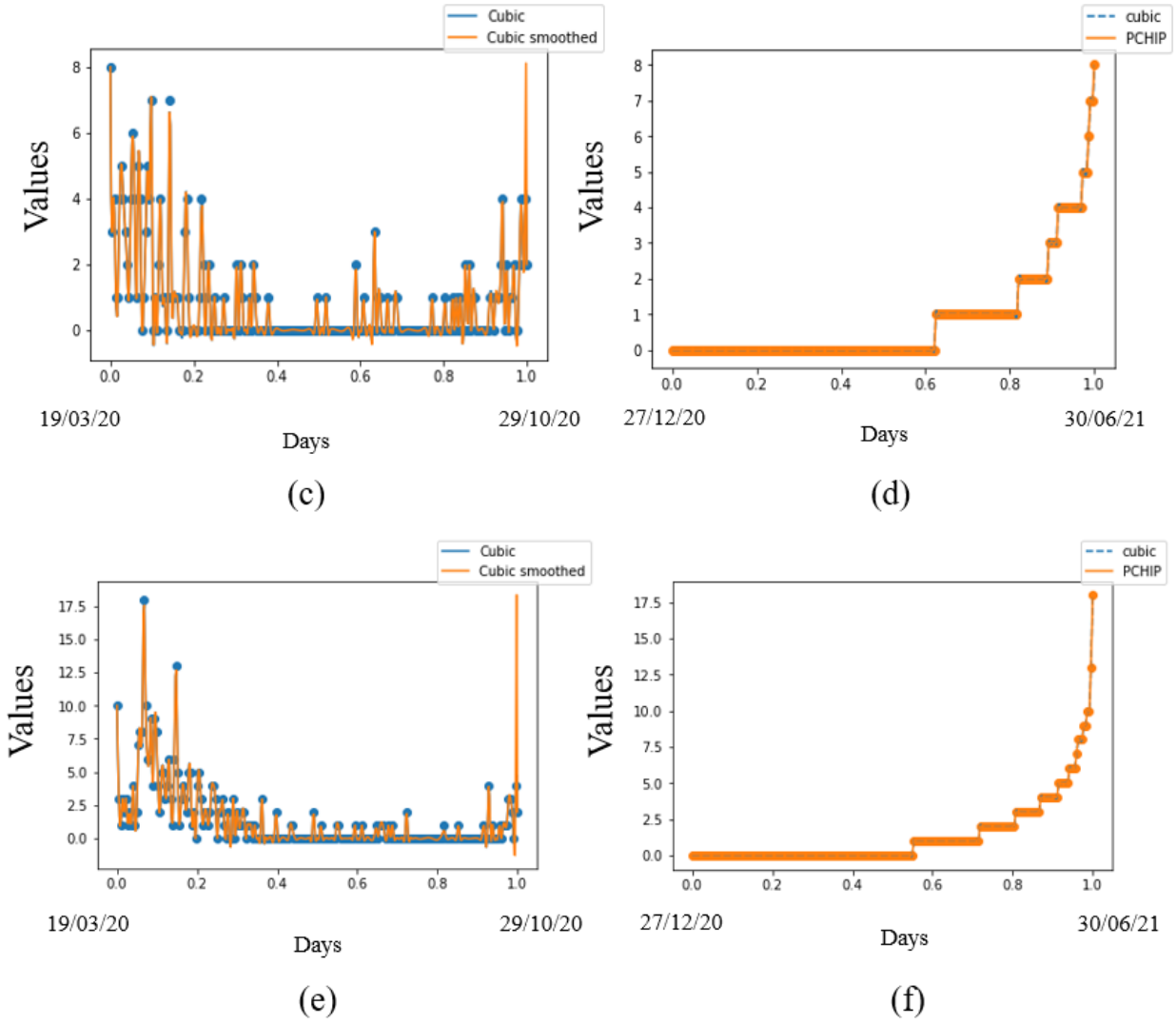


Figure A3. (a) Hospitalized cases interpolation smoothing, (b) hospitalized monotone smoothing, (c) ICU cases interpolation smoothing, (d) ICU cases monotone smoothing, (e) daily deceased interpolation smoothing and (f) daily deceased monotone smoothing. From graphs (a)–(f) one can deduce that cubic spline smoothing curves exhibit oscillations and oscillations are important to know the low and high data thresholds in case the consecutive data points experience a significant change in slope. We also observed that PCHIP is smooth and non-oscillatory despite some sharp increase as the U-shape of the curve deepens.

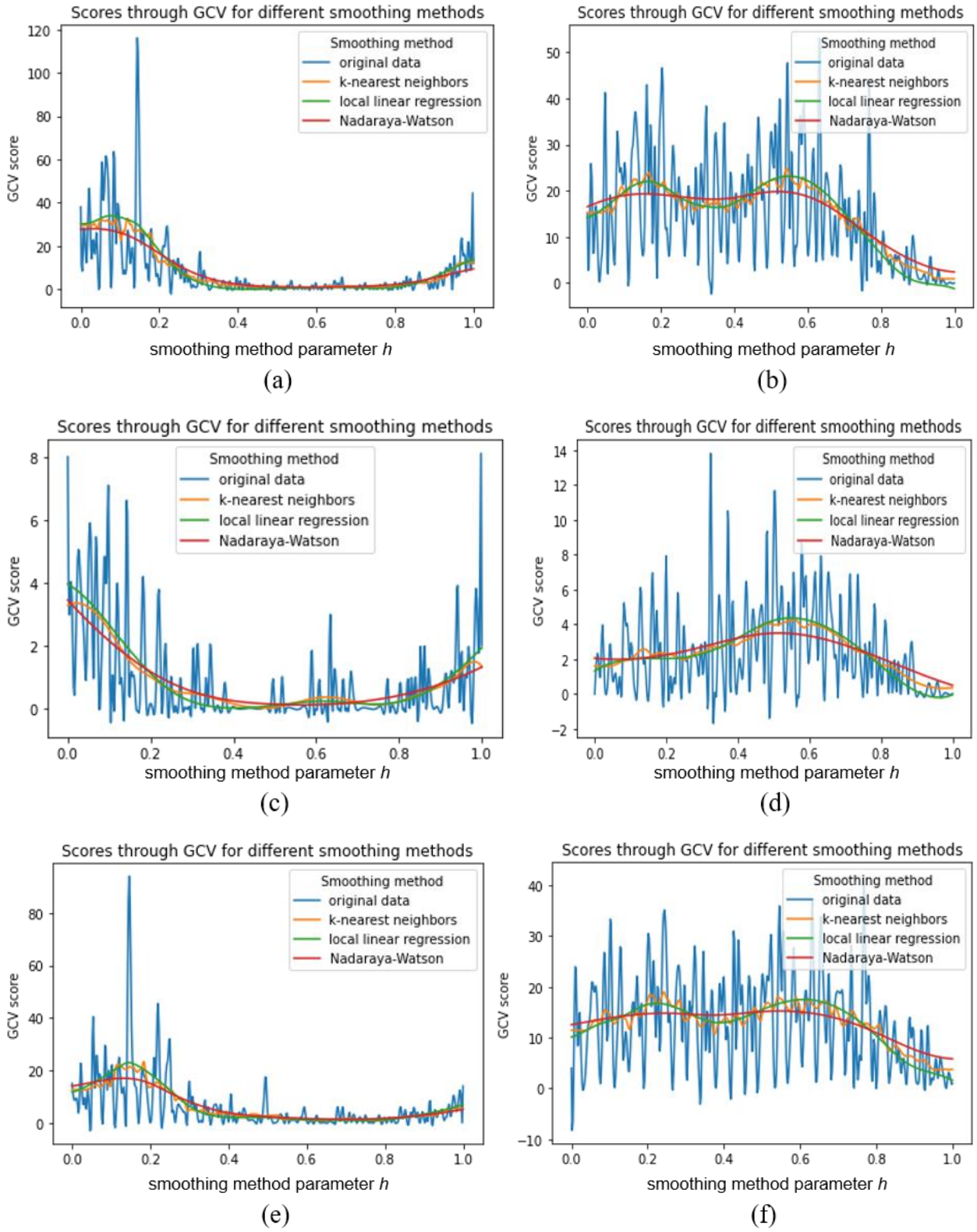
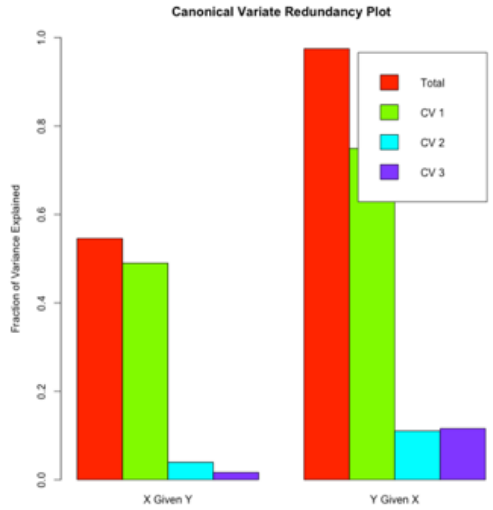
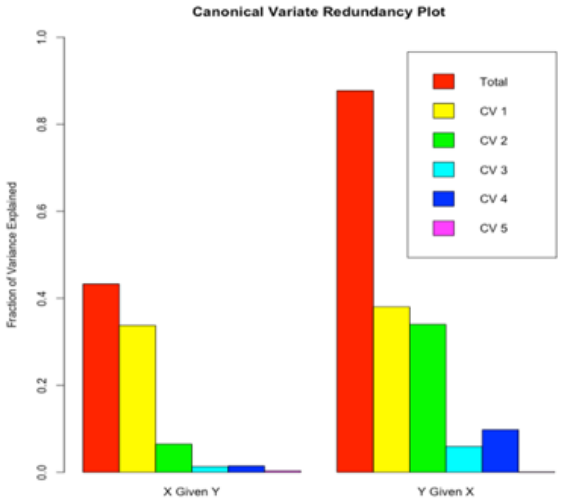


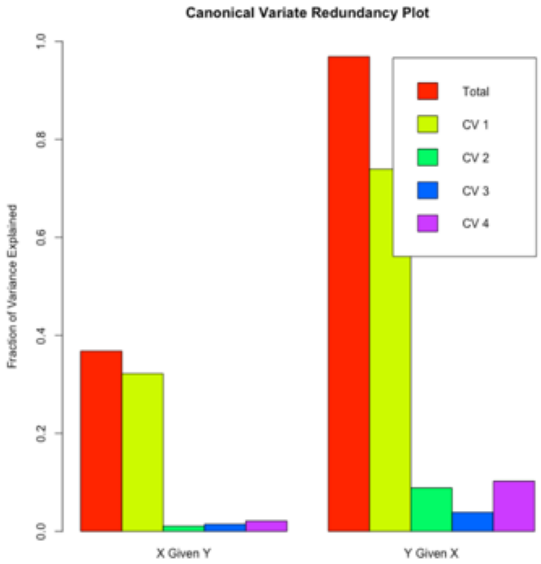
Figure A4. Kernel smoothing method for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home and (f) daily return home when vaccination has started.


Canonical Variate Redundancy Plot
Fraction of Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
Total
CV 1
CV 2
CV 3
X Given Y
Y Given X

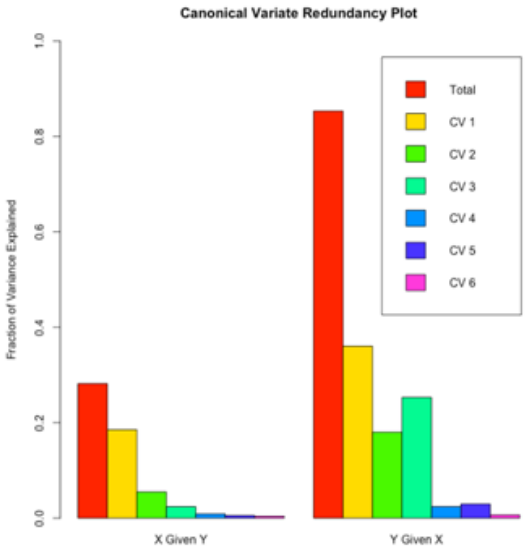
(a)


Canonical Variate Redundancy Plot
Fraction of Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
Total
CV 1
CV 2
CV 3
CV 4
CV 5
X Given Y
Y Given X

(b)


Canonical Variate Redundancy Plot
Fraction of Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
Total
CV 1
CV 2
CV 3
CV 4
X Given Y
Y Given X

(c)


Canonical Variate Redundancy Plot
Fraction of Variance Explained
0.0
0.2
0.4
0.6
0.8
1.0
Total
CV 1
CV 2
CV 3
CV 4
CV 5
CV 6
X Given Y
Y Given X

(d)

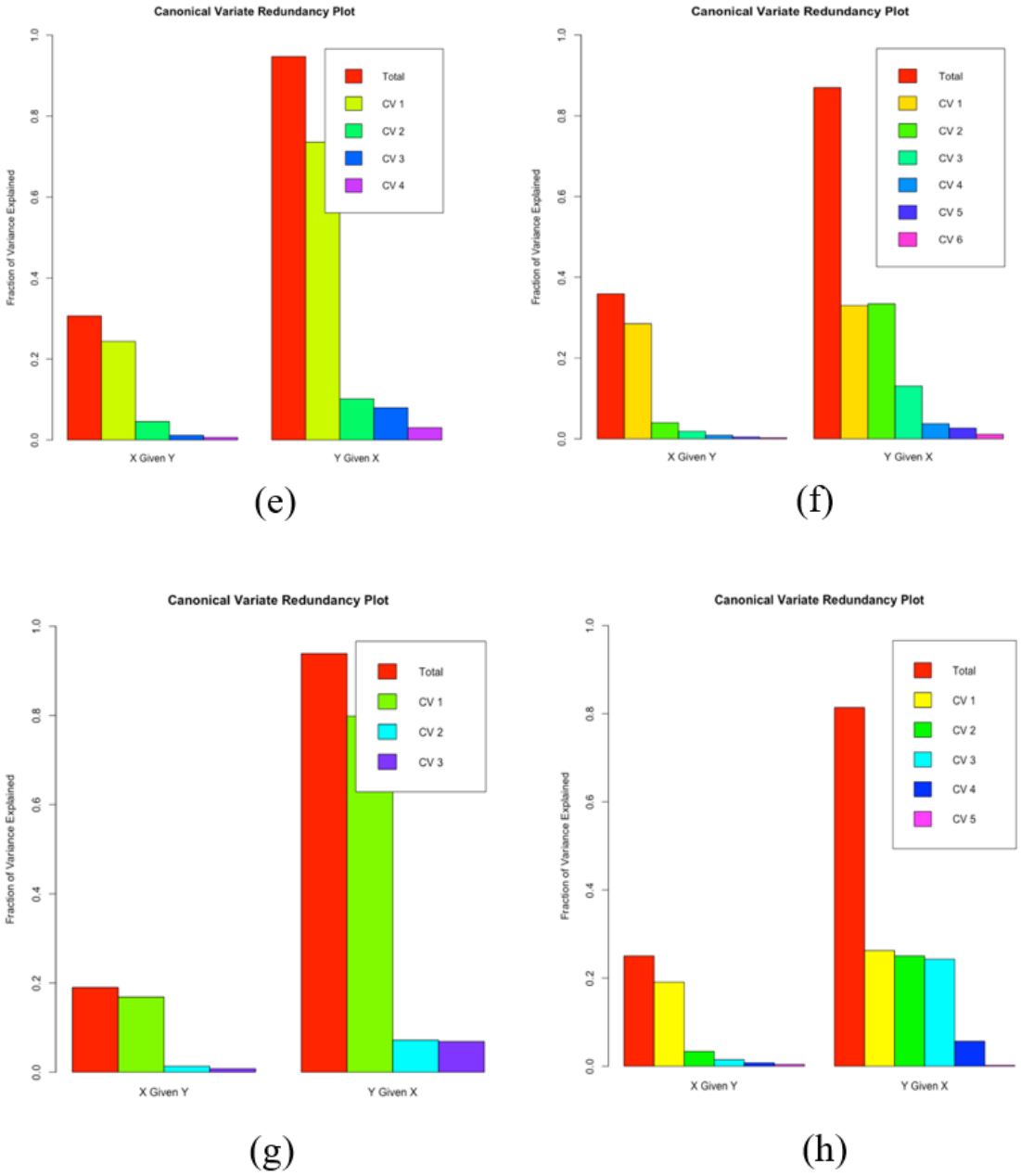


Figure A5. Canonical variate redundancy plot for (a) hospitalized cases, (b) hospitalized when vaccination has started, (c) ICU cases, (d) ICU cases when vaccination has started, (e) daily return home, (f) daily return home when vaccination has started, (g) daily deceased and (h) daily deceased when vaccination has started.