

# Identification of COVID-19 mortality patterns in Brazil by a functional QR decomposition analysis

Jorge C. Lucero\*

March 17, 2021

## Abstract

This paper introduces a functional extension of the QR decomposition of linear algebra and shows its application to identify independent patterns (curves) of COVID-19 mortality in Brazil's states. The problem is treated as a subset selection one, and regions of influence of each pattern are then determined by fitting the mortality curves of the remaining states to the main independent ones. Three main patterns are detected: (1) a two-peak curve in central and southern states, (2) a curve with an early single peak concentrated in the Amazonian state of Roraima, and (3) a curve with an early peak and a large recent increase in the Amazonian, northeastern and southeastern states.

## 1 Introduction

On March 11, 2020, the World Health Organization declared a worldwide pandemic of the COVID-19 disease caused by the new coronavirus SARS-CoV-2. The disease was identified for the first time in Wuhan, People's Republic of China, in December 2019. Until the present date, around 118 million cases have been reported, with 2.6 million deaths [29]. In Brazil, the number of cases reaches 11 millions, with 268 thousands deaths [18].

In order to contain the pandemic, countries throughout the world have implemented measures to enforce adequate hygiene and social distancing, such as the use of facial masks in public places, traveling restrictions, closing of schools, commerce and non-essential services [28, 16, 12]. Naturally, the response to those measures depends on demographic characteristics, timing of the contention measures, compliance of the population, and other factors [10]. Thus, data-driven models of the pandemic propagation are desirable to characterize and analyze underlying patterns, assess the effectiveness of containment policies, forecast its evolution, and a number of them have been proposed [14, 23].

Here, we consider a modeling approach based on the QR decomposition technique of linear algebra [11], in order to identify regions with independent patterns of COVID-19 evolution within Brazil. The QR decomposition is a matrix factorization technique that provides a simple and numerically robust solution to the so-called "subset selection problem". In that problem, a set of observations  $n$  vectors is given and a subset of the  $k$  most independent ones is sought. The subset may be used next as a basis to represent

---

\*Dept. Computer Science, University of Brasília, Brazil. E-mail: [lucero@unb.br](mailto:lucero@unb.br)

the  $n - k$  remaining vectors filtering out any data redundancy. This process has some similarities to the well-known technique of principal component analysis (PCA), in the sense that it achieves a reduction of the dimensionality of the data. However, instead of expressing the data in terms of transformations of the data, it does so in terms of a set of the most nonredundant observation vectors and therefore the results tend to have an easier interpretation [6].

In a previous study [17], the QR decomposition was applied to identify kinematic regions of the face that follow independent motion patterns during speech. The study argued that, whereas PCA could be used to extract facial gestures (i.e., temporal patterns of motion), the QR decomposition approach was more adequate to express the motion of the face in terms of eigenregions which acted as independent biomechanical units. The present study has a similar purpose in the sense that it intends to build a spatial model in terms of regions of independent behavior. Therefore, the same modeling strategy of the previous facial study will be followed, except that a functional extension of the QR decomposition will be considered.

The proposed extension fits within a functional data analysis (FDA) context [20], in which data is expressed as sets of curves instead of discrete numerical values as in traditional statistics. Techniques of FDA have been successfully applied to a variety of problems in biomedicine and public health [26]. In a recent paper, functional principal components analysis (fPCA) combined with functional clustering was used to identify patterns of COVID-19 incidence and mortality across countries [4, 15]. Applications of fPCA to model canonical correlations between confirmed and death cases in the United States [24], mortality patterns in Italian provinces [3], and build spatiotemporal models of infection risk in municipalities of Portugal [2] have also been reported. Further, variations of subset selection problems in functional contexts also have been addressed recently, such as regression analysis with a scalar response and a functional predictor [13], dimension reduction of a functional predictor for a categorical variable [25], subset selection of discrete values from a functional predictor [1], and others [9, 6]. Thus, the present study has the secondary goal of introducing the functional extension of the QR decomposition as an addition to the set of available FDA tools.

## 2 Data

### 2.1 Description and pre-processing

The evolution of the pandemic is assessed in terms of mortality rates (i.e., death counts per day), which provide a more reliable measure than infection rates [27]. Official data of COVID-19 were obtained from a repository at the Ministry of Health of Brazil [18], accessed on March 5, 2021. The data consists of records of deaths counts per day since February 25, 2020, in Brazil’s 27 federative units (26 states and a Federal District). For simplicity, the federative units will be called “states” throughout the analysis.

For each state, the period from the first confirmed death was extracted, and all extracted records were cut to the length of the shortest one (325 days). Then, the records were normalized to population size of each state and expressed in deaths per million individuals,

$$x_{ij} = \frac{\text{Number of deaths at day } j}{\text{Population size}} \times 10^6 \quad (1)$$

for  $i = 1, 2, \dots, 27$  and  $j = 1, 2, \dots, 325$ .

A few isolated mortality values were detected in the records, and those were removed by averaging them with nearby data points, as follows: if  $x_{ij} < 0$ , then

$$x_{ik} = \frac{1}{3} \sum_{\ell=-1}^1 x_{i,j+\ell} \quad (2)$$

for  $k = j - 1, j, j + 1$ .

In addition, and in order to avoid negative mortality values arising from the analysis, a mapping  $[0, \infty)$  into  $(\log \delta, \infty)$  was defined by  $y_{ij} = \log_{10}(x_{ij} + \delta)$ , where  $\delta$  is a small number. All the following processing was applied in the log domain, and mortality values were afterward recovered by a power of 10 transformation. A value of  $\delta = 0.01$  was selected by visual inspection of the data and results, so as to produce non-negative mortality curves while at the same time preventing excessive distortion due to very large negative values in the log domain for mortality values close or equal to 0.

## 2.2 Functional form

The first step of the analysis is to put the discrete data into functional form [20].

For each state  $i$ , the existence of a smooth non-negative real function  $f_i(t)$  is assumed, such that

$$y_{ij} = f_i(t_j) + \varepsilon_{ij}, \quad (3)$$

where  $t_j$  is the time at the end of day  $j$  (with  $t_1 = 0$ ), and  $\varepsilon_{ij}$  is an observational error or noise term. Each mortality function  $f_i$  is defined over the domain  $t \in [0, T]$ , with  $T = 324$  days, and is expressed in a basis expansion form

$$f_i(t) = \sum_{k=1}^K c_{ik} g_k(t) \quad (4)$$

where  $g_k(t)$ ,  $k = 1, 2, \dots, K$  is a set of basis functions and  $c_{ik}$  are the expansion coefficients. The expansion coefficients are computed by minimizing the cost function

$$F_{\lambda, f_i} = \sum_i \left\{ \sum_j [y_{ij} - f_i(t_j)]^2 + \lambda \int_T [D^2 f_i(t)]^2 dt \right\}, \quad (5)$$

where  $\lambda$  is a roughness penalty coefficient and  $D^2$  denotes the second order derivative.

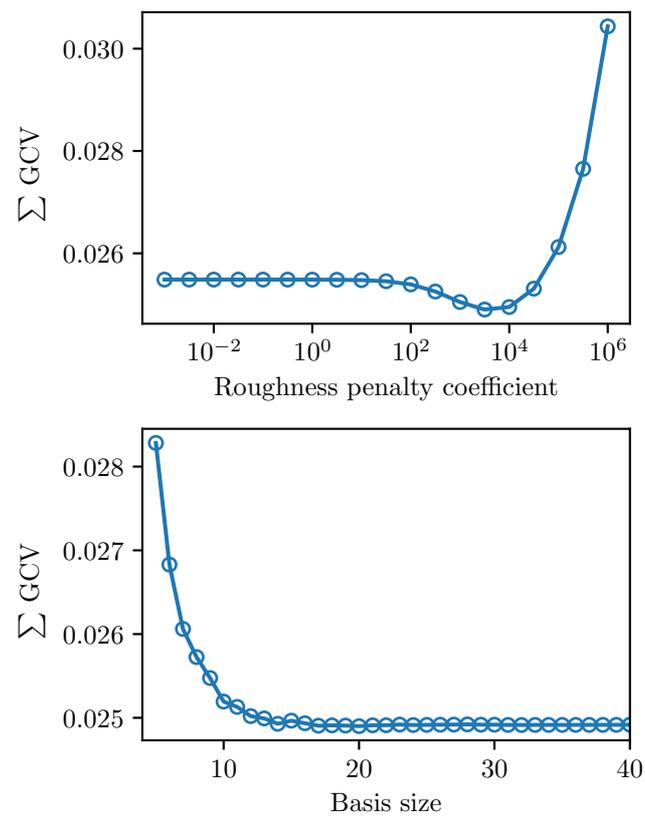
For the basis in Eq. 4, a truncated Fourier cosine series [7] was adopted, i.e.,

$$g_1(t) = 1/\sqrt{T}, \quad (6)$$

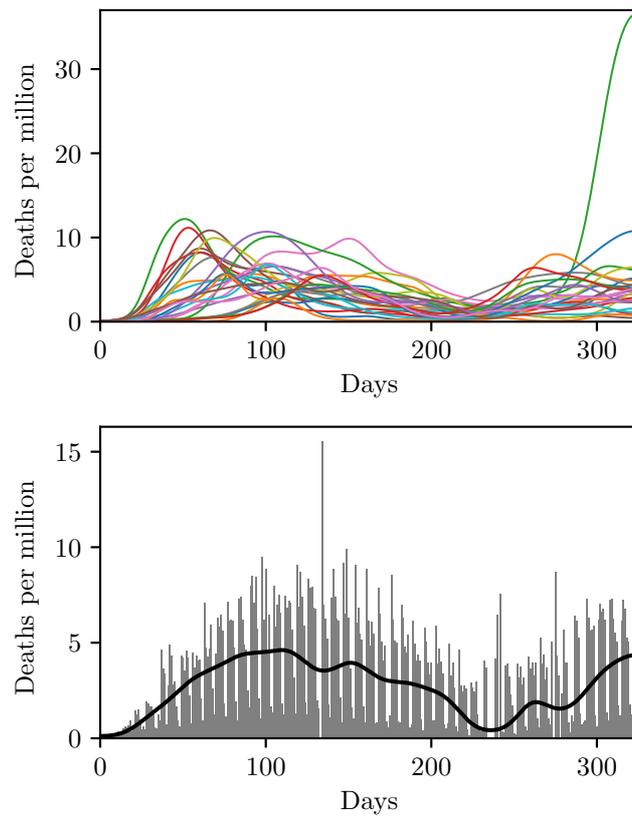
$$g_k(t) = \sqrt{2/T} \cos k\pi t/T, \quad k = 2, 3, \dots, K. \quad (7)$$

This basis was chosen because of its stability, ease of computation, and orthonormality on the interval  $[0, T]$ , which facilitates the QR decomposition in Section 3.2. A B-spline basis system was also tested, but it tended to produce a poorer fit at the extremes of the mortality curves. Optimal values of the basis size  $K$  and the roughness penalty coefficient  $\lambda$  were determined by minimizing the sum of the generalized cross validation measure (GCV) for each  $f_k$  function [8, 20], which produced  $K = 20$  and  $\lambda = 10^{3.5}$  (Fig. 1).

Fig. 2 shows all data in functional form and one example comparing the functional form to the original discrete data. The resultant functions are visually smooth and approximate well the original data, without weekly or short-term fluctuations.



**Figure 1:** Sum of the generalized cross validation measure (GCV) for all records vs. roughness penalty coefficient  $\lambda$  (top) and basis size  $K$  (bottom). Top plot:  $K = 20$ . Bottom plot:  $\lambda = 10^{3.5}$ ; the GCV is minimum at  $K = 20$  and then increases slightly as  $K$  increases.



**Figure 2:** Daily deaths per million inhabitants. Top: functional data. Bottom: original data for the state of Minas Gerais (gray bars) and functional form (black curve).

### 3 Solution of the subset selection problem by the QR decomposition

#### 3.1 Discrete version

In the so-called subset selection problem of linear algebra, a data matrix  $A \in \mathbb{R}^{m \times n}$  and an observation vector  $b \in \mathbb{R}^{m \times 1}$  are given, with  $m \geq n$ , and a predictor vector  $x$  is sought in the least squares sense; i.e., a minimizer of  $\|Ax - b\|_2^2$  [11]. However, instead of using the whole data matrix  $A$  to predict  $b$ , only a subset of its columns is used so as to filter out any data redundancy. This problem may be solved by the QR decomposition with column pivoting [11]. The decomposition expresses  $A$  in the form  $AP = QR$ , where  $P \in \mathbb{R}^{n \times n}$  is a column permutation matrix,  $Q \in \mathbb{R}^{m \times m}$  is an orthogonal matrix, and  $R \in \mathbb{R}^{m \times n}$  is an upper triangular matrix with positive diagonal elements. A simplified variant is the “thin” version, in which  $Q \in \mathbb{R}^{m \times n}$  and  $R \in \mathbb{R}^{n \times n}$ .

The first column of  $AP$  is the column of  $A$  that has the largest 2-norm. The second column of  $AP$  is the column of  $A$  that has the largest component in a direction orthogonal to the direction of first column. In general, the  $k$ th column of  $AP$  is the column of  $A$  with the largest component in a direction orthogonal to the directions of the first  $k - 1$  columns. Thus, the algorithm reorders the columns of  $A$  so as to make its first columns as well conditioned as possible. The first  $k$  columns of  $AP$  may be then adopted as the sought subset of  $k$  least dependent columns. The diagonal elements of  $R$  ( $r_{kk}$ ), also called the “ $R$  values”, measure the size of the orthogonal components, and they appear in decreasing order for  $k = 1, \dots, n$ .

Once the subset of  $k$  columns of  $A$  has been selected, the dimensionality of the data may be reduced by fitting the remaining  $n - k$  columns of  $A$ , as follows. Consider the thin decomposition and define the following block partitions:

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}, \quad P = \begin{bmatrix} P_1 & P_2 \end{bmatrix}, \quad (8)$$

where  $R_{11} \in \mathbb{R}^{k \times k}$ ,  $P_1 \in \mathbb{R}^{n \times k}$ , and the dimensions of the other blocks match accordingly. Then, the remaining  $n - k$  columns of  $AP$  may be approximated as  $AP_2 \approx AP_1 X$ , where  $X \in \mathbb{R}^{k \times (n-k)}$  is the solution of the upper triangular system  $R_{11} X = R_{12}$  [17]. The residual of the approximation is  $\|R_{22}\|_F$ , where subindex F indicates the Frobenius norm.

#### 3.2 Functional extension

In the present case, we have a data set of  $n$  functional observations  $f_i(t)$ . In vector form, the data may be expressed as  $A = [f_1, f_2, \dots, f_n]$ . The functional QR decomposition is defined analogously to the discrete version, as follows.

For functions  $\xi(t), \psi(t)$ , the inner product over the interval  $T$  is defined as

$$\langle \xi, \psi \rangle = \int_T \xi(s)\psi(s)ds. \quad (9)$$

Then,  $\|\xi\|^2 = \langle \xi, \xi \rangle$ , and the projection of  $\xi$  on the direction of  $\psi$  is  $\text{proj}_\psi \xi = \langle \xi, \psi \rangle \psi / \|\psi\|^2$ .

Define next the orthogonal functions

$$u_1 = f_1, \tag{10}$$

$$u_2 = f_2 - \text{proj}_{u_1} f_2, \tag{11}$$

$$\dots \tag{12}$$

$$u_n = f_n - \sum_{j=1}^{n-1} \text{proj}_{u_j} f_n. \tag{13}$$

Thus, matrix  $Q$  has the form  $Q = [q_1(t), q_2(t), \dots, q_n(t)]$ , where  $q_i = u_i/\|u_i\|$ , and matrix  $R$  is

$$R = \begin{pmatrix} \langle q_1, f_1 \rangle & \langle q_1, f_2 \rangle & \dots & \langle q_1, f_n \rangle \\ 0 & \langle q_2, f_2 \rangle & \dots & \langle q_2, f_n \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \langle q_n, f_n \rangle \end{pmatrix} \tag{14}$$

Matrix  $P$  is obtained by reordering the components of  $F$  so that the main diagonal of  $R$  has its elements in decreasing order from top to bottom.

Matrices  $Q$ ,  $R$  and  $P$  may be computed directly from the above equations; however, such a strategy may suffer from the same problems of numerical instability as in the discrete case [11]. A second possibility is to discretize functions  $f_i(t)$  in a number of data points over interval  $[0, T]$ , and so transforming the functional problem into a discrete one [6]. However, a simpler and more convenient strategy is to use the expansion in Eq. (4).

In matrix form, the expansion is

$$A = GC, \tag{15}$$

where  $G = [g_1, g_2, \dots, g_K]$  and  $C$  is a  $K \times n$  matrix of coefficients  $c_{ik}$ . Letting  $AP = QR$  and expressing functions  $g_i(t)$  in the same basis system as functions  $f_i(t)$ , we have  $Q = GB$ , where  $B$  is a  $K \times n$  matrix of coefficients  $b_{ik}$ . Replacing into Eq. (15) and simplifying, we obtain

$$CP = BR \tag{16}$$

We know that matrix  $B$  is orthogonal (since both  $Q$  and  $G$  are orthogonal), and that the QR decomposition with column pivoting is unique [11]. Therefore, Eq. (16) represents the standard (discrete) QR decomposition of matrix  $C$ , which may be computed using available algorithms of matrix algebra.

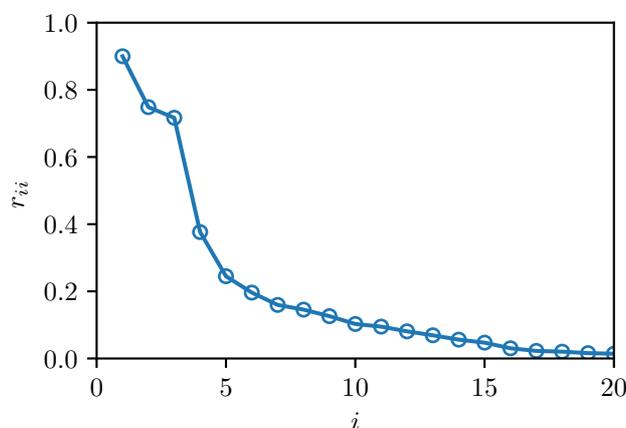
## 4 Results of the COVID-19 data analysis

The states with the most independent log mortality functions are listed in Table I, and Fig. 3 shows the whole set of  $R$  values for the data. The  $R$  values decrease as the number of selected functions increases, with a clear gap between between the third and fourth values. The gap suggests that the main independent functions may be reduced to the first three [22].

The relative error in reconstructing the data set from the first  $k$  selected functions, computed as  $\|R_{22}\|_F/\|R\|_F$  (i.e., in the log domain), is shown in Fig. 4. We can see a large decrease of the error at  $k = 3$  from 68.8% to 38.1 %, matching the pattern of  $R$  values. However, when computing the error on the mortality functions themselves

| #  | Code | Name               | Region       | $R$ value |
|----|------|--------------------|--------------|-----------|
| 1  | MS   | Mato Grosso do Sul | Central-West | 0.90      |
| 2  | RR   | Roraima            | North        | 0.75      |
| 3  | AM   | Amazonas           | North        | 0.72      |
| 4  | DF   | Federal District   | Central-West | 0.38      |
| 5  | GO   | Goiás              | Central-West | 0.24      |
| 6  | CE   | Ceará              | Northeast    | 0.20      |
| 7  | RO   | Rondônia           | North        | 0.16      |
| 8  | AC   | Acre               | North        | 0.15      |
| 9  | SC   | Santa Catarina     | South        | 0.13      |
| 10 | TO   | Tocantins          | North        | 0.10      |

**Table 1:** First 10 states with the most independent log mortality functions.

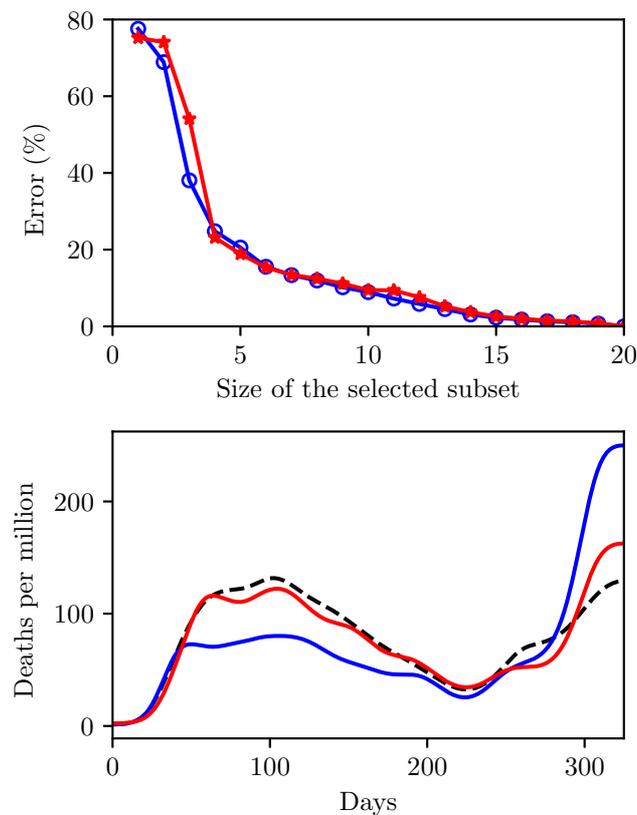


**Figure 3:**  $R$  values.

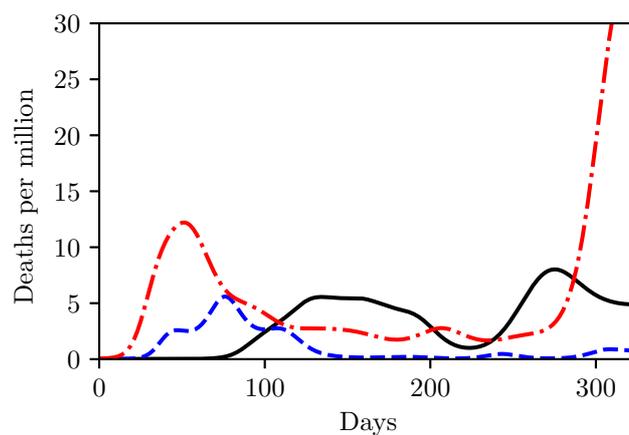
( $f_i$ ), the large decrease appears at  $k = 4$ . The figure also shows an example of the reconstruction of the total mortality function for Brazil using 3 and 4 selected functions, in comparison to the initial functional data. For  $k = 4$ , the general pattern is well approximated. Note that the model is based on a log transformation of the mortality data, and a drawback of such transformation is that it puts more weight in low values of mortality values, particularly those close to zero, and less in high values. Thus, the reconstruction of the mortality curves tend to be poorer in regions with large mortality values, as shown by the bottom plots in Fig. 4.

The first three selected functions correspond to the states of Mato Grosso do Sul (MS), Roraima (RR) and Amazonas (AM), in that order, and the mortality curves are plotted in Fig. 5. Further, Fig. 6 shows the result of fitting the remaining states to the first three ones, as explained in Section 3.1. Since the fit is performed in the log domain, then a negative value of a coefficient (blue states in Fig. 6) corresponds to a low value in the mortality domain.

The first mortality curve, corresponding to Mato Grosso do Sul, shows two local maxima or peaks around days 130-160 and day 275, respectively, indicating two “waves” of the pandemic. The state is located in the Central-West region of Brazil, and its pattern is mostly representative of the central and southern areas of the country.



**Figure 4:** Top: relative error of fitting the data set to the first  $k$  selected functions vs.  $k$ , computed on the log mortality functions (blue circles) and mortality functions (red stars). Bottom: reconstruction of the total mortality curve for the whole country, with  $k = 3$  (solid blue curve),  $k = 4$  (solid red curve) and original functional data (dashed black curve).



**Figure 5:** Main independent mortality curves: Mato Grosso do Sul (solid black curve), Roraima (dashed blue curve), Amazonas (dash-point red curve).

The second mortality curve, corresponding to the state of Roraima, shows a local maxima much earlier than the first pattern, around day 75 and relative small daily deaths afterward. Roraima is Brazil's northernmost state, in the Amazonian region, and borders with both Venezuela and Guyana. It has an intense flow of people across its international borders, which may have contributed to the earlier peak [5]. Other contributing factors may have been its high percentage of indigenous population, more susceptible against contagious diseases, as well as its poorer developed public health care system [19]. According to Fig. 6, its mortality pattern is concentrated in the state itself, but it has some effect in states at the north, northeast and center of the country.

The third mortality curve, corresponding to the state of Amazonas, shows a local maxima even earlier than Roraima, in day 51, but also a very large increase at the right end. The early first peak may have the same causes as in the case of Roraima, namely, international borders, large indigenous population and poorly developed health care system. The large increase at the right end is most likely a consequence of the new lineage P.1 of the SARS-CoV-2 virus detected in Manaus (capital city of Amazonas) at the beginning of 2021, which has higher transmissibility than previous lineages [21]. Fig. 6 shows that this third mortality pattern is prevalent in northern and northeastern states, as well as in southeastern states as São Paulo and Rio de Janeiro.

## 5 Conclusion

This paper has introduced a simple functional extension of the QR decomposition technique of linear algebra, and shown its application to identify independent patterns of COVID-19 evolution in Brazil. Each pattern defines an epidemiological region of influence, and the overall evolution of the pandemic in the country may be then modeled (in the log domain) as a linear combination of the behavior of those regions. Naturally, the accuracy of the model depends on the number of independent patterns considered. Only the first three mortality patterns were discussed here for a general qualitative view of the pandemic evolution in the country; however, a larger number should be included if a more precise representation is desired.

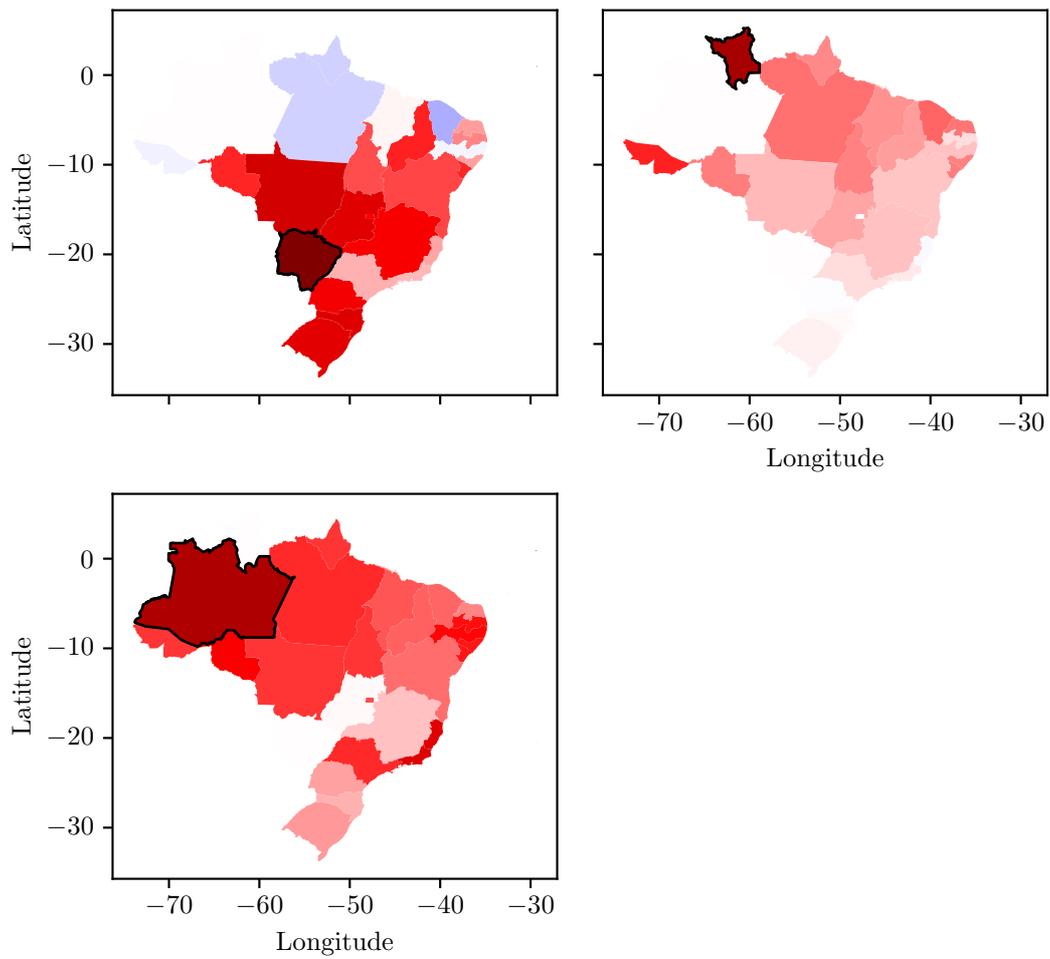
The functional expansion of the data adopted an orthogonal basis to facilitate the computation of the QR decomposition. Nevertheless, further development of the decomposition algorithm to allow for the use of non-orthogonal basis systems, such as the widely used B-splines, would be desired as a next step.

## Acknowledgments

This work was supported by the Committee of Research, Innovation, and Extension to Combat COVID-19 (COPEI) of the University of Brasília.

## References

- [1] G. Aneiros and P. Vieu. Variable selection in infinite-dimensional problems. *Statistics & Probability Letters*, 94:12–20, 2014.
- [2] L. Azevedo, M. J. Pereira, M. Ribeiro, and A. Soares. Spatiotemporal modelling of COVID-19 infection risk in Portugal (preprint). *Research Square*, 2020.



**Figure 6:** Regions of influence of the first three independent functions, corresponding to Mato Grosso do Sul (top left), Roraima (top right), and Amazonas (bottom left). The state corresponding to each function is outlined in black. The value of the square fit coefficient of each state relative to the main selected state is shown in a red (positive) to blue (negative) scale, and the darker the color the larger the magnitude of the coefficient.

- [3] T. Boschi, J. Di Iorio, L. Testa, M. A. Cremona, and F. Chiaromonte. The shapes of an epidemic: using Functional Data Analysis to characterize COVID-19 in Italy (preprint). *arXiv:2008.04700 [stat.AP]*, 2020.
- [4] C. Carroll, S. Bhattacharjee, Y. Chen, P. Dubey, J. Fan, I. Gajardo, X. Zhou, H.-G. Müller, and J.-L. Wang. Time dynamics of covid-19. *Scientific Reports*, 10(1):21040, 2020.
- [5] S. Cimerman, A. Chebabo, C. A. da Cunha, and A. J. Rodríguez-Morales. Deep impact of COVID-19 in the healthcare of Latin America: The case of Brazil. *The Brazilian Journal of Infectious Diseases*, 24(2):93–95, 2020.
- [6] A. Cuevas. A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23, 2014.
- [7] H. F. Davis. *Fourier series and orthogonal functions*. Dover Publications, Mineola, NY, 1989.
- [8] M. Febrero-Bande and M. O. de la Fuente. Statistical computing in functional data analysis: The R package *fda.usc*. *Journal of Statistical Software*, 51(4), 2012.
- [9] R. Fraiman, Y. Gimenez, and M. Svarc. Feature selection for functional data. *Journal of Multivariate Analysis*, 146:191–208, 2016.
- [10] P. D. Giamberardino and D. Iacoviello. Evaluation of the effect of different policies in the containment of epidemic spreads for the COVID-19 case. *Biomedical Signal Processing and Control*, 65:102325, 2021.
- [11] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, 1996.
- [12] M. A. Honein, A. Christie, D. A. Rose, J. T. Brooks, D. Meaney-Delman, A. Cohn, E. K. Sauber-Schatz, A. Walker, L. C. McDonald, L. C. Liburd, J. E. Hall, A. M. Fry, A. J. Hall, N. Gupta, W. L. Kuhnert, P. W. Yoon, A. V. Gundlapalli, M. J. Beach, H. T. Walke, E. Azziz-Baumgartner, S. Bennett, C. Braden, J. Buigut, T. Chiller, C. R. Friedman, C. M. Greene, O. Henao, C. Kosmos, A. MacNeil, B. Marston, G. Massetti, J. Montero, C. G. Perrine, K. Polen, K. Remley, R. Salerno, K. A. Shaw, and I. Williams. Summary of guidance for public health strategies to address high levels of community transmission of SARS-CoV-2 and related deaths. *Morbidity and Mortality Weekly Report*, 69(49):1860–1867, 2020.
- [13] G. M. James, J. Wang, and J. Zhu. Functional linear regression that’s interpretable. *The Annals of Statistics*, 37(5A):2083–2108, 2009.
- [14] E. Kuhl. Data-driven modeling of COVID-19 – Lessons learned. *Extreme Mechanics Letters*, 40:100921, 2020.
- [15] V. Kumar, A. Sood, S. Gupta, and N. Sood. Prevention- versus promotion-focus regulatory efforts on the disease incidence and mortality of COVID-19: A multi-national diffusion study using functional data analysis. *Journal of International Marketing*, 29(1):1–22, 2020.

- [16] J. A. Lewnard and N. C. Lo. Scientific and ethical basis for social-distancing interventions against COVID-19. *The Lancet Infectious Diseases*, 20(6):631–633, 2020.
- [17] J. C. Lucero and K. G. Munhall. Analysis of facial motion patterns during speech using a matrix factorization algorithm. *The Journal of the Acoustical Society of America*, 124(4):2283–2290, 2008.
- [18] Ministry of Health (Brazil). COVID 19 – Painel Coronavírus. Online in <https://covid.saude.gov.br/>, 2021. Accessed on March 6, 2021 00:54 GMT.
- [19] C. V. C. Palamim, M. M. Ortega, and F. A. L. Marson. COVID-19 in the indigenous population of Brazil. *Journal of Racial and Ethnic Health Disparities*, 7(6):1053–1058, 2020.
- [20] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer-Verlag, New York, NY, 1997.
- [21] E. C. Sabino, L. F. Buss, M. P. S. Carvalho, C. A. Prete, M. A. E. Crispim, N. A. Fraiji, R. H. M. Pereira, K. V. Parag, P. da Silva Peixoto, M. U. G. Kraemer, M. K. Oikawa, T. Salomon, Z. M. Cucunuba, M. C. Castro, A. A. de Souza Santos, V. H. Nascimento, H. S. Pereira, N. M. Ferguson, O. G. Pybus, A. Kucharski, M. P. Busch, C. Dye, and N. R. Faria. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *The Lancet*, 397(10273):452–455, 2021.
- [22] M. Setnes and R. Babuska. Rule base reduction: some comments on the use of orthogonal transforms. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 31(2):199–206, 2001.
- [23] N. Talkhi, N. A. Fatemi, Z. Ataei, and M. J. Nooghabi. Modeling and forecasting number of confirmed and death caused COVID-19 in Iran: A comparison of time series forecasting methods. *Biomedical Signal Processing and Control*, 66:102494, 2021.
- [24] C. Tang, T. Wang, and P. Zhang. Functional data analysis: An application to COVID-19 data in the United States (preprint). *arXiv:2009.08363 [stat.AP]*, 2020.
- [25] T. S. Tian and G. M. James. Interpretable dimension reduction for classifying functional data. *Computational Statistics & Data Analysis*, 57(1):282–296, 2013.
- [26] S. Ullah and C. F. Finch. Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 13(1):43, 2013.
- [27] G. L. Vasconcelos, A. M. Macêdo, R. Ospina, F. A. Almeida, G. C. Duarte-Filho, A. A. Brum, and I. C. Souza. Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *PeerJ*, 8:e9421, 2020.
- [28] A. Wilder-Smith and D. O. Freedman. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *Journal of Travel Medicine*, 27(2):taaa020, 2020.
- [29] Worldometer. COVID-19 Coronavirus Pandemic. Online in <https://www.worldometers.info/coronavirus/>, 2021. Accessed on March 9, 2021, 23:40 GMT.