Time warping between main epidemic time series in epidemiological surveillance

Jean-David Morel[1], Jean-Michel Morel[2] and Luis Alvarez[3],

[1] Laboratoire de Physiologie Intégrative et Systémique
Ecole Polytechnique Fédérale de Lausanne,
AI 1144 Station 15 CH-1015 Lausanne Switzerland

[2] Université Paris-Saclay, ENS Paris-Saclay, CNRS,
Centre Borelli, F-94235, Cachan, France.

[3] CTIM. Departamento de Informática y Sistemas,
Universidad de Las Palmas de Gran Canaria. Spain

February 7, 2023

Abstract

The most common reported epidemic time series in epidemiological surveillance are the daily or weekly incidence of new cases, the hospital admission count, the ICU admission count, and the death toll, which played such a prominent role in the struggle to monitor the Covid-19 pandemy. Based on a generalization of the renewal equation, we explore the hypothesis that each pair of such curves is related to the other one by a smooth time varying delay and by a smooth ratio generalizing the reproduction number. This hypothesis is also explored for pairs of curves measuring the same indicator in two different countries. Given two such time series, we develop, based on a signal processing approach, an efficient numerical method for computing their time varying delay and ratio curves, and we verify that its results are consistent. Indeed, they experimentally verify symmetry and transitivity requirements and we also show, using realistic simulated data, that the method is able to properly recover the time delays and ratios ground truth. Moreover, on all real examples considered, the method leads to explainable time delays and ratios. The proposed method generalizes and unifies many recent related attempts to take advantage of the plurality of these health data across regions or countries and time, providing a better understanding of the relationship between them. An implementation of the method is publicly available at the *EpiInvert* CRAN package.

Keywords: COVID-19, time warping, epidemic time series, epidemiological surveillance, mortality ratio, renewal equation, variational models.

1 Introduction

The COVID-19 epidemic has provided us with information, for many countries, on the evolution of a number of key instantaneous epidemic measurements such as the number of cases, deaths, hospitalization, or intensive care

units (ICUs) occupation. In this work, we explore the functional relationship between these time series. Let $f(t)$ and $g(t)$ be a pair of time series, we will assume that there exists a smooth time varying delay, $s(t)$, between the values of both curves (for instance, we expect a temporal delay between cases and deaths). Moreover, we assume that once the delay between both curves is compensated, the ratio between their values, $r(t)$, is a smooth function that actually expresses a causality between both time series. Therefore we look for smooth functions $s(t)$ and $r(t)$ such that

$$r(t)f(t) \approx g(t + s(t)) \tag{1}$$

of course, this problem is ill-posed because there are an infinite number of combinations of functions $r(t)$ and $s(t)$ that satisfy the above equation. However, the problem becomes well-posed if we force the functions $r(t)$ and $s(t)$ to be smooth. The problem is then posed in terms of "we want the equation (1) to verify as well as possible but for $r(t)$ and $s(t)$ as smooth as possible".

Understanding the time delay dynamics between epidemic time series provides insights that are key to discerning changes in the phenotype of the virus, the demographics impacted, the efficacy of treatment, and the ability of the health service to manage large volumes of patients. In [30], authors studied how the pandemy has evolved in the United Kingdom through the temporal changes to the epidemiological time delay distributions for clinical outcomes. In [18] ten time periods are used to track changes in delays and the probabilities of outcomes using parametric statistical models in France. In each period, the variability of the delay is modeled using a parametric distribution. In [16], the difference between in-hospital mortality across four time periods was studied. [10] studied the evolution, in the USA, of in-hospital mortality by month of admission.

Comparison between epidemic waves is also a usual way to study the time evolution of epidemiological indicators. For instance, in [26], authors show that the in-hospital mortality was lower during the second wave of COVID-19 than in the first wave. Mortality predictors and differences in mortality between COVID-19 waves were identified using logistic regression in [6].

In this paper, we propose a different approach where instead of separating the epidemic time interval in different periods, we model, using a signal processing approach, the evolution of the delay $s(t)$ and the ratio $r(t)$ between indicators as smooth functions defined in the whole epidemic time interval.

The rest of the paper is organized as follows: in Section 1 we propose an analysis of the relevant literature. We start by a short account of time warping techniques in section 1.1. We then review in section 1.2 the numerous propositions that compare measured signals by estimating a time warping between them and we end in section 1.3 by relating time warping to a classical epidemiological model, the renewal equation. This model gives a simple key to interpret model (1). Section 2 details the numerical method to solve (1) including the imposition of regularity to $r(t)$ and $s(t)$. Section 3 shows several significant applications to various epidemic time series of COVID-19. Section 4 is a conclusion of the proposed approach.

1.1 Dynamic time warping: the technique

Dynamic time warping is a method to match two signals $f(t)$ and $g(t)$ by finding a nondecreasing function shift function $w(t)$ such that $f(t)$ is as close as possible to $g(w(t))$. Variants of dynamic time warping like the one proposed in [15] match the slopes of the signals rather than the signals themselves. If the signals are discretized, thus encoded as two vectors, Bellman's dynamic programming principle [5] applies and yields an optimal solution for any time discrete version of the cost $E(w) =: \int c(f(t) - g(w(t))^2 dt$, where c is a positive cost function. As developed in [22], a discrete path optimization is performed in the 2D space of time correspondances, where only diagonal $(t, s) \to (t + 1, s + 1)$ horizontal $(t, s) \to (t + 1, s)$, and vertical $(t, s) \to (t, s + 1)$ moves are authorized. This discrete setting is efficiently solved by dynamic programming, and it can be extended to an

optimal warping with time variable penalty [7]. We refer to [27] for an algorithm review. In this paper, we shall prefer a continuous formulation to the discrete one, in the line of [28], where a continuous time warping functional is defined for matching multiple copies of the same audio by a signal alignment that is robust to the presence of sparse outliers with arbitrary magnitude. The continuous formulations allow for a continuous iterative optimization by gradient descent and remove limitations on the form of the functional or discretization of the warp. As proposed in [32], a continuous penalty on the derivative of the warp can be used to ensure that this derivative does not vanish and does not become too high.

In the same line, the warping of signals has a 2D classic equivalent in the optical flow problem introduced in a seminal paper [13]. Applied to two successive images of a video, the method finds a vector field (2D warp) matching both images and enforces the warping field to be smooth by combining a regularity term with the matching fidelity term. In its initial formulation, the differential method worked only if the motion does not exceed one pixel. A more general computation of the optical flow is developed in [4] as a hierarchical multiscale implementation of the optical flow using the Laplacian pyramid to represent both compared images at different scales. These ideas are developed in [21] which implements a multiscale version of Horn-Schunk optical flow. The same hierarchical differential method is obviously applicable to pairs of similar signals.

1.2 Epidemiological applications of time warping

In [29] the authors used dynamic warping to forecast the future spread of COVID-19 by exploiting the identified lead-lag effects between different countries. The past relation among nations is determined through dynamic time warping between cumulative case curves. Then the cumulative case curve of the leading country is used to predict the Covid-19 spread in the following nation. In [14] the authors similarly present an algorithm for the clustering of confirmed COVID-19 cases at the county level in the United States. Dynamic time warping is used as a k-means clustering distance metric. The obtained clusters enable retrospective interpretation of the pandemy and informative inputs for case prediction models. In [17], the same method is developed for the Polish voivodeships Similarities in time series are found by dynamic time warping for those infected with, and those who died of, COVID-19 in pairs of vovoideships . The authors argue that a time warping method is important because the coronavirus epidemic did not start in all voivodeships at the same time. The method jointly analyzes the number of infected and deceased people in each province. Dynamic warping is applied to match pairs of (smoothed) incidence curves, pairs of cumulative case curves, and pairs of cumulative death curves.

The main objective of the study in [9] is to assess the similarity between the time series of energy commodity prices and the time series of daily COVID-19 cases. The authors argue that the pandemy affects all aspects of the global economy and propose to assess the connections between the number of COVID-19 cases and the energy commodities sector. These connections are achieved by using the Dynamic Time Warping (DTW) method to compute DTW distances between all time series—and use them to group the energy commodities according to their price change. In that way, the authors demonstrate that some commodities such as natural gas are strongly associated with the development of the pandemy. In [11] dynamic time warping is used to compare temporal patterns in patient disease trajectories. This enables the assessment of comorbidities in population-based studies, as it permits to identify more complex disease patterns apart from the pairwise disease associations. The disease-history vectors of patients of a regional Spanish health dataset are represented as time sequences of ordered disease diagnoses. Statistically significant pairwise disease associations are identified and their temporal directionality is assessed. Subsequently, an unsupervised clustering algorithm, based on Dynamic Time Warping, is applied on the common disease trajectories in order to group them according to the temporal patterns that they share.

1.3 The renewal equation

The reproduction number $r(t)$ (in the literature often written R_t or Rt) is a key epidemiological parameter evaluating the transmission rate of a disease over time. It is defined as the average number of new infections caused by a single infected individual at time t in a partially susceptible population [25]. The reproduction number $r(t)$ can be computed from the daily observation of the incidence curve $i(t)$, but requires empirical knowledge of the probability distribution $k(s)$ of the delay between two infections [12, 3].

There are two different models for the incidence curve and its corresponding infection delay $k(s)$. In a theoretical model, $i(t)$ would represent the real daily number of new infections, and $k(s)$ is sometimes called *generation time* [31, 8] and represents the probability distribution of the time between infection of a primary case and infections in secondary cases, this time delay distribution being assumed stationary (independent of t). In practice, neither parameter is easily observable because the infected are rarely detected before the appearance of symptoms and tests will be negative until the virus has multiplied over several days. What is routinely recorded by health organizations is the number of *new detected, incident cases*. When dealing with this real incidence curve, $k(s)$ is called *serial interval* [31, 8]. The serial interval is defined as the delay between the onset of symptoms in a primary case and the onset of symptoms in secondary cases [8].

A fundamental equation links $r(t)$ to $i(t)$ and $k(s)$. It is the *renewal equation*, first formulated for birth-death processes in a 1907 note of Alfred Lotka [19]. We adopt the Nishiura et al. formulation [23, 24],

$$i(t) = \sum_{s=f_0}^{f} r(t-s)i(t-s)k(s) \tag{2}$$

where f_0 and f are the maximal and minimal observed times between a primary and a secondary case. In the renewal equation, the incidence curve $i(t)$ is causally linked to itself through an internal delay represented by the probability distribution $k(s)$ and a reproducing factor $r(t)$. We now consider generalizations of this renewal that link time series that can still be related by causality, time delay, and a reproducing factor.

Our first example is the reproducing link from cases to deaths. Consider the time delay probability distribution $k_t(s)$ that a person reported dead at time t was reported positive at time $t-s$. Then we call $r(t)$ the probability that a person diagnosed at time t will eventually die from the disease. With these assumptions, we obtain a *generation equation* linking cases to deaths, namely

$$d(t) = \sum_{s\geq 0} i(t-s)r(t-s)k_t(s), \tag{3}$$

where $d(t)$ denotes the daily death toll and $i(t)$ the incidence curve. A strong simplification of the model may replace k_t by the Dirac mass concentrated at its center of mass $\sigma(t) =: \sum_s sk_t(s)$. Then (3) boils down to

$$d(t) = i(t-\sigma(t))r(t-\sigma(t)), \tag{4}$$

which amounts to assuming that all persons dead at time t have been detected at time $t-\sigma(t)$. Let us now relate eq. (4) to the simpler equation (1). This can be done by setting $\hat{t}(t) = t - \sigma(t)$. Then (4) rewrites

$$d(t(\hat{t})) = i(\hat{t})r(\hat{t})$$

where $t(\hat{t})$ is the reverse change of variable. Setting $s(\hat{t}) = t(\hat{t}) - \hat{t}$ we get

$$d(\hat{t} + s(\hat{t})) = i(\hat{t})r(\hat{t})$$

which is precisely (1). It remains to ask why the change of variable is licit. This is ensured if $|\sigma'(t)| < 1$, so that $\hat{t}(t)$ is strictly increasing. We shall ensure such a condition in our formalization of the problem.

2 Mathematical formalization and resolution of (1)

Given two causally related time series $f(t)$ and $g(t)$, we introduced the problem of finding a smooth time warping $s(t)$ and a smooth reproduction number $r(t)$ such that

$$r(t)f(t) \approx g(t + s(t)) \tag{5}$$

This leads us to solve the following variational formulation of the problem where $r(t)$ and $s(t)$ are estimated by minimizing the error functional

$$E(r,s) = \int_0^T (r(t)f(t) - g(t+s(t)))^2 + w_r \int_0^T r'(t)^2 dt + w_s \int_0^T s'(t)^2 dt + w_o \int_0^T (s_{min} - s(t))_+^2 + (s(t) - s_{max})_+^2 dt. \tag{6}$$

The first term of the energy (6) penalizes a poor matching between $r(t)f(t)$ and $g(t + s(t))$. The second and third terms of the energy impose that $r(t)$ and $s(t)$ be smooth, and the strength of this smoothing is controlled by the weights w_r and w_s. The larger the values of these weights the more regular $r(t)$ and $s(t)$ will be. The fourth term penalizes warps $s(t)$ that would fall outside an expected interval $[s_{min}, s_{max}]$ (the function $(x)_+$ is defined as $(x)_+ = x$ if $x > 0$ and $(x)_+ = 0$ otherwise).

To compute the minima of the energy (6), we first normalize the time series $f(t)$ and $g(t)$ by dividing them by their medians. In this way, the minimization is independent of the time series magnitudes. Once the energy is minimized, the obtained value of $r(t)$ is scaled to compensate for this normalization. Next, we express the energy in a discrete setting as

$$E(\{r_k, s_k\}) = \sum_{k=0}^{K} (r_k f_k - g_k^s)^2 + w_r \sum_{k=1}^{K} (r_k - r_{k-1})^2 + w_s \sum_{k=1}^{K} (s_k - s_{k-1})^2 + w_o \sum_{k=1}^{K} \left((s_k - s_{\max})_+^2 + (s_{\min} - s_k)_+^2 \right), \tag{7}$$

where $r_k = r(t_k)$, $s_k = s(t_k)$, $f_k = f(t_k)$, and $g_k^s = g(t_k + s_k)$. We use an alternate iterative gradient descent type method to compute $\{r_k^n\}$ and $\{s_k^n\}$. Given $\{s_k^n\}$ we compute $\{r_k^n\}$ by expressing the derivatives of the energy with respect to r_k^n and equating them to 0. This yields a system of linear equations in $\{r_k^n\}$,

$$\left(r_k^n f_k - g_k^{s^n} \right) f_k + w_r (2r_k^n - r_{k-1}^n - r_{k+1}^n) = 0 \qquad \text{for } k = 1, .., K-1;$$

the cases $k = 0$ and $k = K$ are managed using the boundary conditions. If the user has selected a value $\tilde{r}_0$ or $\tilde{r}_K$ for r_0 or r_K, then we add to the system the associated equation $r_0 = \tilde{r}_0$ or $r_K = \tilde{r}_K$. Otherwise we add the equations

$$\left(r_0^n f_0 - g_0^{s^n} \right) f_0 + w_r (r_0^n - r_1^n) = 0$$

or

$$\left(r_K^n f_K - g_K^{s^n} \right) f_K + w_r (r_K^n - r_{K-1}^n) = 0.$$

Given $\{r_k^n, s_k^n\}$, we compute s_k^{n+1} using a first order approximation of $g(t_k + s_k^{n+1})$. We express $s_k^{n+1} = s_k^n + h_k^n$ (then the unknown becomes h_k^n) and we replace in the energy $g_k^{s^{n+1}} = g(t_k + s_k^{n+1})$ by $g(t_k + s_k^n) + g'(t_k + s_k^n) h_k^n = g_k^{s^n} + g_k'^{,s^n} h_k^n$, (where $g_k'^{,s^n} = g'(t_k + s_k^n)$). Differentiating the energy with respect to h_k^n we obtain for $k = 1, .., K-1$, a linear equation in $\{h_k^n\}$,

$$-\left(r_k^n f_k - g_k^{s^n} - g_k'^{,s^n} h_k^n \right) g_k'^{,s^n} + w_s (2s_k^n - s_{k-1}^n - s_{k+1}^n + 2h_k^n - h_{k-1}^n - h_{k+1}^n) + w_o (s_k^n - s_{\max})_+ - w_o (s_{\min} - s_k^n)_+ = 0,$$

where the boundary conditions in $k = 0$ and $k = K$ are managed in the same way as for r_k^n. We perform iterations of the alternate scheme to compute $\{r_k^n, s_k^n\}$ until convergence. Notice that we need to initialize $\{s_k^0\}$ to start the iteration. We use two approaches to get an initialization of $\{s_k^0\}$. The first one is to put in correspondence the local maxima of $f(t)$ and $g(t)$. That provides the value of $\{s_k^0\}$ as the shift between the local maxima. For the rest of the points, we use linear interpolation to compute $\{s_k^0\}$ between the shift obtained for the local maxima. The second approach to initialize $\{s_k^0\}$ is to divide $[0, K]$ in equally spaced subintervals $[k_m, k_{m+1}]$ and for each subinterval, we compute $s_{k_m}^0$ as the value $s \in [s_{\min}, s_{\max}]$ which maximizes the covariance between $\{f_k\}_{k=k_m}^{k_{m+1}}$ and $\{g_{k+s}\}_{k=k_m}^{k_{m+1}}$. For the rest of points $k \neq k_m$ we initialize $\{s_k^0\}$ using linear interpolation. To get the best result for $\{r_k, s_k\}$ we minimize the energy using both mentioned different strategies for the initialization of $\{s_k^0\}$ and we keep the one providing the lowest value of the energy.

The main parameters of the energy are w_r, w_s, which determine the regularity of the ratio $r(t)$ and of the time shift $s(t)$, the interval $[s_{min}, s_{max}]$ with determines the expected range of values for $s(t)$, and the optional boundary values of $s(t)$ and $r(t)$ at the ends of the time interval $[0, T]$. All of these parameters can be tuned by the user. Unless otherwise indicated, the values of these parameters used in the experiments presented in this work are $w_r = 1000$, $w_s = 10$, $[s_{min}, s_{max}] = [-10, 25]$ and, by default, optional boundary values are not used. The parameter w_o only intervenes in the case where $s(t)$ goes out of the interval $[s_{min}, s_{max}]$. Since an interval $[s_{min}, s_{max}]$, large enough to include the possible values of $s(t)$ is normally used, the term in the energy with w_o is rarely used. We fixed $w_o = 10^{10}$ in the energy minimization implementation.

3 Results

All the experiments presented in this paper can be reproduced using the implementation of the method (named *EpiIndicators*) publicly available at the *EpiInvert* R CRAN package [2]. We used the values of the following time series provided by the World in data organization [20] :

- *new_cases* : new daily confirmed cases of COVID-19
- *new_deaths* : new daily deaths attributed to COVID-19
- *icu_patients* : number of COVID-19 patients in intensive care units (ICUs) on a given day
- *hosp_patients* : number of COVID-19 patients in hospital on a given day
- *weekly_icu_admissions* : number of COVID-19 patients newly admitted to intensive care units (ICUs) in a given week (reporting date and the preceding 6 days)
- *weekly_hosp_admissions* : number of COVID-19 patients newly admitted to hospitals in a given week (reporting date and the preceding 6 days)

The raw registered number of daily cases and deaths are very irregular curves that contain a strong weekly seasonality due to the way countries register the number of cases and deaths every day of the week. To remove the noise and the weekly seasonality of these curves, we use, *EpiInvert*, the method proposed in [1] which estimates smooth trend curves for the incidence and death. In the experiments presented below, we use these restored versions of the incidence and deaths as the epidemic time series.

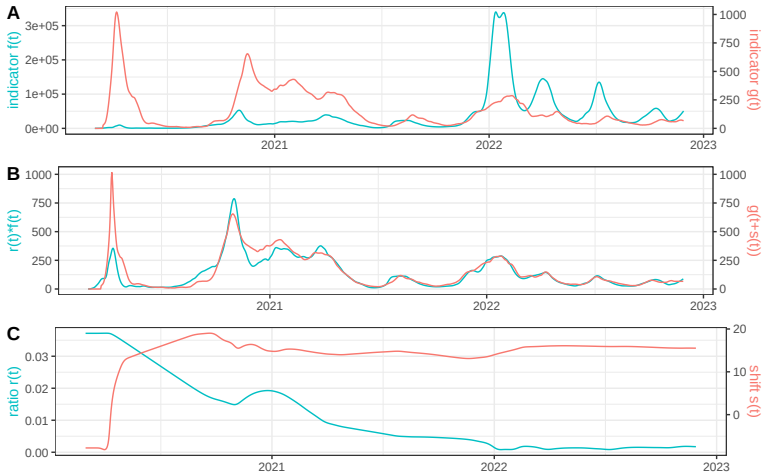

Figure 1: For France, we use as time series $f(t)$: the restored number of daily cases using EpiInvert and $g(t)$: the restored number of daily deaths using EpiInvert. In **(A)** we plot $f(t)$ and $g(t)$. In **(B)** we plot the estimated $r(t)f(t)$ and $g(t+s(t))$. In **(C)** we plot $r(t)$ and $s(t)$.

Comparison of cases and deaths. For France, we compare in Figure 1 the cases and deaths. There is a good agreement between $r(t)f(t)$ and $g(t+s(t))$, except at the beginning of the epidemic where $r(t)f(t)$ is much smaller than $g(t+s(t))$, and $s(t)$ is negative. This is likely due to a very significant underestimation of the number of cases at the beginning of the epidemic. We highlight that the mortality rate, $r(t)$ decreases sharply up to 2022, and then it stabilizes around 0.15%. The delay $s(t)$ increases sharply up to 2021 and then stabilizes around 15 days.

Influence of the weights w_r and w_s. For France again, Figure 2, we compare the cases and deaths when varying w_r and w_s. As expected, decreasing the value of w_r and w_s improves the agreement between $r(t)f(t)$ and $g(t+s(t))$, but $r(t)$ and $s(t)$ are more irregular.

Comparison of hospitalizations and ICU's new admissions. We compare hospitalizations' weekly admissions and ICU's weekly admissions for France in Figure 3. We observe a very good agreement between $r(t)f(t)$ and $g(t+s(t))$. This is likely due to the fact that the quality of these indicators is higher than the quality of the reported cases used above. The maximum of $r(t)$ is attained in 2021 around 22% and in 2022 stabilizes around 9%. The delay $s(t)$ is always small, and at the end of the period stabilizes around 1 day.

Comparison of epidemiological waves between countries. In Figure 4, we now pass to comparing the deaths between Italy and France to estimate the delay between the COVID-19 waves in both countries We know that the epidemy started in Italy before France. This is reflected in that $s(0) \approx 11.8$ days. Then $s(t)$ decreases sharply and stabilizes near zero, indicating that there is currently only a short delay between the epidemic waves in France and Italy. We also know that the first wave of the epidemy in Italy generated a large number of deaths; this is reflected in that $r(0) \approx 1.12$. The median of $r(t)$ for the whole period is 0.885, which is a value very close to the ratio between the populations of Italy and France (around 0.875), which indicates

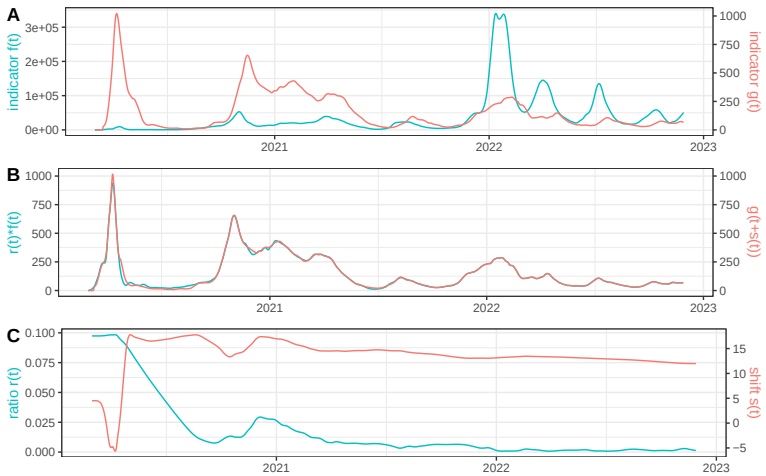

Figure 2: Same experiment as in Figure 1 but using $w_r = 10$ (instead of the default value $w_r = 1000$) and $w_s = 1$ (instead of the default value $w_s = 10$).

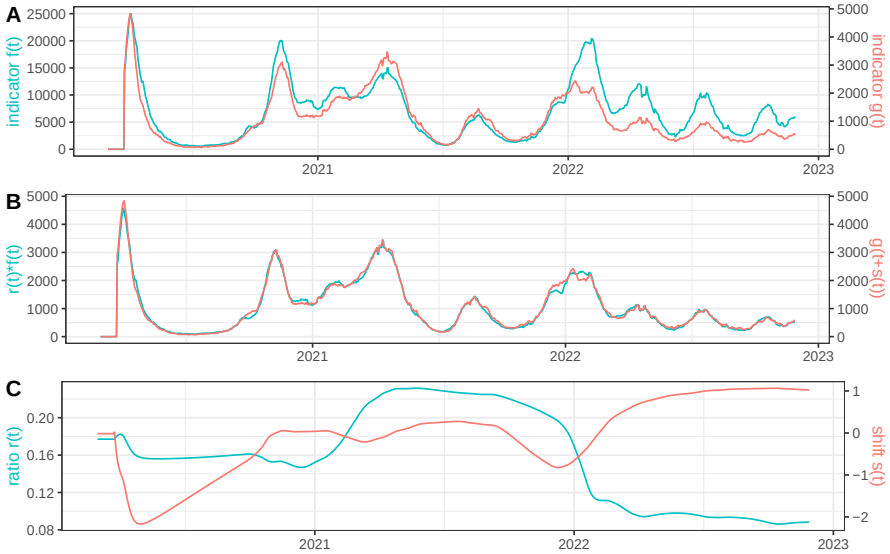

Figure 3: For France, we use as time series $f(t)$: the weekly hospitalization admission and $g(t)$: the weekly ICU's admission. In **(A)** we plot $f(t)$ and $g(t)$. In **(B)** we plot the estimated $r(t)f(t)$ and $g(t + s(t))$. In **(C)** we plot $r(t)$ and $s(t)$.

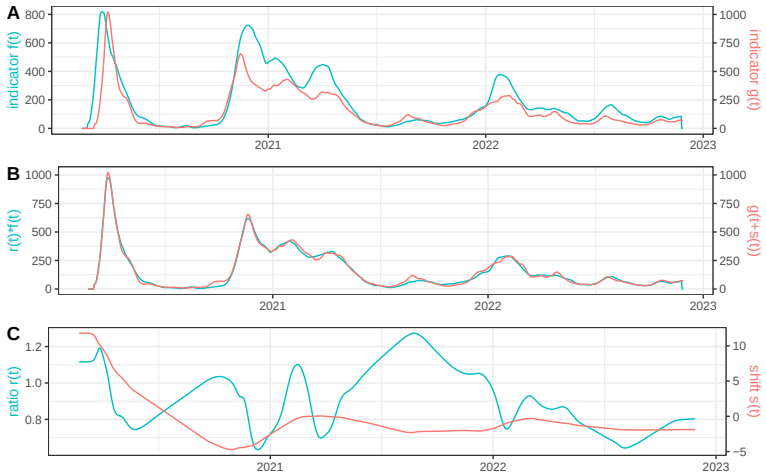

Figure 4: We use as time series $f(t)$: the restored number of daily deaths in Italy using EpiInvert and $g(t)$: the restored number of daily deaths in France using EpiInvert. In **(A)** we plot $f(t)$ and $g(t)$. In **(B)** we plot the estimated $r(t)f(t)$ and $g(t+s(t))$. In **(C)** we plot $r(t)$ and $s(t)$.

that, globally, the mortality rate in both countries has been similar.

Comparing the result of *EpiInvert* with the smooth values of cases provided by [Our World in data]. The [Our World in data][1] dataset includes smooth values of the cases and deaths that we have stored in the **owid** dataset included in the EpiInvert CRAN package [2]. In Figure 5 we compare this smooth data with the one provided by *EpiInvert* for the new cases in France. We observe that the [Our World in data][2] smooth version of cases is more irregular than *EpiInvert*, it can be zero sometimes and the median of the delay $s(t)$ is about 1.81 days. So, in general, to smooth the raw values of cases and deaths, it is better to use *EpiInvert* instead of the [Our World in data] smooth data.

Verifying the model by a simulation study. To evaluate the capacity of *EpiIndicators* to correctly estimate $r(t)$ and $s(t)$ we developed a realistic simulation framework where both functions were known. We used as second time serie $g(t)$ the restored number of deaths in France using *EpiInvert* and we generated the first time serie $f(t)$ from simulated $r(t)$ and $s(t)$ using the formula

$$f(t) = \frac{g(t+s(t))}{r(t)}. \tag{8}$$

in this way, we obtained a perfect matching between both curves. The function $r(t)$ was obtained using 'atan()' type functions where we simulated a decreasing behavior for $r(t)$, $s(t)$ was simulated in the same way with an increasing behavior. After simulating $f(t)$ and $g(t)$ in that way with a "ground truth" $(s(t), r(t))$, we applied *EpiIndicators* to the pair $f(t), g(t)$ and compared the obtained values of $r(t)$ and $s(t)$ with the ground truth. The results are shown in Figures 6 and 7. We observe a good agreement between the estimated values of $r(t)$,

[1] `https://github.com/owid/covid-19-data/tree/master/public/data`
[2] `https://github.com/owid/covid-19-data/tree/master/public/data`

Figure 5: We use as time series $f(t)$: the restored number of daily cases in France using EpiInvert and $g(t)$: the smooth values of daily cases provided by [Our World in data][4]. In **(A)** we plot $f(t)$ and $g(t)$. In **(B)** we plot the estimated $r(t)f(t)$ and $g(t+s(t))$. In **(C)** we plot $r(t)$ and $s(t)$.

$s(t)$, and the ground truth. Notice that we cannot expect the values to be exactly equal, because in the areas where $g(t+s(t))$ is very small or constant, there is not enough information to compute $s(t)$. In that case, the energy regularity term dominates, causing the estimated value to be slightly smoother than the ground truth. On the contrary, at the peaks of the epidemic waves, the information is more robust and we can expect the calculation of $r(t)$ and $s(t)$ to be more accurate.

Next, we studied, using these simulated data, the influence of the order in which the time series are taken. Let $\tilde{r}(t)$ and $\tilde{s}(t)$ be the ratio and delay obtained using $g(t)$ as the first curve and $f(t)$ as the second curve. Then, accordingly with our model, we have that

$$\tilde{r}(t)g(t) \approx f(t+\tilde{s}(t)) \approx \frac{g(t+\tilde{s}(t)+s(t+\tilde{s}(t)))}{r(t+\tilde{s}(t))} \tag{9}$$

To check if this relation is satisfied we studied the normalized error distribution

$$\frac{r(t+\tilde{s}(t))\tilde{r}(t)g(t) - g(t+\tilde{s}(t)+s(t+\tilde{s}(t)))}{median(g)}. \tag{10}$$

The results are shown in Figure 8. The small size of the normalized error (10) shows the performance of the method and its stability when we change the role of the time series. In the beginning, we find larger errors due to boundary effects in the estimations.

A sanity check: verifying the transitivity of the estimations of $r(t)$ and $s(t)$. In this paragraph, we end this study by verifying the transitivity of the estimation when three time series are involved. Let $f_1(t)$, $f_2(t)$ and $f_3(t)$ be these three time series. We took as example the France incidence curve $f_1(t)$ restored using *EpiInvert*; $f_2(t)$ is the number of new hospital admissions and $f_3(t)$ is the number of new deaths restored

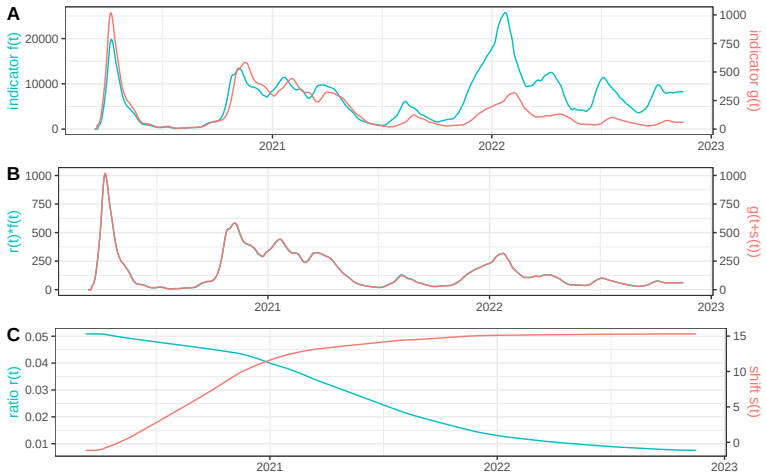

Figure 6: We use as time series $g(t)$ the restored number of deaths in France using *EpiInvert* and $f(t)$ is defined as $f(t)\frac{g(t+s(t))}{r(t)}$ where $s(t)$ and $r(t)$ are a simulated ground truth. In **(A)** we plot $f(t)$ and $g(t)$. In **(B)** we plot the estimated $r(t)f(t)$ and $g(t+s(t))$ using the recovered values for $r(t)$ and $s(t)$. In **(C)** we plot the recovered values for $r(t)$ and $s(t)$.

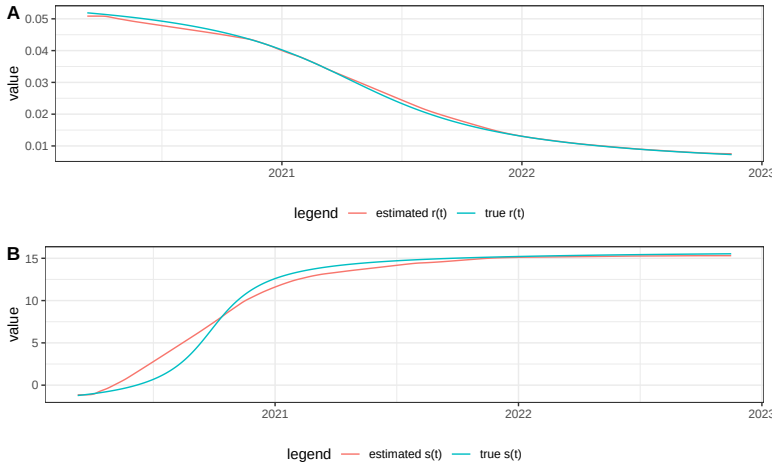

Figure 7: Comparison of the values of $r(t)$ and $s(t)$ estimated by the model and their ground truth using the simulated data.

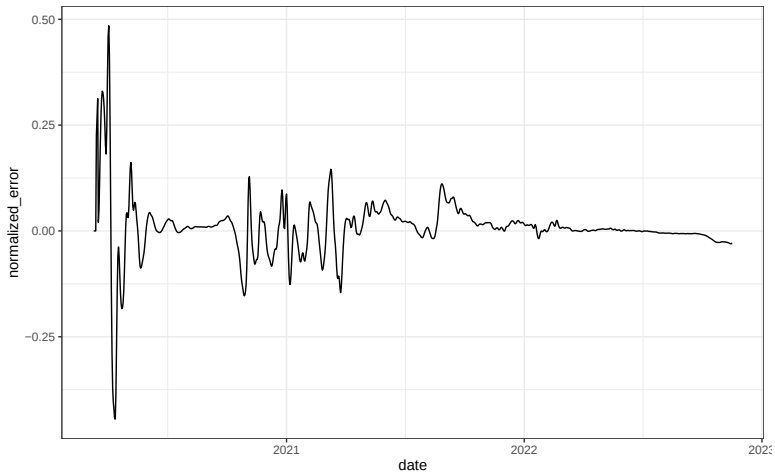

Figure 8: Normalized error (10) using the simulated data.

using *EpiInvert*. The hospitalization data was taken from the COVID-19 European Hub. We will study the relationship between the ratio and delay between the three time series. Let $r_{i,j}(t)$ and $s_{i,j}(t)$ be the ratio and delay between time series $f_i(t)$ and $f_j(t)$. Then we have

$$r_{i,j}(t)f_i(t) \approx f_j(t + s_{i,j}(t)). \tag{11}$$

Using this relation for $\{i, j\} = \{1, 2\}, \{2, 3\}, \{1, 3\}$ we obtain the following transitivity relation:

$$r_{2,3}(t + s_{1,2}(t))r_{1,2}(t)f_1(t) \approx r_{2,3}(t + s_{1,2}(t))f_2(t + s_{1,2}(t)) \approx f_3(t + s_{1,2}(t) + s_{2,3}(t + s_{1,2}(t))); \tag{12}$$

therefore the transitivity condition can be expressed using the equations

$$r_{1,3}(t) \approx r_{2,3}(t + s_{1,2}(t))r_{1,2}(t); \tag{13}$$

$$s_{1,3}(t) \approx s_{1,2}(t) + s_{2,3}(t + s_{1,2}(t)). \tag{14}$$

Next, we check if the relations (13) and (14) were satisfied for the proposed example. To do so we estimated $r_{i,j}(t)$ and $s_{i,j}(t)$ using the proposed method. Then, in Figure 9, we plotted $r_{1,3}(t)$ and $r_{2,3}(t+s_{1,2}(t))r_{1,2}(t)$ in (**A**) and $s_{1,3}(t)$ and $s_{1,2}(t)+s_{2,3}(t+s_{1,2}(t))$ in (**B**). We notice a reasonable transitivity between the estimations of the ratio and delay for the three time series, except at the beginning of the epidemic likely, due to the poor quality of the initial incidence curve.

4 Conclusions

The results of the method are informative and realistic: we found for example blindly a fairly constant and plausible time delay of 16 days between cases and deaths, a very good correspondence between hospitalizations and ICU's time series, and an extraordinarily smooth correspondence between the death curves of France and

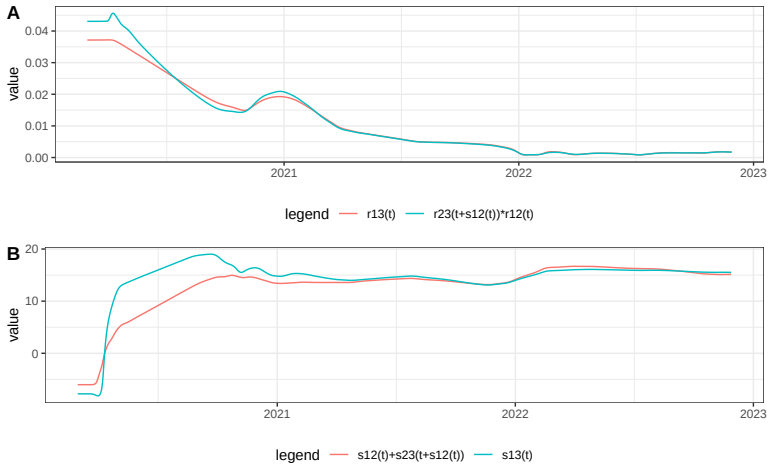

Figure 9: A sanity check: verifying the transitivity of the estimations of $r(t)$ and $s(t)$ when three time series are involved: $f_1(t)$, the France incidence curve, $f_2(t)$, the number of new hospital admissions and $f_3(t)$ the number of new deaths. In (**A**) we plot $r_{1,3}(t)$ and $r_{2,3}(t + s_{1,2}(t))r_{1,2}(t)$ and in (**B**) we plot $s_{1,3}(t)$ and $s_{1,2}(t) + s_{2,3}(t + s_{1,2}(t))$.

Italy. Our method leads to propose generalizations of the renewal equation that enable relating curves of different data, thus generalizing the concept of reproducing number R_t to heterogeneous classes.

The main objection that can be raised by the method is its over-determination. Indeed, there are multiple solutions of (1), some trivial like $s(t) = 0$ and $r(t) = \frac{g(t)}{f(t)}$, and many more. Indeed we can fix an arbitrary and smooth shift $s(t)$, and obtain again a solution pair $r(t), s(t)$ to (1) as $r(t) = \frac{g(t+s(t))}{f(t)}$. Thus, the joint imposition of regularity to $r(t)$ and $s(t)$ is crucial for our purposes. The good results obtained in the simulated study where a realistic ground truth for $r(t), s(t)$ is properly recovered by the method, and the symmetry and transitivity of the method confirm that a sufficient regularity imposition on $r(t)$ and $s(t)$ properly solves the over-determination problem. From the modeling viewpoint, we verified that the behavior of $r(t)$ and $s(t)$ was interpretable and linked to health policies, which is reassuring.

Funding. The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Declaration of Competing Interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Luis Alvarez, Miguel Colom, Jean-David Morel, and Jean-Michel Morel. Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 118(50):1–10, 2021.

[2] Luis Alvarez, Jean-David Morel, and Jean-Michel Morel. EpiInvert R package. *CRAN*, 2022. `https://CRAN.R-project.org/package=EpiInvert`.

[3] Peter Ashcroft, Jana S Huisman, Sonja Lehtinen, Judith A Bouman, Christian L Althaus, Roland R Regoes, and Sebastian Bonhoeffer. COVID-19 infectivity profile correction. *arXiv preprint arXiv:2007.06602*, 2020.

[4] Steven S. Beauchemin and John L. Barron. The computation of optical flow. *ACM computing surveys (CSUR)*, 27(3):433–466, 1995.

[5] Richard Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, 1954.

[6] Carbonell R, Urgelés S, Rodríguez A, et al. Mortality comparison between the first and second/third waves among 3,795 critical COVID-19 patients with pneumonia admitted to the ICU: A multicentre retrospective cohort study. *Lancet Reg Health Eur.*, 11:100243, 2021.

[7] David Clifford, Glenn Stone, Ivan Montoliu, Serge Rezzi, François-Pierre Martin, Philippe Guy, Stephen Bruce, and Sunil Kochhar. Alignment using variable penalty dynamic time warping. *Analytical chemistry*, 81(3):1000–1007, 2009.

[8] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology*, 178(9):1505–1512, 2013.

[9] Krzysztof Dmytrów, Joanna Landmesser, and Beata Bieszk-Stolorz. The connections between covid-19 and the energy commodities prices: Evidence through the dynamic time warping method. *Energies*, 14(13):4024, 2021.

[10] Finelli L., Gupta V., Petigara T., Yu K., Bauer KA., Puzniak LA. Mortality Among US Patients Hospitalized With SARS-CoV-2 Infection in 2020. *JAMA Netw Open*, 4(4), 2021.

[11] Alexia Giannoula, Alba Gutierrez-Sacristán, Álex Bravo, Ferran Sanz, and Laura I Furlong. Identifying temporal patterns in patient disease trajectories using dynamic time warping: a population-based study. *Scientific reports*, 8(1):1–14, 2018.

[12] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*, 26(5):672–675, 2020.

[13] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.

[14] Qixuan Jin. Time warping clustering for the forecast and analysis of covid-19. In *2020 IEEE MIT Undergraduate Research Technology Conference (URTC)*, pages 1–5. IEEE, 2020.

[15] Eamonn J Keogh and Michael J Pazzani. Derivative dynamic time warping. In *Proceedings of the 2001 SIAM international conference on data mining*, pages 1–11. SIAM, 2001.

[16] Kurtz P, Bastos LSL, Dantas LF, et al. Evolving changes in mortality of 13,301 critically ill adult patients with COVID-19 over 8 months. *Intensive Care Med*, 47(5):538–548, 2021.

[17] Joanna Małgorzata Landmesser et al. The use of the dynamic time warping (dtw) method to describe the covid-19 dynamics in poland. *Oeconomia Copernicana*, 12(3):539–556, 2021.

[18] Lefrancq N., Paireau, J., Nathanaël H. et al. . Evolution of outcomes for patients hospitalised during the first 9 months of the SARS-CoV-2 pandemic in France: A retrospective national surveillance data analysis. *Lancet Reg Health Eur.*, 5:100087, 2021.

[19] Alfred J Lotka. Relation between birth rates and death rates. *Science*, 26(653):21–22, 1907.

[20] Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina, and Max Roser. Coronavirus pandemic (covid-19). *Our World in Data*, 2020. https://ourworldindata.org/coronavirus.

[21] Enric Meinhardt-Llopis and Javier Sánchez. Horn-schunck optical flow with a multi-scale strategy. *Image Processing on line*, 2013.

[22] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[23] Hiroshi Nishiura. Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918–19. *Theoretical Biology and Medical Modelling*, 4(1):20, 2007.

[24] Hiroshi Nishiura and Gerardo Chowell. *The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends*, pages 103–121. Springer Netherlands, Dordrecht, 2009.

[25] Patsarin Rodpothong and Prasert Auewarakul. Viral evolution and transmission effectiveness. *World Journal of Virology*, 1(5):131, 2012.

[26] Roelens M., Martin A., Friker B. et al. Evolution of COVID-19 mortality over time: results from the Swiss hospital surveillance system (CH-SUR). *Swiss Med Wkly*, 151:w30105, 2021.

[27] Pavel Senin. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1-23):40, 2008.

[28] Pablo Sprechmann, Alex Bronstein, Jean-Michel Morel, and Guillermo Sapiro. Audio restoration from multiple copies. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 878–882. IEEE, 2013.

[29] Johannes Stübinger and Lucas Schneider. Epidemiology of coronavirus covid-19: Forecasting the future incidence in different countries. In *Healthcare*, volume 8, page 99. MDPI, 2020.

[30] Ward T. and Johnsen A. Understanding an evolving pandemic: An analysis of the clinical time delay distributions of COVID-19 in the United Kingdom. *PLoS ONE*, 16(10), 2021.

[31] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6):509–516, 2004.

[32] Kongming Wang and Theo Gasser. Alignment of curves by dynamic time warping. *The annals of Statistics*, 25(3):1251–1276, 1997.