Learning from the past: a short term forecast method for the COVID-19 incidence curve

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

[1] Laboratory of Integrative Systems Physiology, Ecole Polytechnique Fédérale de Lausanne, EPFL/IBI/LISP - Station 15, Lausanne, CH-1015, Switzerland ; jean-david.morel@epfl.ch
[2] ENS Paris-Saclay, CNRS, Centre Borelli, Université Paris-Saclay, Gif-sur-Yvette, F-91190, France ; jean-michel.morel@ens-paris-saclay.fr
[3] Departamento de Informática y Sistemas, Universidad de Las Palmas de G.C., Las Palmas de G.C., 35017, Spain; lalvarez@ulpgc.es
* Correspondence: jean-david.morel@epfl.ch;

Simple Summary: Forecasting the short time evolution of the COVID-19 daily incidence is a key issue in the epidemic decision making policy. We propose a machine learning method which forecasts the future values of the daily incidence trend based on the evolution of other incidence trend curves that were similar to the current one in the past. Using comparison performed by the European Covid-19 Forecast Hub with the current state of the art forecast methods, we verify that the proposed global learning method, *EpiLearn*, compares favorably to methods forecasting from a single past curve.

Abstract: The COVID-19 pandemy has created a radically new situation where most countries provide raw measurements of their daily incidence and disclose them in real time. This enables new machine learning forecast strategies where the prediction might no longer be based just on the past values of the current incidence curve, but could take advantage of observations in many countries. We present such a simple global machine learning procedure using all past daily incidence trend curves. Each of the 27,418 COVID-19 incidence trend curves in our database contains the values of 56 consecutive days extracted from observed incidence curves across 61 word regions and countries. Given a current incidence trend curve observed over the past four weeks, its forecast in the next four weeks is computed by matching it with the first four weeks of all samples, and ranking them by their similarity to the query curve. Then the 28 days forecast is obtained by a statistical estimation combining the values of the 28 last observed days in those similar samples. Using comparison performed by the European Covid-19 Forecast Hub with the current state of the art forecast methods, we verify that the proposed global learning method, *EpiLearn*, compares favorably to methods forecasting from a single past curve.

Keywords: Incidence curve ; trend curve ; pandemic ; COVID-19 ; renewal equation ; machine learning ; forecasting

MSC: 92C60 ; 92C55 ; 62M20 ; 65K10

1. Introduction

The COVID-19 epidemic has provided us with information on the evolution of the daily incidence in many different countries and epidemic scenarios. Given the enormous global impact of COVID-19, a large number of researchers have studied the problem of predicting the incidence curve. For example, the European Covid-19 Forecast Hub [1] gathers a variety of prediction models based on many different techniques that we will review in this paper.

These methods observe the past of daily incidence in a given country and forecast its future evolution in the weeks to come. The prediction is generally made for the next four weeks. Most of these methods base their forecast on the observation of only the past values of the current incidence curve, that is, the one that they want to extend towards the future. We introduce in the present paper a prediction method that learns the future of a given incidence trend curve from the past evolution of other many incidence trend curves. In a nutshell, the learning method uses all past incidence trend curves that are similar on 28 consecutive days to the last 28 last days of the trend incidence curve that is to be extended towards the future. To demonstrate the method, we use as learning database a collection of 27,418 COVID-19 past incidence trend curves across 61 world regions and countries. These trend curves are computed by the *EpiInvert* method [2] from the original raw incidence curves communicated by the governments. A raw incidence curve is not the adequate input for forecasting because of its high noise and weekly oscillation. The weekly seasonality depends on each country, thus hindering comparison between raw incidence curves. Trend curves instead, being freed from seasonality and noise, are much more suitable to forecasting. Nevertheless, as we will show later, a daily forecast of the raw incidence can be deduced from its forecasted trend using the estimated seasonality.

Let us denote by $\mathbf{s} = (s_1, s_2, ..., s_{28})$, the last 28 values of the current incidence trend that we want to extrapolate, and by $\mathbf{s}^f = (s^f_1, s^f_2, ..., s^f_{28})$ the forecast for the next 28 days proposed in this work. Each of the 27,418 incidence trend curves in our database contains the values of 56 consecutive days extracted from observed past incidence curves. We predict the evolution of the current incidence trend curve from the median of the 28 last days of the 27,418 database curves, where the median is computed on the 121 most similar curves. The similarity to the query of these candidate curves is measured on its first 28 days, which are matched to the 28 last observed days of the query curve $\mathbf{s}$ that we want to extrapolate. In summary, the 28 future samples $\mathbf{s}^f$ of the current curve $\mathbf{s}$ are obtained as the median of the corresponding days 29 to 56 of the most similar past curves[1].

We also compute empirical confidence intervals for the incidence trend forecast by applying the proposed method to the incidence curves of our database and obtaining a distribution of the forecast error as a function of the number of days passed from the current day (the last day of the used incidence curve). In Fig. 1 we illustrate the results of the proposed method for four countries, using their incidence curves up to May 5, 2022. This figure displays in black the raw input incidence curves, which show a strong weekly periodic bias. In the case of France for example, there is a strong deficit on week-ends compensated by a peak on Mondays. For our prediction, we therefore use a smooth incidence trend curve (in red), that is easier to extend and forecast than the original raw incidence. The usual way to compute a incidence trend curve is to apply a 7 or 14 days sliding average to the original raw incidence, which reduces the weekly effects [3]. In our method, we use the more sophisticated *EpiInvert* method introduced in [2,4] and available as R package [5]. This method is summarily described in section 3.2. Fig. 1 shows in blue the forecast curve, that can be compared to the magenta ground truth that became later available. In light blue, the figure also displays the predicted raw incidence curve where the weekly bias learned by *EpiInvert* in the immediate past is also applied. In these relatively favorable examples, picked from large countries with large incidence and at a time of regular daily measurements, the error between ground truth and prediction seems acceptable. Nevertheless, the error on the fourth week can exceed 25%. This is not surprising, given the high variability of the possible futures illustrated for the same countries and times in Fig. 2. In this introduction we do not present the many alternative forecasting methods. Instead, we review them in detail in the discussion section. In section 2.6, the methods that were publicly available through the European Covid-19 Forecast Hub are quantitatively compared with our method through the unbiased metrics of the hub. Our learning technique is widely different from previous methods introduced in

[1] Alternatively, we also tested a weighted average of all curves instead of the median, but it has a slightly inferior performance.

the literature. We involve no parametric model for the incidence curve. Our method produces a daily forecast of the future, whereas most COVID-19 incidence analysis methods [3,6] aim to forecast the 7-day sliding average of the daily incidence.

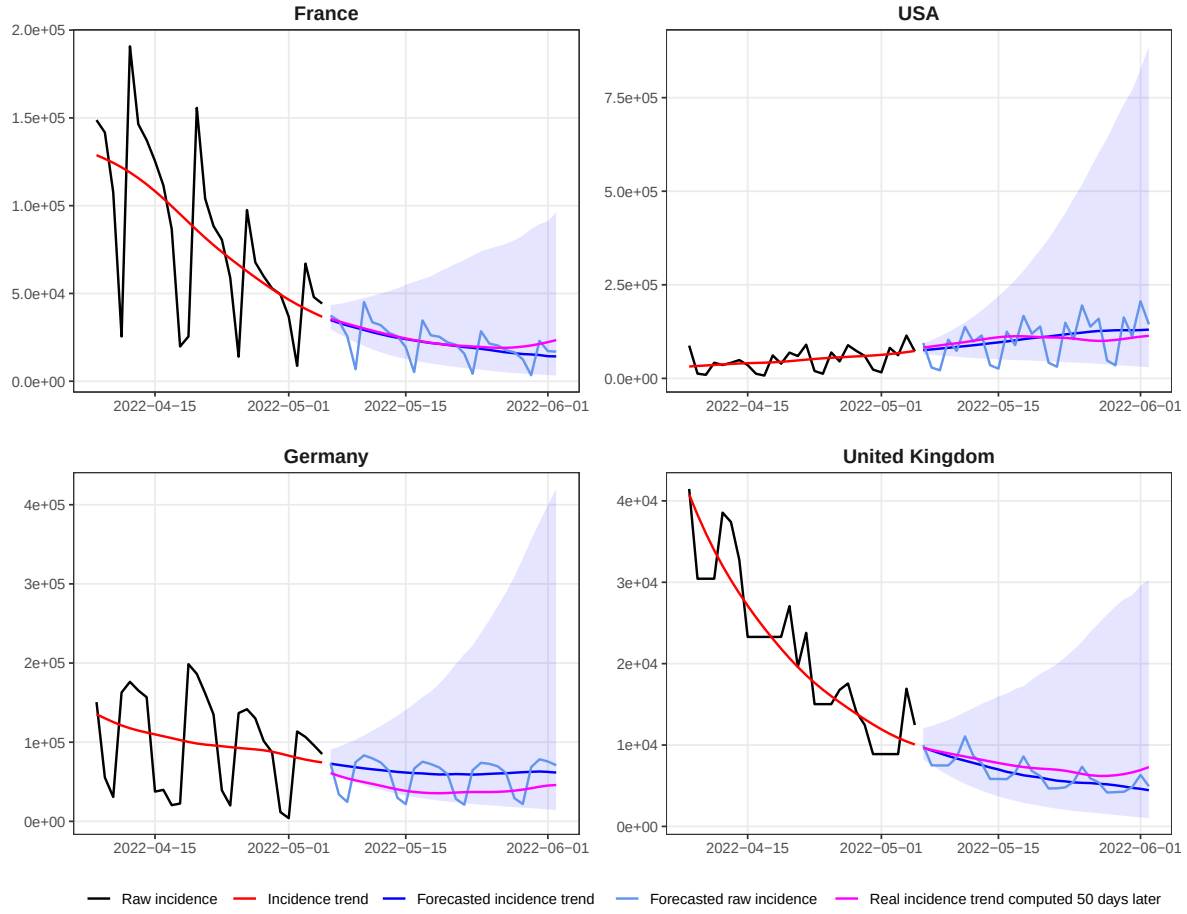

Figure 1. 28-day forecast of the daily incidence for four countries, using the data up to May 5, 2022. In **black** the current original raw incidence curve which suffers from periodic weekly effects. In **red** the current incidence trend computed by *EpiInvert* [2], in **blue** the forecast of the incidence trend curve by *EpiLearn*, in **magenta** the ground truth given by the incidence trend curve obtained 50 days later and in **light blue** the forecast of the raw incidence using equation (5). The **cyan** shaded area represents a 95% empirical confidence interval of the incidence trend forecast. The discontinuity at the past-future junction in Germany is due to a sharp drop of the incidence after the last observed day. When recalculating the incidence trend curve, the values of the past days are also changed by smoothing, thus creating the observed gap.

2. Results

2.1. Incidence trend curves database construction using EpiInvert

Our proposed method, *EpiLearn*, uses a world-wide database of raw incidence curves from 61 countries and regions up to May 5, 2022. For each country or region the last n days of the raw incidence data sequence were iteratively removed (with $n = 0, 1, 2, 3....$). Then, the resulting curve is further processed by applying the *EpiInvert* incidence decomposition algorithm [2] (see section 3.2) and we keep the last 56 values of the estimated incidence trend curve. To add a curve of this type to the database, two conditions were imposed: the first was that the minimum time interval of the resulting sequence to apply *EpiInvert* was 150 days. The second condition was that the mean of the 56 values of

the sequence must be larger than 1000. (Small averages can correspond to nonthreatening or neglected stages of the epidemic, and the resulting incidence curves are often unreliable). Using this procedure we built a database of 27,418 incidence trend curves.

2.2. Normalization of the database incidence curves

EpiInvert is magnitude-invariant, that is, multiplying the raw incidence values by a scalar factor multiplies the estimated EpinInvert incidence trend values by the same scalar factor. Our forecast method preserves this magnitude-invariance by normalizing the magnitude of the incidence trend curves.

Let N be the number of incidence trend curves stored in the database (in our case N =27,418). For $k = 1, 2, .., N$, let us denote by $\mathbf{i}^k = (i_1^k, i_2^k, ..., i_{56}^k)$ the last 56 days of the incidence trend curve computed by *EpiInvert* and stored in the database. Each $\mathbf{i}^k$ has been normalized by multiplying it by a scale factor so that the average of the first 28 values be equal to 1:

$$\frac{\sum_{m=1}^{28} i_m^k}{28} = 1. \tag{1}$$

2.2.1. Computing the distance between curves

We denote by $\hat{\mathbf{s}}$ the present-day incidence trend curve for the country being predicted, that has been normalized in the same way, so that

$$\hat{\mathbf{s}} = \frac{28}{\sum_{m=1}^{28} s_m}\mathbf{s}. \tag{2}$$

We compare the normalized vectors $\hat{\mathbf{s}}$ and $\mathbf{i}^k$ through the following magnitude-invariant distance average,

$$d(\hat{\mathbf{s}}, \mathbf{i}^k, \mu) = \frac{\sum_{m=1}^{28} e^{-\mu(28-m)} |\hat{s}_m - i_m^k|}{28}, \tag{3}$$

where the parameter $\mu \geq 0$ governs the exponentially weighted moving average. The larger the value of μ, the lower this weight for the more remote days, as is classical in control theory [7] and in epidemiological forecasting [8].

2.3. Forecasting using a median of the closest database curves

First, we select in the database the N_{median}, curves $\mathbf{i}^{k_n}_{n=1,..,N_{median}}$ that are closest to the current one, using the similarity criterium (3). N_{median} is a parameter of the method. The median forecast of $\mathbf{s}^f = (s_1^f, s_2^f, ..., s_{28}^f)$ for the next 28 days is defined by

$$s_m^f = median \left\{ \frac{s_{28}}{i_{28}^{k_n}} i_{m+28}^{k_n} \right\}_{n=1,..,N_{median}} \qquad \text{for } m = 1, .., 28. \tag{4}$$

As *EpiInvert* also computes multiplicative weekly seasonality correction factors, q_t, we additionally compute a forecast, $\mathbf{s}^{0,f}$, of the raw incidence curve, $\mathbf{s}^0$, by dividing the forecasted incidence trend curve by the corresponding seasonality factors,

$$s_m^{0,f} = \frac{s_m^f}{q_{22+m\%7}} \qquad \text{for } m = 1, .., 28. \tag{5}$$

By using $q_{22+m\%7}$ as future seasonality factors we are simply making a 7-periodic extrapolation of the last seasonality factors estimated by *EpiInvert*.

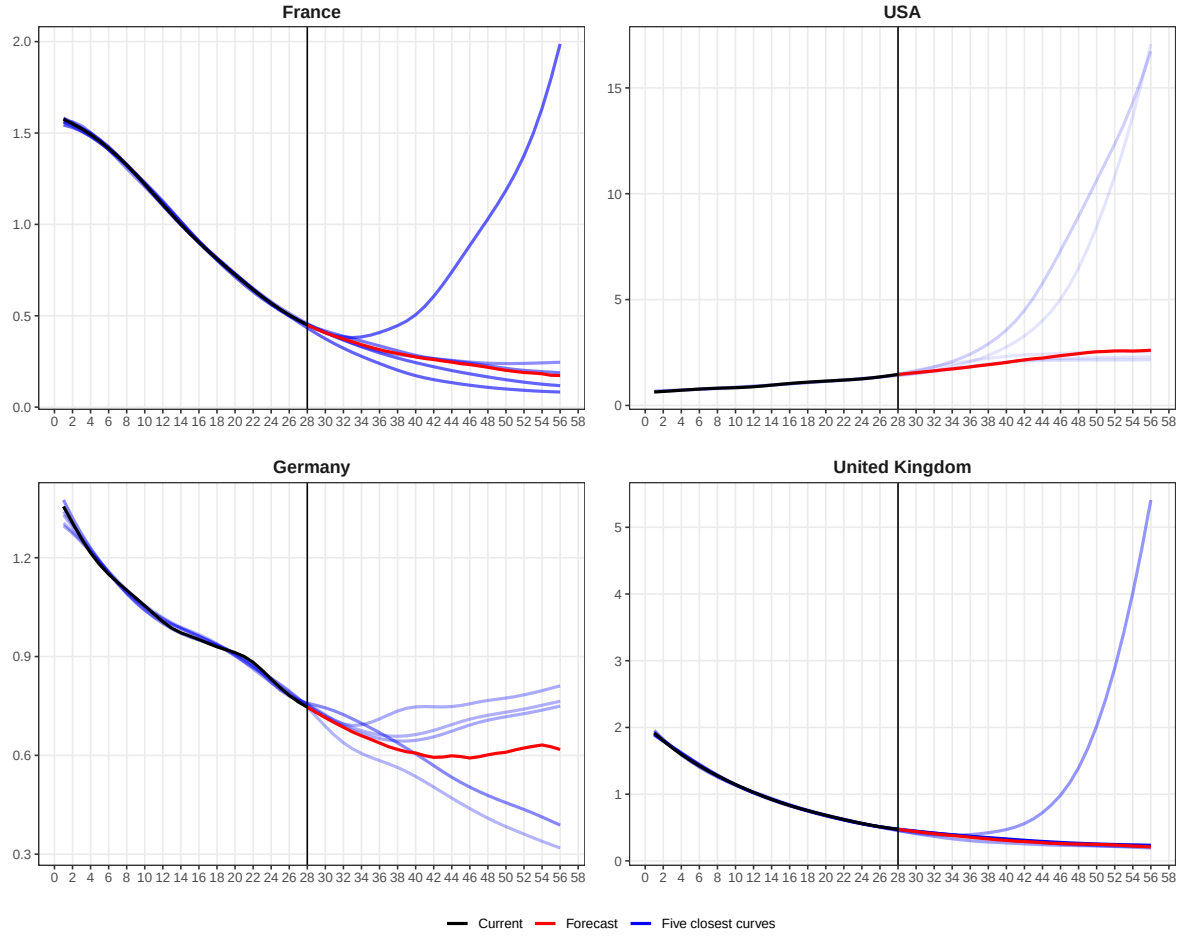

Figure 2. For France, the USA, Germany and the United Kingdom: in **black**, the normalized curve $\hat{\mathbf{s}}$ of the last 28 values of the incidence trend curve up to May 5, 2022, in red the normalized forecasting curve $\hat{\mathbf{u}}$ obtained by *EpiLearn*. Are also displayed in a blue scale the five curves $\mathbf{i}^k$ in the database with the lowest distance $d(\hat{\mathbf{s}}, \mathbf{i}^k, \mu)$ to the incidence trend curve $\hat{\mathbf{s}}$. The lighter the blue, the larger the distance to the input curve.

Figure 2 illustrates the proposed learning procedure. For four countries, it shows the current incidence trend by EpinInvert, its 5 closest curves in the database for their first four weeks, and the forecast, computed as the median of the 121 closest curves in their last four weeks. For France, the UK and the USA, we can observe that among the most similar curves there are curves with a strong growth. These curves correspond to the first wave of the omicron variant in Romania, Hungary and Italy that occurred by the end of 2021. These examples show that very close curves in the past can evolve very differently in the future. In particular, the methods studied in this paper, which forecast the evolution of the incidence only using past incidence data, may be subject to large errors in forecasting.

2.4. Choice of the method parameters

We have to choose N_{median} and μ. Set, for each $\mathbf{i}^k$ in the database,

$$\hat{u}^k_m(N_{median}, \mu) = median \left\{ \frac{\hat{s}^k_{28}}{i^{k_n}_{28}} i^{k_n}_{m+28} \right\}_{n=1,..,N_{median}} \quad \text{for } m = 1, .., 28, \tag{6}$$

where $\mathbf{i}^{k_n}$ are the N_{median} closest curves, in the database, to $\mathbf{s}^k$ (removing from the choice a neighborhood of k in the database). For each forecast day $m = 1, .., 28$, the relative forecast error is given by

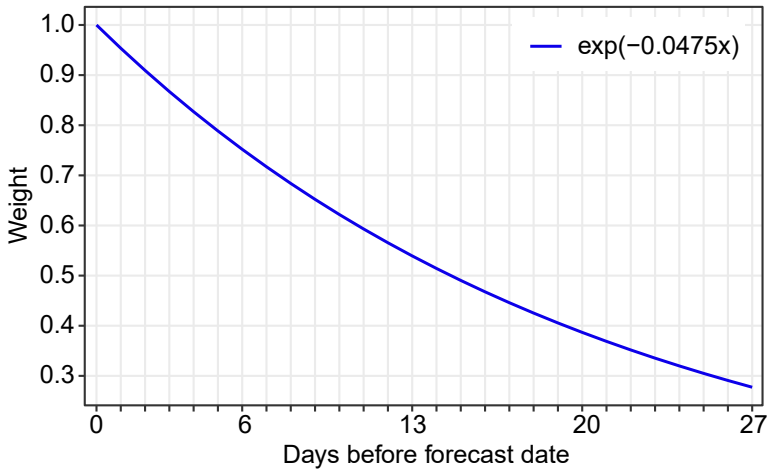

Figure 3. Shape of the functions $e^{-0.0475x}$ which determines the weight assigned to each day in the past in the distance estimation (3) for the proposed forecast method.

$$e^{m,k}(N_{median},\mu) = \frac{|\hat{u}^k_{m+28}(N_{median},\mu) - i^k_{m+28}|}{\hat{u}^k_{m+28}(N_{median},\mu)}. \tag{7}$$

We define the method's median forecast error by

$$\text{ForecastError}(N_{median},\mu) = median\left\{\frac{1}{14}\sum_{m=1}^{14} e^{m,k}(N_{median},\mu)\right\}_{k=1,..,N}. \tag{8}$$

By minimizing this median error, we obtained the optimal values $N_{median} = 121$ and $\mu = 0.0475$. Fig. 3 shows the function $f(x) = e^{-0.0475x}$ which determines the weight assigned to each day in the past in the distance estimation.

2.5. Empirical confidence intervals.

For each curve, $\mathbf{i}^k$, of the database, we denote by $\hat{\mathbf{u}}^k$ the approximation of $\mathbf{i}^k$ obtained by applying the proposed forecasting method (using the optimal values of the parameters) to the curve $\mathbf{s}^k$ given by the incidence trend curve obtained 28 days before $\mathbf{i}^k$ was computed. Moreover, to eliminate the bias due to the fact that $\mathbf{i}^k$ is also in the database, when comparing $\mathbf{s}^k$ with the rest of vectors $\mathbf{i}^n$ in the database we remove a 15-day neighborhood of $\mathbf{i}^k$ in the database. For each forecast day m, we shall study the distributions of relative errors, $E_m = \{e_{m,k}\}_{k=1}^N$ given by

$$e_{m,k} = \frac{i^k_{m+28} - \hat{u}^k_{m+28}}{\hat{u}^k_{m+28}} \qquad \text{for } m = 1,..,28. \tag{9}$$

Assuming that the distribution of the relative forecast error for the current incidence trend curve $\mathbf{s}$ is similar to the one obtained for the database and determined by E_m, we can empirically approximate the percentiles of the forecast distribution, F_m, of the current curve, using the percentiles of E_m. Indeed, let us denote by $P_p(X)$ the p-th percentile of a distribution X, then

$$P_p(F_m) \approx s^f_m + s^f_m P_p(E_m) \tag{10}$$

where s^f_m is the forecast estimated by the proposed method. A 95% central confidence interval for the incidence trend value is for example given by $(P_{0.025}(F_m), P_{0.0975}(F_m))$. In Fig. 4 we display the

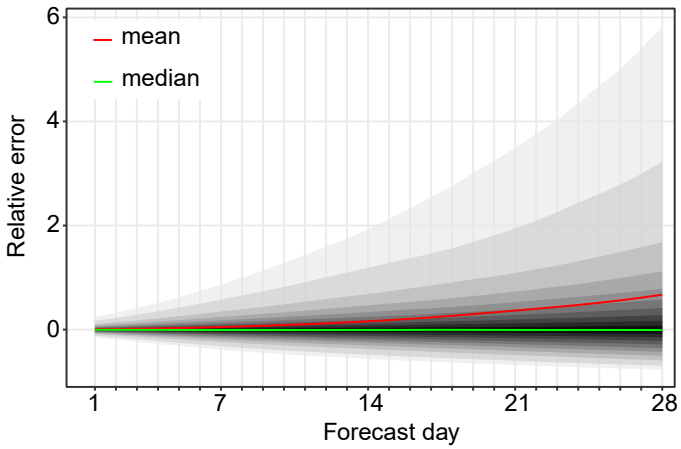

Figure 4. Illustration of some statistics of the $E_m = \{e_{m,k}\}$ distribution defined by (9) for the entire database: the **red** curve indicates the mean of the distribution that is greatly affected by the skewness of the distribution, which justifies using the median (the curve in **green**) instead of the mean. The median is indeed very close to zero, which proves the consistency of the approximation adopted in equation (10). From the outside to the inside, the shaded areas represent the estimated $(1-\alpha_k)\times 100\%$ central prediction intervals $(l_{\alpha_k}, u_{\alpha_k})$ for $\alpha_k = 0.05, 0.1, 0.2, ..., 0.9$.

confidence intervals of E_m for the proposed forecast method. As expected, the size of the confidence intervals increases with the forecast day m and is quite large after 28 days. Notice that the mean and the median $(P_{0.50}(E_m))$ of E_m are very different due to the asymmetry of the distribution E_m. The mean is closer to the upper end of the forecast interval than the median. The fact that the median of the error is very close to zero confirms the consistency of the method.

2.6. Comparative results in the context of the European Covid-19 Forecast Hub

The question arises of how to compare all methods, in theory and in practice. For a practical comparison, we take advantage of the fact that a wide variety of forecasts are submitted to the European COVID-19 Forecast Hub [9] and to the COVID-19 Forecast Hub [10]. A study on the methodology to evaluate and compare forecast has been proposed in [11], using the data of this Hub. We shall address the theoretical comparison in section 3. As developed in [1], the European Covid-19 Forecast Hub provides short-term forecasts of Covid-19 cases and deaths across Europe. It is supported by teams working on pandemic modeling and sharing their forecast of the weekly accumulated incidence with horizons of 1 to 4 weeks. Each week starts on Sunday and ends on Saturday. At the time of writing, many countries do not provide data during the week-end, and some countries only provide a weekly estimate. This fact has no influence for method preprocessing the data by a 7 day sliding average. Nevertheless, since we use daily estimates, a single weekly estimate has a negative impact on the quality of our forecast. To address this issue, when a country provides data on a day, but not on the previous days, we distribute equally the last accumulated value over the previous uninformed days before applying *EpiInvert*.

Since *EpiLearn* forecasts the daily incidence, the weekly forecast is obtained by summing the forecasted raw daily incidence given by (5). The quantiles of the associated weekly distributions are computed on the registered database of incidence curves by extending the procedure of section 2.5 which computes the confidence intervals of the forecasted incidence curve.

The European Hub encourages teams to provide, for each model, m, each horizon week, $h = 1, 2, 3, 4$, and each forecast target, n, the prediction of the weekly incidence, $f_{m,h,n}$, and 23 quantiles of the associated distribution. These quantiles correspond to the predictive median, M, and eleven $(1-\alpha_k)\times 100\%$ central prediction intervals $(l_{\alpha_k}, u_{\alpha_k})$, with $\alpha_k = 0.02, 0.05, 0.1, 0.2, ..., 0.9$, where l_{α_k} and

u_{α_k} are (respectively) the $\alpha_k/2$ and $(1-\alpha_k/2)$ quantiles of F. The following weighted interval score, $WIS_{m,h,n}$ (see [12]), is proposed to evaluate the distribution accuracy:

$$WIS_{m,h,n} = \frac{\frac{1}{2}|o_{h,n} - M| + \sum_{k=1}^{11} \frac{\alpha_k}{2}(u_{\alpha_k} - l_{\alpha_k}) + (l_{\alpha_k} - o_{h,n})_+ + (o_{h,n} - u_{\alpha_k})_+}{11.5} \tag{11}$$

where $o_{h,n}$ is the observed outcome, $(.)_+$ is defined as $(x)_+ = x$ if $x > 0$ and 0 otherwise. The lower the value of $WIS_{m,h,n}$, the better the score associated to the forecast distribution determined by the quantiles of F.

The prediction accuracy of a model is measured using two indicators: the first one is $|f_{m,h,n} - o_{h,n}|$, that is, the absolute value of the difference between the observed value $o_{h,n}$ and the prediction $f_{m,h,n}$. The second indicator measures the quality of the confidence intervals and is given by $WIS_{m,h,n}$. To compare the prediction accuracy of different models, we have to take into account that, in general, each team provides a different number of forecast targets. We started for example submitting forecast to the European Hub by August 2022, but other teams started submitting up to 2 years earlier. Furthermore, not all teams provide a forecast for all horizons and for all countries. Thus, defining a fair comparison of models requires some caution. To address this issue, the European Hub uses the following procedure (we explain the procedure for the comparison of $|f_{m,h,n} - o_{h,n}|$, but the comparison of $WIS_{m,h,n}$ is equivalent). Consider two models m and m', a week horizon $h \in \{1,2,3,4\}$ and $\{(f_{m,h,n}, f_{m',h,n}, o_{h,n})\}_{k=1}^{N_{m,m',h}}$ where $N_{m,m',h}$ is the number of forecast targets that have been handled by both models. The pairwise comparison of both models is then defined by the ratio

$$\theta_{m,m',h} = \frac{\sum_{k=1}^{N_{m,m',h}} |f_{m,h,n} - o_{h,n}|}{\sum_{k=1}^{N_{m,m',h}} |f_{m',h,n} - o_{h,n}|}, \tag{12}$$

which is smaller than 1 if m' is more accurate than m, and larger than 1 otherwise. Subsequently, we compute for each model m the geometric mean of the results achieved for all different pairwise comparisons,

$$\theta_{m,h} = \left(\prod_{m'=1}^{M'} \theta_{m,m',h} \right)^{\frac{1}{M'}}, \tag{13}$$

where M' is the number of models, m', which have forecast targets in common with model m. It follows that $\theta_{m,h}$ is a measure of the relative skill of model m with respect to the set of all other models in the week horizon h. The relative performance of model m is computed with respect to $\theta_{b,h}$, the score of the baseline model, as

$$\theta^*_{m,h} = \frac{\theta_{m,h}}{\theta_{b,h}}, \tag{14}$$

where the baseline model b is nothing but the constant prediction extending the last observed weekly value [13].

The ratio $\theta^*_{m,h}$ is called the relative MAE, *rel_ae*, of model m in the week horizon h. A score of $0 < rel\_ae < 1$ means that model m is better than the baseline; a score of $rel\_ae > 1$ means that the baseline is better. In the case of $WIS_{m,h,n}$, we use the same procedure (replacing $|f_{m,h,n} - o_{h,n}|$ by $WIS_{m,h,n}$) and call *rel_wis* the associated indicator. Every week, the Hub publishes the scores *rel_ae* and *rel_wis* for all team models and for the *ensemble* model, which is a forecast estimate obtained by aggregating the predictions of all teams. In each weekly evaluation report, the Hub publishes the scores obtained using the information of the last 10 weeks and the scores obtained using all available data.

Given that we have recently started to submit forecasts to the Hub, we used the 10 week forecast scores. The box plots of Figure 5 give the distribution of *rel_ae* and *rel_wis* for the weeks from 2022-08-29 to 2022-10-31 which correspond to weeks during which our model has been included in

the 10-week European Hub forecast evaluation reports. The eight forecasting methods that are being compared are the only ones that provided reasonable forecasts over the whole time period considered. In this sense, we removed the ILM-EKF model from the comparison because the scores published by the Hub for this model were unreasonable (in some cases its score is greater than 10). In these box plots comparing the models for these dates, the baseline equal to 1 is the constant prediction, which is expected to be beaten by all more sophisticated methods. Yet, the two last displayed methods actually give worse estimates. Among the four methods beating the baseline, *EpiLearn* (named AMM-EpiInvertForecast in the Hub evaluation reports) ranks first for both quality indicators, *rel_ae* and *rel_wis*. In the first week horizon, the Ensemble method is the second best, and in the second, third and fourth week the MUNI-ARIMA method, which is very close to *EpiLearn*, ranks second and the Ensemble method third. The result of this ensemble model is usually considered as the best option as argued in [12]. A great advantage of using the scores published by the European Hub is that such scores cannot be manipulated. They represent a fair quality comparison framework for the models performance.

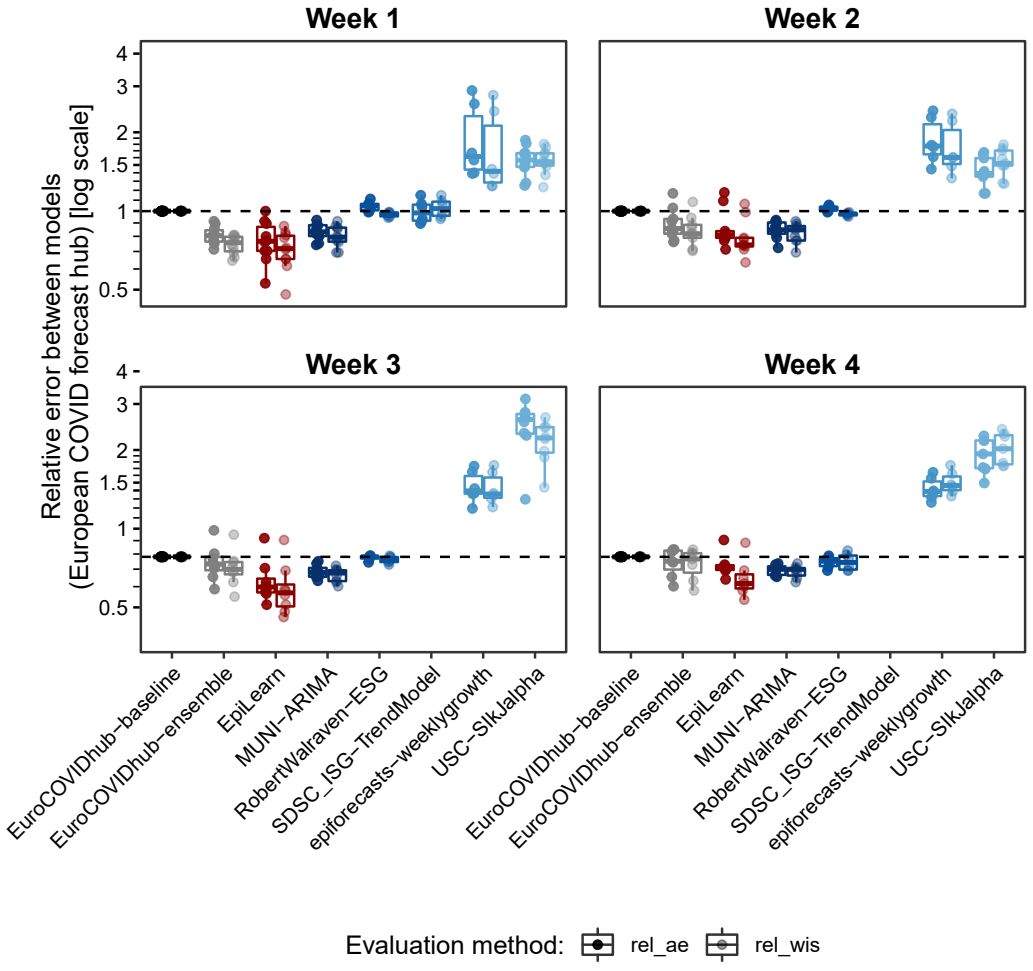

Figure 5. Box plots of the European Hub *rel_ae* and *rel_wis* overall model scores with target horizons 1, 2, 3 and 4 weeks published by the European Hub using the weekly evaluation reports from 2022-08-29 to 2022-10-31. It is the period during which *EpiLearn* (named AMM-EpiInvertForecast in the Hub) was included in the weekly evaluation reports.

3. Discussion

In this section we review and discuss the properties and assumptions of the most relevant forecasting methods, and link them when possible to methods and results published weekly in the European Covid-19 Forecast Hub [9].

3.1. ARIMA

The ARMA (AutoRegressive Moving Average) and ARIMA (AutoRegressive Integrated Moving Average) models are the backbone of many forecasting methods and are implemented through the popular R package [14], with an automatic selection mode of the best parameters, usually by the Akaike information criterion (AIC). The ARMA model with parameters (p, q) can be written

$$I_t - \alpha_1 I_{t-1} - \cdots - \alpha_p I_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where parameters p and q are non-negative integers, p is the order (number of time lags) of the autoregressive model and q is the order of the moving-average model. The ARMA models are often written compactly using the lag operator $LI_t := I_{t-1}$ as

$$(1 - \alpha_1 L - \alpha_2 L^2 \cdots - \alpha_p L^p) I_t = (1 + \theta_1 L + \cdots + \theta_q L^q) \varepsilon_t.$$

The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean. In the the European Hub Forecast initiative, the *epiforecasts-weeklygrowth* team [15] uses a Bayesian AR model using weekly incidence data. Both the incidence and the growth rate are assumed to be AR(1) processes with the growth rate being differenced and scaled by a decay parameter". The fact that they are differenced implies that they obey an ARIMA model. Its results were illustrated in Figure 5.

The ARIMA model is a specific ARMA model with parameters (p, d, q):

$$(1 - \alpha_1 L - \alpha_2 L^2 \cdots - \alpha_p L^p)(1 - L)^d I_t = (1 + \theta_1 L + \cdots + \theta_q L^q) \varepsilon_t, \tag{15}$$

where d is called degree of differencing, which is the number of times the data have had past values subtracted. If $d \geq 1$ this model removes trend and seasonal structures that negatively affect the regression model. Taking $d = 1$ corresponds to a linear trend and $d = 2$ to a quadratic trend. A thorough description of ARIMA is given in the online book [16].

ARIMA is arguably the most used forecasting model for COVID-19, and has been applied with a country-specific optimization of parameters. For example the MUNI-ARIMA [17] team participating to the European Hub Forecast initiative also uses an "ARIMA model with outlier detection fitted to transformed weekly aggregated series". This method is one of the best performing methods as illustrated in Figure 5. Table 1 details the (p, d, q) parameters, the forecasting period and the model selection method used in 14 articles proposing applications of this forecasting method.

3.1.1. Discussion

In most applications of the ARIMA model we see that $p + d$ ranges from 2 to 5 while the values $q = 0, 1, 2$ are dominant in the empirical models. The values 2 to 6 for $p + d$ cannot be used with raw incidence curves that show a 7 days periodicity. To address this, most models are being applied after a pre-processing of the raw incidence curve, such as a 7-day mean and a subsequent subsampling of the input series. The main feature of Table 1 is that $d = 2$ is the more frequent differentiating parameter, meaning that the trend of the pandemic is assumed to be quadratic, and that stationary variations from this quadratic trend are being estimated. This means that ARIMA (with $d = 2$) is effective only in time intervals where the trend is not changing between concave and convex, and is actually having a

Paper	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]
p	2	1	0-4	0-1	6	9	5-10	1	1	2	0-4	2-3	2	2,6
d	2	2	2	2	1	0	1-2	0	1	1	2	1-2	2	1
q	2	0	1-8	0-1	0	8	2-9	3-4	0	1	1-2	0-1	2	3,7
days	20	7	180	16	10	60	14	21	50	28	50	21	14	7
data	D	D	D	D	MA	D	GF	D	D	D	D	D	D	D
criterion	MAPE	ADF	RMS	MAPE	AIC	AIC	MAE	ADF	BIC	AIC	MAE	AIC	BIC	AIC

Table 1. Parameters of several representative ARIMA models that were used for Covid-19. The orders p and q are in lines 2 and 4, the degree in line 3. The forecast interval length ranges from 2 to 180 days. The orders and degrees of ARIMA models were selected automatically, mostly by the Akaike information criterion (AIC), often complemented by the Bayesian information criterion (BIC) and average errors such as the root mean square error (RMSE), the mean absolute Error (MAE) or the mean absolute percentage error (MAPE). The augmented Dickey-Fuller (ADF) unit-root test is also often used to test the stationarity of the sequence after differencing. Given the small order of most processes, the ARIMA parameters were estimated on a past time series of between 30 and 100 days. The data row indicates if the input time series has been processed. In absence of any information in the paper we assume that "D" (daily raw data) are being processed. Some methods indicate that they preprocess the incidence curve with a moving average of 7 days and others with a Gaussian filter (GF).

constant second derivative, hence a parabolic shape. A quick examination of the examples given in Figure 2 suggests that this assumption is only compatible with a short prediction.

3.2. Seasonal models

An extension of ARIMA, SARIMA (seasonal ARIMA) is a combination of two ARIMA models. Let us rewrite the ARIMA equation (15) as

$$P_p(L)(1-L)^d I_t = Q_q(L)\varepsilon_t,$$

where P_p and Q_q are polynomials of degrees p and q respectively. Then a seasonal ARIMA model writes

$$P_p(L)\Pi_\pi(L^s)(1-L)^d(1-L^s)^\delta = Q_q(L)\Phi_\phi(L^s)\varepsilon_t,$$

where (p,d,q) are ARIMA degrees, (π,δ,ϕ) are the corresponding degrees for the seasonal part, and s is the season length, typically 7, 30 or 365. This method was tested for forecasting the global COVID-19 incidence in [23], with p ranging from 6 to 9, $d=0$, q ranging from 0 to 8, $\pi=\delta=0$, ϕ ranging from 1 to 2, $s=3,7,12$. Surprisingly, the displayed experimental predictions show no seasonal oscillation.

This method can be put in a more general framework, as done in an exemplary treatise [16] dedicated to forecasting. The chapter of this volume dedicated to the Holt-Winters method shows recursive equations with a seasonal component which can be additive or multiplicative. This method comprises a forecast equation and three smoothing equations — one for a level variable, a second one for the trend and one for the seasonal component. It has been used and compared to ARIMA for COVID-19 forecast in [32].

The *EpiInvert* method [2,4,5] handles the COVID-19 incidence curves in a similar setting by decomposing them into three terms: a) a quasi-periodic multiplicative term with a weekly period, b) an additive white noise term and d) a smooth incidence trend, which obeys closely the renewal equation. The renewal equation can be viewed as an ARMA model dictated by the pandemy's reproducing kernel, also called *time interval*.

EpiInvert proceeds by a variational technique which estimates the weekly seasonal bias, the time varying reproduction number R_t, and finally a smooth incidence trend curve constrained to satisfy

closely the pandemy's renewal equation. The method computes trend values of the incidence up to the present days, which gives it a $\simeq 3$ days in advance over other methods based on a sliding average[2].

3.3. Compartmental epidemiological models (SIR, SEIR, SIRD, SEIARD and SUIHTER)

Compartmental models are *in silico* simulation models that consider the population as a collection of compartments, for example in the case of SEIARD : S susceptible, E exposed, I infected, A asymptomatic, R recovered and D dead. In the simplest form, the SIR model, writes

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta IS}{N}, \\ \dfrac{dI}{dt} = \dfrac{\beta IS}{N} - \gamma I, \\ \dfrac{dR}{dt} = \gamma I, \end{cases} \tag{16}$$

where S is the stock of susceptible population, I is the stock of infected, R is the stock of removed population (either by death or recovery), and N is the sum of these three. The ratio $R_0 = \frac{\beta}{\gamma}$ is the basic reproduction number. Initially designed for epidemic modeling, the SIR model and its variants have since been adapted to forecasting the future evolution of the pandemic from an estimated starting point. The model's parameters are estimated for the past incidence, and the model is then applied forward to simulate the feature. This method has been developed for SIR [33,34], SEIR [35,36], SIRD [37], SEIARD [38] and SUIHTER [39,40].

3.3.1. Discussion

In the SIR model (16) and its variants applied over a short period of time (a few weeks), many of the modeled compartments are either constant over a short period of time, or can be inferred from the incidence with a time delay. For example, since most forecasts predict at most a month of the epidemic, the number S of susceptible people may be considered as nearly constant in the second equation of (16), namely $S = S_0$. Thus, the model boils down to the ODE

$$\frac{dI}{dt} = \frac{\beta IS_0}{N} - \gamma I, \tag{17}$$

and its basic reproduction number is $R_0 = \frac{\beta}{\gamma}$. This observation holds true with more complex models such as SEIR, SEIRD, SEIARD, SUIHTER: Indeed, the dependence of I from the other compartments occurs through S. In short, when neglecting the variation of S on a learning and forecast period which does not exceed several weeks, all models are equivalent to the linear "I-model" (17). This model depends on a single parameter β. Indeed, the other parameter γ is linked to the duration of the infectious period, which is constant for a given virus type. Thus, using a SIR model for a few weeks forecasting implies that R is implicitly kept constant. Note that in all mentioned models, the evolution of I is governed by the first two equations of the SIR model (16) only. Adding more compartments makes the model more complete, but does not change the behavior of the incidence.

3.4. Regression models

In [41] the authors generate short-term forecasts in real-time using three phenomenological models that have been previously used to derive short-term forecasts for a number of epidemics for several infectious diseases [42]. The generalized logistic growth model (GLM) extends the simple

---

[2] *EpiInvert* method can be run using the R package [5], or in real-time for any country in the world at https://ipolcore.ipol.im/demo/clientApp/demo.html?id=77777000032

logistic growth model to accommodate sub-exponential growth dynamics with a scaling of growth parameter, p. The Richards model also includes a scaling parameter, a to allow for deviation from the symmetric logistic curve. The authors also include a recently developed sub-epidemic wave model that supports complex epidemic trajectories, including multiple peaks. In this approach, the observed reported curve is assumed to be the aggregate of multiple trajectories. We describe briefly in the sequel these regression models and mention other papers using them for COVID-19 forecasting.

The simple logistic growth model (SLM) is a common S-shaped curve sometimes called sigmoid function with equation

$$N(t) = \frac{K}{1 + e^{-r(t-t_0)}},$$

where t_0 is the sigmoid's midpoint, K is the curve's maximum value, and r is the logistic growth rate or steepness of the curve. This model describes the curve of cumulative cases. The corresponding incidence curve is its derivative $i(t) = N'(t)$ which satisfies

$$i(t) = N'(t) = rN\left(1 - \frac{N(t)}{K}\right).$$

The generalized logistic growth model (GLM) is an extension of the simple logistic growth model that includes an additional parameter, p, to allow for *scaling of growth*; $p = 1$ indicates early exponential growth, $p = 0$ represents constant growth, and $0 < p < 1$ accommodates early sub-exponential or polynomial growth. The GLM is defined by the differential equation

$$i(t) = N^{'}(t) = rN(t)^p(1 - \frac{N(t)}{K})$$

where $N(t)$ represents the cumulative number of cases at time t, r is the growth rate, p is the scaling of growth parameter, and K is the carrying capacity or final epidemic size.

The Richards model

The Richards model [43] is a 2-parameter extension of the simple logistic growth model including a scaling parameter and defined by

$$i(t) = N'(t) = rN(t)\left(1 - (\frac{N(t)}{K})^{a}\right)$$

where $N(t)$ represents the cumulative number of cases at time t, r is the growth rate, K is the final epidemic size, and the exponent a measures the deviation from the symmetric s-shaped dynamics of the simple logistic curve. This model is used in [44] as a parametric regression model for the modeling of incidence indicators. The incidence distribution is modeled by an appropriate Poisson or Negative Binomial. It is also used in [45] for estimating the regional propagation of COVID-19 in Italy and in [46] for recurrent forecasting in Europe.

The Gompertz model is used for COVID-19 forecast in [47] and [48]. The Gompertz model models the cumulative cases of Covid-19. It was originally proposed to explain human mortality curves has been further employed in the description of growth processes. The Gompertz equation reads

$$N(t) = Ke^{-\log\left(\frac{K}{N_0}\right)e^{-at}},$$

where the parameter K is the final number of cases, N_0 the initial number of cases, and parameter a is the rate of decrease in the initially exponential growth. Since the Gompertz function shows the cumulative cases its temporal derivative

$$i(t) = N'(t) = aKe^{-\log\left(\frac{K}{N_0}\right)e^{-at}} \log\left(\frac{K}{N_0}\right) e^{-at}$$

gives an estimate of the incidence curve. The fitting of the Gompertz function to the data can be done by the minimum least squares method. This amounts to make a forecast based of the growth parameter a. For its forecast, the BIOCOMSC-Gompertz [47] team participating to the European Hub Forecast initiative fits the Gompertz model to past data.

Composite models The composite logistic growth model (CLM) [41] can be written as

$$N(t) = \sum_{i=1}^{p} \frac{K_i}{1 + A_i e^{-r_i(t-\tau_i)}},$$

where $N(t)$ is the cumulative number of cases, the number of waves is p, and the four parameters $((K_i, A_i, r_i, \tau_i)$ for each wave are estimated by minimization of the objective function, which is the sum of squares of residuals. The RobertWalraven-ESG [49] team participating to the European Hub Forecast iniative uses a variant of CLM, making a "multiple skewed Gaussian distribution peaks fit to raw data" where the skewed Gaussians have the form

$$i(t) = Ae^{-\frac{(t-t_0)^2}{(c(1+d(t-t_0)))^2}}.$$

Its results were illustrated in Figure 5.

The sub-epidemic model

The most flexible extension of the previous models used for forecasting [41] is the sub-epidemic wave model which supports complex epidemic trajectories by shaped by multiple underlying sub-epidemics modeled by the GLM, where the growth rate r and scaling parameter p are the same across sub-epidemics. An epidemic wave is composed of n overlapping sub-epidemics as follows: $N_i'(t) = rA_{i-1}(t)N_i(t)^p\left(1 - \frac{N_i(t)}{K_i}\right)$, where $N_i(t)$ is the cumulative number of infections for sub-epidemic i, and K_i is the size of the i^{th} sub-epidemic ($i = 1, \ldots, n$).

3.4.1. Discussion

All of the above simple regression non-composite models depend on one or two growth parameters and can only be used for a short term forecast. The Gompertz model actually adds one more degree of freedom. So it can learn more from the past, and therefore can be used for a larger forecast period. The last two composite models use a global modeling of a long interval including several waves but, since each term has exponential decay, forecasting will in practice be based on the two parameters of the last term, which limits their ability to learn from the past incidence.

3.5. Short term prediction by the renewal equation, linear extrapolation

The approach proposed in [50] to forecasting future COVID-19 cases involves 1) modeling the observed incidence cases using a Poisson distribution for the daily incidence number, and a gamma distribution for the series interval; 2) estimating the effective reproduction number assuming its value stays constant during a short time interval (by the EpiEstim method [3]); and 3) using the renewal equation, drawing future incidence cases from their posterior distributions, assuming that the current transmission rate will stay the same, or change by a certain degree.

A similar forecast method is involved in [51] which compares human and machine forecasts in Germany and Poland. The authors use a Bayesian model from the EpiNow2 R package (version 1.3.3) to predict reported cases. Epinow [6] estimates the effective reproduction number R_t. The future infections are computed by the Fraser renewal equation as a weighted sum of past infection multiplied by R_t. In the comparison, R_t is assumed to stay constant beyond the forecast date. The conclusion of this paper is that an average of human experts' forecasts performs better. Similarly, the USC-SIkJalpha [8] and ILM-EKF [52] teams participating to the European Hub Forecast initiative use the renewal

equation (the second mentioned groupe also involves a Kalman filter in its prediction). Its results were illustrated in Figure 5.

We mentioned *EpiInvert* in section 3.2 as a similar method using the renewal equation as forcing term to restore a coherent incidence curve after compensation of the weekly biases.

Last but not least, the SDSC_ISG-TrendModel [53] team, also participating to the European Hub Forecast initiative, is a trend extrapolation which starts, like *EpiInvert*, by decomposing the incidence curve into three components: the trend, a seasonal component and noise. Then the model predicts daily cases using linear extrapolation on the linear or log scale of the underlying trend estimated by a robust LOESS seasonal-trend decomposition model. Its results were illustrated in Figure 5, where only the results for the first two weeks are available.

3.5.1. Discussion

Like for the logistic models, these extrapolation models have the merit of simplicity. Also, they avoid estimating reproducing parameters by imposing those of the renewal equation. Yet, when used to extrapolate, they maintain a constant growth parameter (the reproducing number R_t for its forecast. Hence, they are not adapted to forecasting in intervals where a change of trend is likely to happen.

3.6. Aggregation of estimators

The idea of aggregation methods, sometimes also called *ensemble* methods, is to aggregate estimates stemming from diversified statistical methods. In [54], an ensemble method of regression learners was utilized to predict the incidence of COVID-19 in different regions. The idea of ensemble learning is to build a prediction model by combining the strengths of a collection of simpler base models called weak learners. At every step, the ensemble fits a new learner to the difference between the observed response and the aggregated prediction of all learners grown previously. One of the most commonly used loss functions is the least-squares (LS) error [55]. In this study, the model employed a set of individual Least-squares boosting (LSBoost) learners trying to minimize the mean squared error (MSE). The output of the model in step m, $F_m(x)$, was calculated using

$$F_m(x) = F_{m-1}(x) + \rho_m h(x, a_m),$$

where x is the input variable and $h(x; a)$ is the parameterized function of x, characterized by parameters a. The values of ρ and a were obtained from

$$(\rho_m, a_m) = \arg\min_{a,\rho} \sum_{i=1} N[\tilde{y}_i - \rho h(x_i, a)]^2,$$

where N is the number of training data and y_i is the difference between the observed response and the aggregated prediction up to the previous step.

The European Covid-19 forecast Hub and its ensemble method

The European Covid-19 Forecast Hub [9] also proposes "an ensemble, or model average, of submitted forecasts to the European COVID-19 Forecast Hub", described in [12]. In it, the teams submit weekly forecasts for COVID-19 cases and deaths in up to 32 countries for the next week and the three following weeks. The teams also submit standardized quantiles of their predictive distribution. In the *ensemble forecast*, each predictive quantile is calculated as the equally-weighted median of all individual models' predictive quantiles. The performance of each model is evaluated with the relative Weighted Interval Score (WIS), comparing a models' forecast accuracy relative to all other models (see section 2.6 for the formula of WIS).

In [12], the authors report that

> the ensemble performed better on relative WIS than 84% of participating models' forecasts of incident cases (with a total N=862), and 92% of participating models' forecasts of deaths

> (N=746). Across a one to four week time horizon, ensemble performance declined with longer forecast periods when forecasting cases, but remained stable over four weeks for incident death forecasts. In every forecast across 32 countries, the ensemble outperformed most contributing models when forecasting either cases or deaths, frequently outperforming all of its individual component models. Among several choices of ensemble methods we found that the most influential and best choice was to use a median average of models instead of using the mean, regardless of methods of weighting component forecast models.

In view of this, we shall pay a special attention to the comparison of the model proposed here with the ensemble model.

3.7. Global learning

The idea of Global learning is to predict jointly an ensemble of time series with similar characteristics. This method is described in [56] and summarized in the following terms.

> The information of multiple time series can be shared in a single model via a large dimensional manifold embedding. In addition to Europe death series, the regions with the largest average daily deaths are added to reduce the variance of the model estimation and share information (the regions more advanced in the pandemic can help forecast the others). Each time series is time-delay embedded and stacked together before for fitting a single linear autoregressive model. The dimension of the embedding is tuned by temporal validation, the best dimension of the last 4 weeks.

The same method is used in [57], which proposes to estimate a time lag between two countries after finding an optimal dynamic time warping between their incidence curves. This procedure allows an elastic adjustment of the time axis to find similar but phase-shifted sequences. Then the incidence curve of the leading country is used to extend toward future the incidence curve of the other.

3.7.1. Discussion

This group of methods can seen as a direct antecedent of the method proposed here. Indeed, our method (implicitly) estimates time lags between past incidence curves of different countries and the one that we want to extend before exploiting the "future" samples of these time shifted incidence curves to predict the future of our target incidence.

4. Conclusion

Given the large number of factors that can influence a future evolution, forecasting the evolution of the incidence curve is clearly difficult. We saw in section 3 that most standard approaches estimate the parameters of an evolution model (ARIMA, SIR, a logistic curve). In this work, we proposed *EpiLearn*, a method following a more empirical approach that estimates the forecast by a learning procedure using many samples of past incidences evolution in many countries. Using *EpiInvert*, an incidence decomposition method, we removed first the strong administrative weekly bias from the original raw incidence to estimate a smooth incidence trend curve. Using a large database of incidence trends, the forecast is computed as the median of the closest curves, in the past, to the current incidence trend curve. We observed that the size of the estimated empiric confidence interval grows quickly with the number of forecast days. For a 28 days forecast the size of the confidence interval becomes very large, and this is confirmed weekly by our results in the European hub [9]. These results place *EpiInvert* among the very best methods. We observed that the prediction of all methods may miss the forecast target by a large margin in the three and four weeks horizon. Nevertheless, they seem to be reliable and useful to predict the pandemic in a two week horizon.

5. Material and Methods

EpiLearn, the forecasting model presented in this work, introduced in section 2, is implemented in the publicly available *EpiInvert* CRAN R package [5]. In this package, *EpiLearn* is executed using the *EpiInvertForecast* R function. In the vignette https://ctim.ulpgc.es/covid19/EpiInvertForecast.html a description, with examples, of *EpiInvertForecast* usage is presented.

The incidence trend database has been built using the daily incidence data, up to May 5, 2022, provided in https://covid.ourworldindata.org/data/owid-covid-data.csv for the following countries and regions: Argentina, Austria, Bangladesh, Belgium, Brazil, Canada, Chile, Colombia, Cuba, Czech Republic, Denmark, Germany, France, Greece, Hungary, India, Iraq, Iran, Ireland, Israel, Italy, Japan, Jordan, Kazakhstan, Malaysia , Mexico, Nepal, Netherlands, Peru, Philippines, Poland, Romania, Russia, Serbia, Slovakia, South Africa, South Korea, Spain, Sweden, Switzerland, Thailand, Tunisia, Turkey, Ukraine, United Arab Emirates, United Kingdom, USA , Vietnam, Africa, South America, North America, Asia, Europe, European Union, Oceania, and the world.

The comparative results with other methods, presented in section 2.6 have been obtained by using the weekly evaluation reports published by the European Hub, [1] in the repository https://github.com/covid19-forecast-hub-europe/covid19-forecast-hub-europe/tree/main/evaluation/weekly-summary. The eight forecasting models that are being compared are the only ones that provided reasonable forecasts over the whole time period considered. In this sense, we excluded the ILM-EKF model from the comparison because the scores published by the Hub for this method were unreasonable (in some cases its score is greater than 10).

Author Contributions: Conceptualization, J-D.M., J-M.M. and L.A.; methodology, J-D.M., J-M.M. and L.A.; software, L.A.; validation, J-D.M., J-M.M. and L.A.; formal analysis, J-D.M., J-M.M. and L.A.; investigation, J-D.M., J-M.M. and L.A.; resources, J-D.M., J-M.M. and L.A.; data curation, J-D.M., J-M.M. and L.A.; writing–original draft preparation, J-D.M., J-M.M. and L.A.; writing–review and editing,J-D.M., J-M.M. and L.A.; visualization, J-D.M.; supervision, L.A.; project administration, J-D.M., J-M.M. and L.A.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets of the incidence curve were analyzed in this study. These data can be found in [58]. The method evaluation performance uses the weekly evaluation reports published in the European CoviD-19 Hub [1].

Conflicts of Interest: The authors declare no conflict of interest.

1. Sherratt, K.; others. European COVID-19 Forecast Hub. *Zenodo* **2022**. doi:10.5281/zenodo.7006980.
2. Alvarez, L.; Morel, J.D.; Morel, J.M. Modeling COVID-19 incidence by the renewal equation after removal of administrative bias and noise. *Biology* **2022**, *11(4)*, 1–22.
3. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* **2013**, *178*, 1505–1512.
4. Alvarez, L.; Colom, M.; Morel, J.D.; Morel, J.M. 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* **2021**, *118(50)*, 1–10.
5. Alvarez, L.; Morel, J.D.; Morel, J.M. EpiInvert R package. *CRAN* **2022**. https://CRAN.R-project.org/package=EpiInvert.
6. Bosse, N.; Abbott, S.; Funk, S. EpiNow2 (epiforecasts). *London School of Hygiene and Tropical Medicine* **2022**.
7. Hunter, J.S. The exponentially weighted moving average. *Journal of quality technology* **1986**, *18*, 203–210.
8. Srivastava, A. The Variations of SIkJalpha Model for COVID-19 Forecasting and Scenario Projections. *arXiv preprint arXiv:2207.02919* **2022**.
9. European Covid-19 Forecast Hub. https://covid19forecasthub.eu/. Accessed: 2022-11-02.

10. The COVID-19 Forecast Hub. https://covid19forecasthub.org. Accessed: 2022-11-02.
11. Bosse, N.I.; Gruson, H.; Cori, A.; van Leeuwen, E.; Funk, S.; Abbott, S. Evaluating Forecasts with scoringutils in R. *arXiv preprint arXiv:2205.07090* **2022**.
12. Sherratt, K.; others. Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. *medRxiv* **2022**, [https://www.medrxiv.org/content/early/2022/06/16/2022.06.16.22276024.full.pdf]. doi:10.1101/2022.06.16.22276024.
13. Baseline model automatically generated using past truth data. https://github.com/reichlab/covidModels/blob/master/R-package/R/quantile_baseline.R. Accessed: 2022-09-25.
14. Hyndman, R.; Athanasopoulos, G.; Bergmeir, C.; Caceres, G.; Chhay, L.; O'Hara-Wild, M.; Petropoulos, F.; Razbash, S. Forecasting functions for time series and linear models. *R package version* **2015**, *6*.
15. epiforecasts-weeklygrowth. https://github.com/seabbs/ecdc-weekly-growth-forecasts. Accessed: 2022-09-25.
16. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: principles and practice*; OTexts, 2018.
17. MUNI-ARIMA. https://krausstat.shinyapps.io/covid19global/. Accessed: 2022-09-25.
18. Tandon, H.; Ranjan, P.; Chakraborty, T.; Suhag, V. Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. *arXiv preprint arXiv:2004.07859* **2020**.
19. Kufel, T. ARIMA-based forecasting of the dynamics of confirmed COVID-19 cases for selected European countries. *Equilibrium. Quarterly Journal of Economics and Economic Policy* **2020**, *15*, 181–204.
20. Katoch, R.; Sidhu, A. An application of ARIMA model to forecast the dynamics of COVID-19 epidemic in India. *Global Business Review* **2021**, p. 0972150920988653.
21. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of The Total Environment* **2020**, *729*, 138817.
22. Demongeot, J.; Oshinubi, K.; Rachdi, M.; Hobbad, L.; Alahiane, M.; Iggui, S.; Gaudart, J.; Ouassou, I. The application of ARIMA model to analyze COVID-19 incidence pattern in several countries. *J. Math. Comput. Sci.* **2021**, *12*, Article–ID.
23. Malki, Z.; Atlam, E.S.; Ewis, A.; Dagnew, G.; Alzighaibi, A.R.; ELmarhomy, G.; Elhosseini, M.A.; Hassanien, A.E.; Gad, I. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Computing and Applications* **2021**, *33*, 2929–2948.
24. Fatimah, B.; Aggarwal, P.; Singh, P.; Gupta, A. A comparative study for predictive monitoring of COVID-19 pandemic. *Applied Soft Computing* **2022**, *122*, 108806.
25. Benvenuto, D.; Giovanetti, M.; Vassallo, L.; Angeletti, S.; Ciccozzi, M. Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief* **2020**, *29*, 105340.
26. Khan, F.M.; Gupta, R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience* **2020**, *1*, 12–18.
27. Alzahrani, S.I.; Aljamaan, I.A.; Al-Fakih, E.A. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of infection and public health* **2020**, *13*, 914–919.
28. Perone, G. An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. *MedRxiv* **2020**.
29. Sharma, R.R.; Kumar, M.; Maheshwari, S.; Ray, K.P. EVDHM-ARIMA-based time series forecasting model and its application for COVID-19 cases. *IEEE Transactions on Instrumentation and Measurement* **2020**, *70*, 1–10.
30. Roy, S.; Bhunia, G.S.; Shit, P.K. Spatial prediction of COVID-19 epidemic using ARIMA techniques in India. *Modeling earth systems and environment* **2021**, *7*, 1385–1391.
31. Duan, X.; Zhang, X. ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. *Data in brief* **2020**, *31*, 105779.
32. Panda, M. Application of ARIMA and Holt-Winters forecasting model to predict the spreading of COVID-19 for India and its states. *medRxiv* **2020**.
33. Dehning, J.; Zierenberg, J.; Spitzner, F.P.; Wibral, M.; Neto, J.P.; Wilczek, M.; Priesemann, V. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **2020**, *369*, eabb9789.
34. Dehning, J.; others. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **2020**, *369*, eabb9789.

35. Šmíd, M.; Berec, L.; Kuběna, A.A.; Levínskỳ, R.; Trnka, J.; Tuček, V.; Zajíček, M. SEIR filter: A stochastic model of epidemics. *medRxiv* **2021**.
36. Rodiah, I.; others. age-structured and extended SEIR model. *Helmholtz Zentrum fuer Infektionsforschung, Leibniz Universitaet Hannover, Technische Universitaet Kaiserslautern* **2022**.
37. Srivastava, A.; Xu, T.; Prasanna, V.K. Fast and Accurate Forecasting of COVID-19 Deaths Using the SIkJ \$\alpha\$ Model. *arXiv preprint arXiv:2007.05180* **2020**.
38. de León , U.A.P.; .; Pérez, Á.G.C.; Avila-Vales, E. A data driven analysis and forecast of an SEIARD epidemic model for COVID-19 in Mexico. *Big Data and Information Analytics* **2020**, *5*, 14–28. doi:10.3934/bdia.2020002.
39. Parolini, N.; Dede', L.; Antonietti, P.F.; Ardenghi, G.; Manzoni, A.; Miglio, E.; Pugliese, A.; Verani, M.; Quarteroni, A. SUIHTER: A new mathematical model for COVID-19. Application to the analysis of the second epidemic outbreak in Italy. *Proceedings of the Royal Society A* **2021**, *477*, 20210027.
40. Ardenghi, G.; others. epiMOX-SUIHTER. *MOX, Dipartimento di Matematica, Politecnico di Milano* **2022**.
41. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.M.; Yan, P.; Chowell, G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling* **2020**, *5*, 256–263.
42. Chowell, G.; Tariq, A.; Hyman, J.M. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. *BMC medicine* **2019**, *17*, 1–18.
43. Richards, F. A flexible growth function for empirical use. *Journal of experimental Botany* **1959**, *10*, 290–301.
44. Alaimo Di Loro, P.; Divino, F.; Farcomeni, A.; Jona Lasinio, G.; Lovison, G.; Maruotti, A.; Mingione, M. Nowcasting COVID-19 incidence indicators during the Italian first outbreak. *Statistics in Medicine* **2021**, *40*, 3843–3864.
45. Mingione, M.; Di Loro, P.A.; Farcomeni, A.; Divino, F.; Lovison, G.; Maruotti, A.; Lasinio, G.J. Spatio-temporal modelling of COVID-19 incident cases using Richards' curve: An application to the Italian regions. *Spatial Statistics* **2022**, *49*, 100544.
46. Mingione, M.; AND, P.A.D.L. Statgroup19-richards. *University of Rome "La Sapienza"* **2022**.
47. Català, M.; Alonso, S.; Alvarez-Lacalle, E.; López, D.; Cardona, P.J.; Prats, C. Empirical model for short-time prediction of COVID-19 spreading. *PLOS Computational Biology* **2020**, *16*, 1–18.
48. Beckmann, D. DirkBeckmann-Gompertz **2022**.
49. RobertWalraven-ESG. https://github.com/seabbs/ecdc-weekly-growth-forecasts. Accessed: 2022-09-25.
50. Zhao, H.; Merchant, N.N.; McNulty, A.; Radcliff, T.A.; Cote, M.J.; Fischer, R.S.; Sang, H.; Ory, M.G. COVID-19: Short term prediction model using daily incidence data. *PloS one* **2021**, *16*, e0250110.
51. Bosse, N.I.; Abbott, S.; Bracher, J.; Hain, H.; Quilty, B.J.; Jit, M.; van Leeuwen, E.; Cori, A.; Funk, S.; others. Comparing human and model-based forecasts of COVID-19 in Germany and Poland. *medRxiv* **2021**.
52. ILM-EKF. https://github.com/Stochastik-TU-Ilmenau. Accessed: 2022-09-25.
53. Krymova, E.; Béjar, B.; Thanou, D.; Sun, T.; Manetti, E.; Lee, G.; Namigai, K.; Choirat, C.; Flahault, A.; Obozinski, G. Trend estimation and short-term forecasting of COVID-19 cases and deaths worldwide. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2112656119.
54. Ahouz, F.; Golabpour, A. Predicting the incidence of COVID-19 using data mining. *BMC public health* **2021**, *21*, 1–12.
55. Friedman, J.H. Greedy function approximation: a gradient boosting machine. *Annals of statistics* **2001**, pp. 1189–1232.
56. Montero-Manso, P.; Hyndman, R.J. Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting* **2021**, *37*, 1632–1653.
57. Stübinger, J.; Schneider, L. Epidemiology of coronavirus COVID-19: Forecasting the future incidence in different countries. Healthcare. MDPI, 2020, Vol. 8, p. 99.
58. H. Ritchie and others. Coronavirus Pandemic (COVID-19), OurWorldInData.org, . Accessed May 6, 2022.