Supplementary Information for

Trend estimation and short-term forecasting of COVID-19 cases and deaths worldwide

E. Krymova, B. Béjar, D. Thanou, T. Sun, E. Manetti, G. Lee, K. Namigai, C. Choirat, A. Flahault, G. Obozinski

Corresponding Author
E-mail: ekaterina.krymova@epfl.ch

This PDF file includes:

Supplementary text
Figs. S1 to S14 (not allowed for Brief Reports)
Tables S1 to S4 (not allowed for Brief Reports)
SI References

Supporting Information Text

A. Evaluation

Global comparison with the baseline. We evaluate our method by reporting, for each country the RMAE, the RmedianAE, the relative improvement in mean total coverage (see Section E), and average WIS. The relative improvement in mean total coverage is computed as $\mathrm{RC} = (\mathrm{MC}(s) - \mathrm{MC}(b))/\mathrm{MC}(b)$ from the coverages MC of the methods, since one aims for higher coverage by the estimates of the confidence intervals.
In Fig. S5, we illustrate the evaluation for the 80 countries with reliable data. In Fig. S6, we illustrate the evaluation score for the 101 countries that did not pass our evaluation criteria. The evaluation on the remaining countries (with the exclusion of 11 countries which had a particularly low number of cases) shows that there remains 42 countries, for which our method still obtains a better MAE than the baseline, only 29 for which we improve in median AE, 63 for which RWIS is improved and 74 with improved total coverage. Many of the countries for which our method performs clearly worse than the baseline are in fact countries with a fairly low number of cases, which were not the focus of our modelling efforts. This is justified by the fact that accurate forecasts are not critical for these countries as long as their number remains low. Furthermore, the corresponding time series of a significant number of these countries have numerous irregularities of the reporting and backlogs which makes it harder to associate the target average weekly number with the true underlying trend. In that case, the simple baseline forecast, which is the average value of the previous week, appears to be closer to the target than the forecast obtained after trend estimation. Different smoothing techniques would be needed to produce better trend estimation for these countries, which take into account the discrete nature of the count and their Poissonian distribution.

Comparison with the forecasts submitted to European Covid Forecast Hub, cases. The list of methods includes (with abbreviations used in the Tables S1-S4 below):

- EuroCOVIDhub-ensemble (EUHub-ens, https://covid19forecasthub.eu/visualisation.html),
- EuroCOVIDhub-baseline (https://covid19forecasthub.eu/visualisation.html),
- MUNI-ARIMA (MUNI, https://krausstat.shinyapps.io/covid19global/),
- IEM_Health-CovidProject (IEM_Health, https://iem-modeling.com/),
- USC-SIkJalpha (USC, https://scc-usc.github.io/ReCOVER-COVID-19/#/),
- RobertWalraven-ESG (RW, http://rwalraven.com/COVID19/Model),
- ILM-EKF (ILM, https://github.com/Stochastik-TU-Ilmenau),
- the proposed method (SDSC_ISG, https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/).

Fig. S1-S3 show histograms of the errors of the methods with respect to the baseline with 0.1-wide bins. For visualization purposes, errors greater than 1.55 were set to 1.55 and therefore contribute to the last bin. One week ahead forecasts MAE measured in multiples of the baseline MAE and average WIS measured in multiples of the baseline WIS are presented in Tables S1 and S2. Two week ahead forecasts MAE measured in multiples of the baseline MAE and average WIS measured in multiples of the baseline WIS are presented in Tables S3 and S4.

Death forecasting: motivation and comparison with European Covid Forecast Hub submissions. At first sight, in a very simple theoretical model, the number of deaths should be related to the number of cases, and correspond simply to the fraction of the cases that did not survive. The strategy of estimating the deaths from cases has been particularly successful for the USA at the country and state levels. Among the models participating in the US COVID Forecast Hub, `epiforecasts-ensemble1` contains a model which estimates deaths from a convolution of cases, the model `MIT_Crit_Data-GBCF` takes 3-weeks lagged deaths and cases numbers as a part of the input, etc. Forecasting the number of deaths from a lagged case curve was one of our first approaches, but the diversity of situations encountered across the world and in time for a particular region makes it that this strategy fails in a number of cases. The relation between lagged cases and the number of deaths is sometimes quite unclear: for example, if we consider the evolution of the number of cases and deaths in Egypt in November-December 2021 that we show in Fig. S7, the number of cases is almost not changing while the number of death is increasing and then decreasing. There are many reasons why the relation between the number of cases might be more complicated, be non-stationary and potentially change relatively quickly: as the virus circulates it affects different groups in the population who are more or less fragile and who protect their senior more or less well, the testing and reporting policies of some countries have sometimes changed quite quickly (including reporting policies for deaths), there is an effect of the vaccination (which is however on a sufficiently slow timescale that it can be reestimated over time), there is the emergence of new variants, etc. Taking into account all of the above, we use the same strategy for the number of deaths forecasting as for cases, i.e. we estimate the trend based solely on the previous deaths observations and predict future numbers by the simple extrapolation.

We provide a brief comparison of the performance of our strategy to forecast deaths (individually in the same way as cases) with a few methods from the European forecast Hub for 31 European countries in a similar way to what we did for cases. We identified two forecasting methods IEM_Health-CovidProject and RobertWalraven-ES, which were regularly submitting to the

European COVID forecast Hub apart from EuroCOVIDhub-ensemble and EuroCOVIDhub-baseline. The results demonstrate that for 1-week ahead forecast our methodology obtains levels of performance comparable to those of other methods submitted to EU COVID Forecast Hub: on Fig. S8, S10 and Fig. S9, S11 one can see that mean absolute error (MAE) and WIS normalized by the respective errors of the EU COVID hub baseline of our method (SDSC_ICG) are aligned with the other methods.

B. Raw data preprocessing

In this section, we describe the preliminary data cleaning steps and smoothing for the daily cases and deaths.

Negative values. When a negative count is reported following a reassessment of previously reported cases or deaths, we substitute the negative value with the estimate x_t computed from the daily observations a week before multiplied by the growth factor computed from two weekly observations: during a week before and two weeks before the negative value had occurred, i.e. with

$$x_t \frac{X_{t-1}}{X_{t-8}}$$

where x_t and $X_t = \sum_{k=0}^{6} x_{t-k}$ are daily and weekly observations respectively. Next we reduce the counts in the whole previous history by a constant multiplicative factor c to match the cumulative counts obtained by removing this negative quantity from the cumulative counts. That is we compute $c = (\sum_{s=0}^{t} x_t)/(\sum_{s=0}^{t-1} x_t)$ and x_t is updated to $x_t \leftarrow c\, x_t$. For the case of large and significantly delayed reports or reassessment leading symmetrically to a large positive spike in the data, this is ignored at this stage, and will be addressed by the trend-estimation method.

Identifying last missing daily reports. One of the difficulties with the data sources that we are using is that no distinction is made between a missing report and an existing report stating that no new cases should be reported on a given day: in both cases the database contains a zero. This is of course because, in practice, there is often no distinction made in the reporting protocols. It is however important for our models to be able to distinguish between these two situations. To distinguish missing values from actual zeros, we proceed as follows: if the last observation in the data is zero, we compute an estimate of the Poisson rate by taking the average of observations during the week before zero occurs. If the probability of observing zero new cases given this estimated rate is too low, we consider the zero value to be a missing report one and exclude it from the history for further trend estimation and forecasting. The forecasting starts from the next day from the last initial observation.

Imputations. Many countries have seasonal patterns in which no data is reported on certain days, typically during the weekend, and where all the new cases that appeared during these days are reported all together on the next reporting day (typically a Monday). We use as a preprocessing a simpler imputation scheme, which consists in reassigning the data declared on that last day uniformly over the previous days of missing report and the following reporting day.

C. Details of the piecewise STL algorithm

Since the seasonal pattern might evolve over time due to changes in the reporting pattern we propose to apply STL in a piecewise fashion. This allows our method to better adapt to changes in the seasonal component, as the hyperparameters defining the smoothing levels can then change in each separate segment modelled. To estimate the trend locally in the whole period of observations, we split the observed time interval into half-overlapping intervals of 6 weeks. These time intervals are defined from the end of the time series backwards, so that the time series ends with the last segment of 6 months.
First, we apply STL to estimate the trend of the last subinterval $[T-L+1, T]$, where $L = 42$ (corresponding to 4 weeks). For the last subinterval, the STL trend is rescaled to preserve the number of observations in the last $L/2$ days to obtain the estimate s_{-1} in $[T-L+1, T]$. Next, the trend estimation proceeds as follows. For the two overlapping subintervals, e.g. consider $[T-L+1, T]$ and $[T-3L/2+1, T-L/2]$, we take the estimate s_{-1} and we estimate the trend $\tilde{s}$ in $[T-3L/2+1, T-L/2]$. In order to smoothly join two local trends s_{-1} and $\tilde{s}_{-2}$ in the interval $[T-L+1, T-L/2]$ we use a simple weighted interpolation and obtain the trend in $[T-3L/2+1, T]$:

$$s_{-2}(t_0+\tau) = \begin{cases} \tilde{s}(t_0+\tau), & \tau = -L/2+1, \dots, 0, \\ \sigma(\tau)\, \tilde{s}(t_0+\tau) + (1-\sigma(\tau))\, s_{-1}(t_0+\tau), & \tau = 1, \dots, L/2, \\ s_{-1}(t_0+\tau), & \tau = L/2+1, \dots, L, \end{cases}$$

where $t_0 = T-L+1$, and $\sigma(\tau) = (1+\exp(a(\tau-1)-b))^{-1}$ with $a = 21.1/L$, $b = 5.46$. We additionally apply rescaling to redistribute the possible outliers, removed by trend estimation: we compare the sum of the numbers so far estimated by the trend, e.g. $S_{-2} = \sum_{i=0}^{3L/2-1} \tilde{s}_{-2}(T-i)$, with the corresponding number of raw daily observations $\kappa_{-2} = \sum_{i=0}^{3L/2-1} x_{T-i}$: if the excess $\kappa_{-2} - S_{-2}$ is positive, it is added to the observations before $T - 3L/2 + 1$ by rescaling, otherwise the trend is rescaled such that the sum of estimated numbers meets κ_{-2}. The procedure continues with the next local trend estimate ($\tilde{s}$) from the corrected data. Note that κ is always computed from raw observations. Local trend estimation with rescaling repeats backward until we reach the beginning of the time interval. As a result, we get a smooth trend, the sum of which is equal to the sum of raw daily observations.

D. Outlier detection scheme

One of the criteria for the inclusion of a country into our main evaluation set is that there are not too many large delayed reports. We assimilate these as outliers and use a simple estimate for each time series of the number of outliers. The corresponding outlier detection scheme is based on the Median Absolute Deviation (MAD), which is defined as $\mathrm{MAD} = \mathrm{median}(|x_i - \mathrm{median}(x_i)|)$ for daily observations x_i. For each country, we detect outliers using a sliding window MAD estimate. More precisely, for each daily value, we compute the MAD in a symmetric window of 22 days around that day. For the MAD to be a consistent estimator for the standard deviation, we multiply it by a constant scale factor of 1.4826, which relates the MAD to the standard deviation for a Gaussian distribution. If the daily value differs from the median in the window by more than 2 standard deviations, we consider it as an outlier. Note that this procedure is only used for the construction of the set of countries in the evaluation set, and not for the trend estimation or forecast.

E. Total Coverage

We define the *total coverage* of a probabilistic forecast P the sum over all levels considered of the coverage of the intervals $[q_k, q_{2K+2-k}]$ as defined in (1):

$$C(P, A, \xi) = \sum_{k=1}^{K} \mathbb{1}\{q_k \leq \xi \leq q_{2K+2-k}\},$$

where $\mathbb{1}\{\cdot\}$ is 1 if the condition is satisfied and zero otherwise. We the define the *mean total coverage* as

$$\mathrm{MC} = \frac{1}{T} \sum_{t=1}^{T} C(P_t, A, X_t),$$

where X_t are the weekly total number of cases/deaths.

F. Growth rate analysis

To be able to estimate the growth rate of the trend, we first compute an independent estimate of the trend using cubic B-splines on the weekly data and next compute the growth rate as the slope of the trend normalized by the trend value. To aggregate AE values for different countries, we use the MAPE (the mean of $\mathrm{AE}(F_t)/X_t$) on the weekly forecasts in the evaluation period instead of the MAE, to bring the errors for each country on a comparable scale. Given that the growth rate as a measure of the slope is comparable across countries, we pool the data from all countries to obtain Fig. S4. The baseline performs best when there are no changes in the number of cases/deaths (i.e., the growth rate of 0 or constant trend). However, our method outperforms the baseline predictor as soon as the growth rate is larger than 3% in absolute value, which shows that the proposed forecast is informative as soon as the trend is not flat.

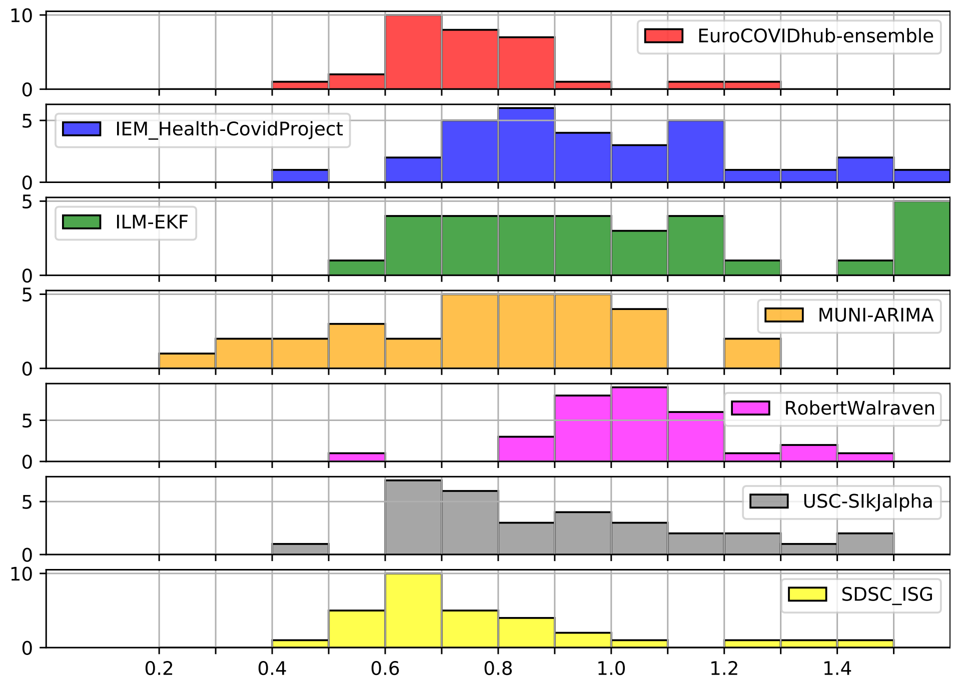

Fig. S1. Histograms for the MAE (in x-axis) based on 1 week ahead cases forecasts for 31 European countries

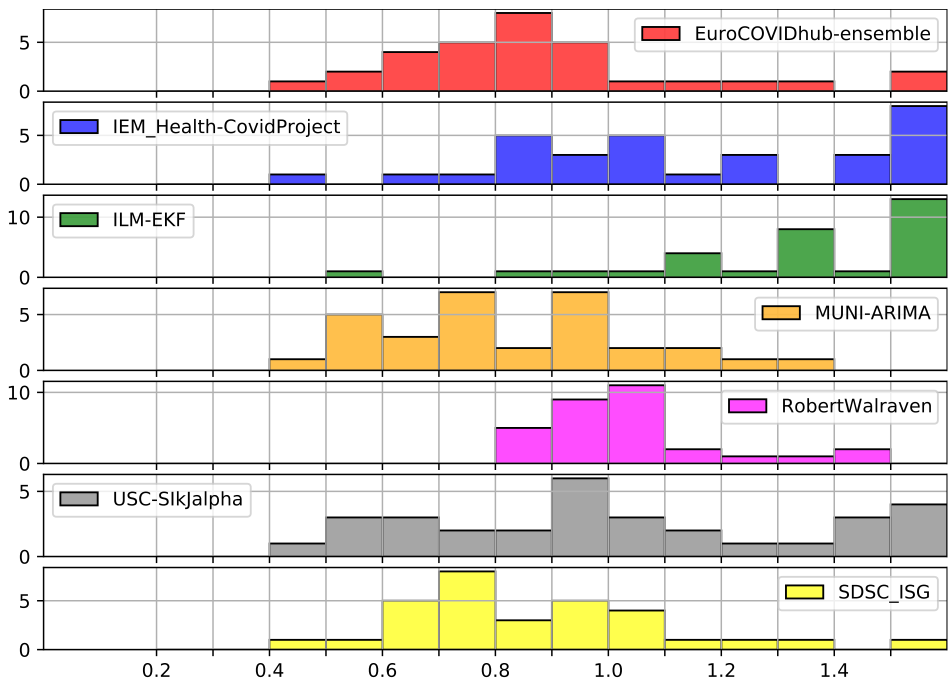

Fig. S2. Histograms for the MAE (in x-axis) based on 2 week ahead cases forecasts for 31 European countries

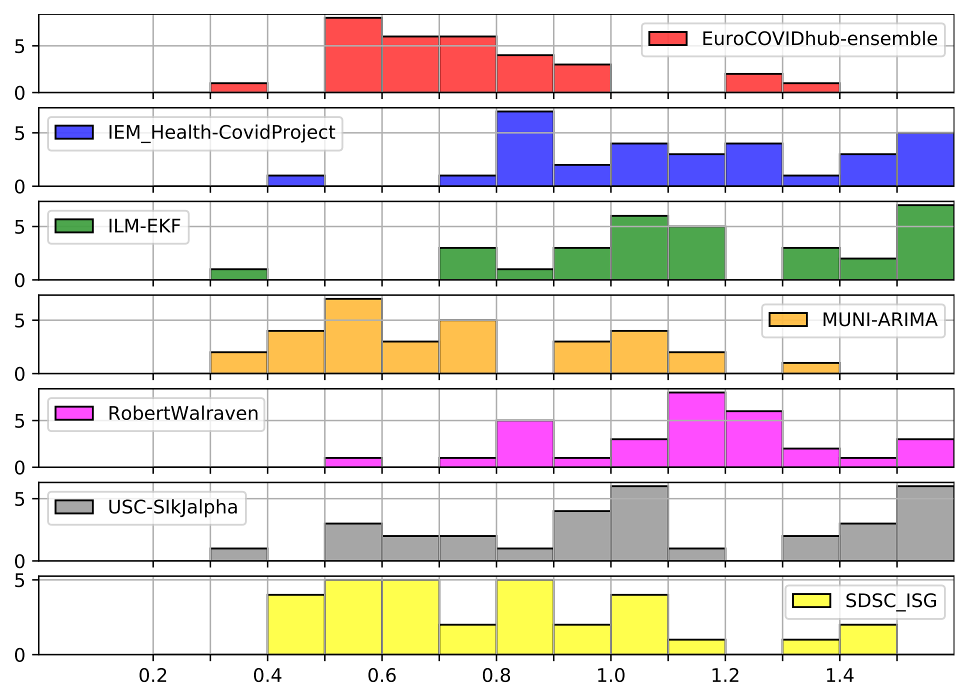

Fig. S3. Histograms for the average WIS (in x-axis) based on 2 week ahead cases forecasts for 31 European countries

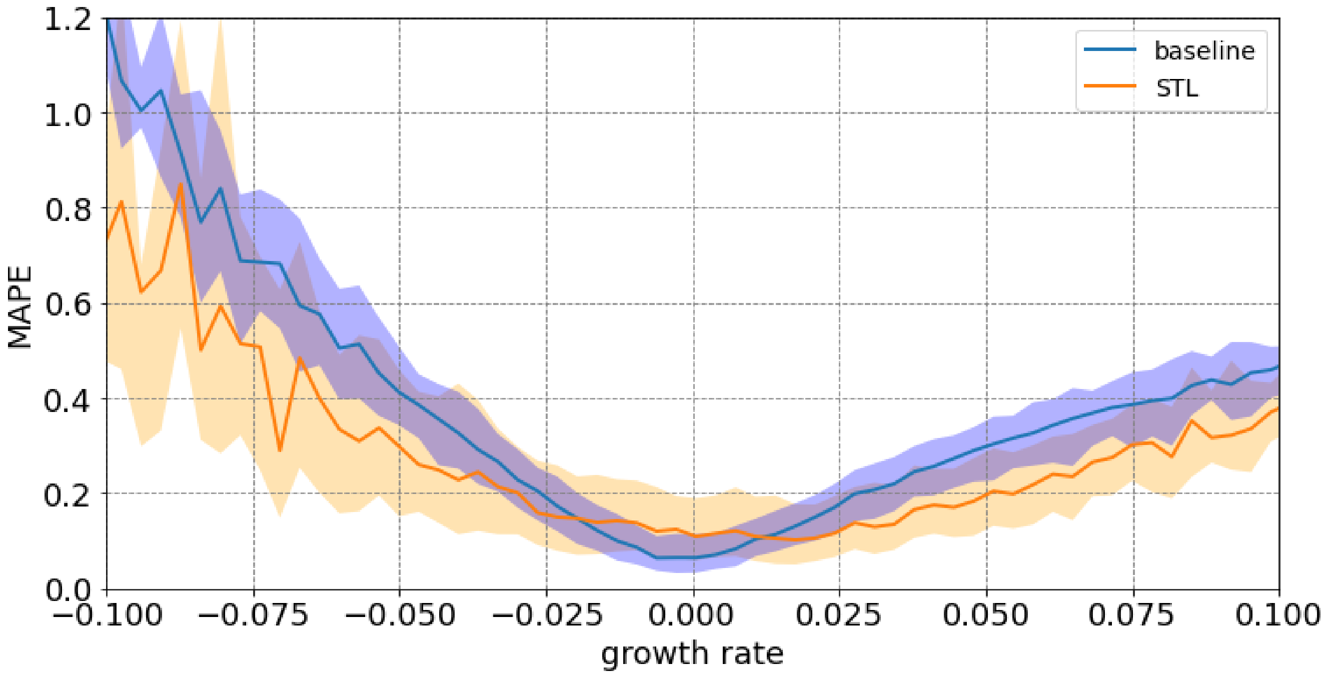

Fig. S4. Dependence of the error on the relative slope of the trend: Median (solid line) and interquartiles (shaded region) over all countries (this excludes 11 countries with more than 90% of zero daily observations: Marshall Islands, Grenada, Vanuatu, Tanzania, Fiji, Saint Kitts and Nevis, Micronesia, Samoa, the Holy See, Solomon Islands, Laos) of the MAPE of each forecasting algorithm (blue: baseline, orange: proposed forecast) as a function of the growth rate (aka relative slope) of the trend. Since the baseline assumes zero slope it has lower median error when the absolute growth rate is less than 3%, but larger median error otherwise.

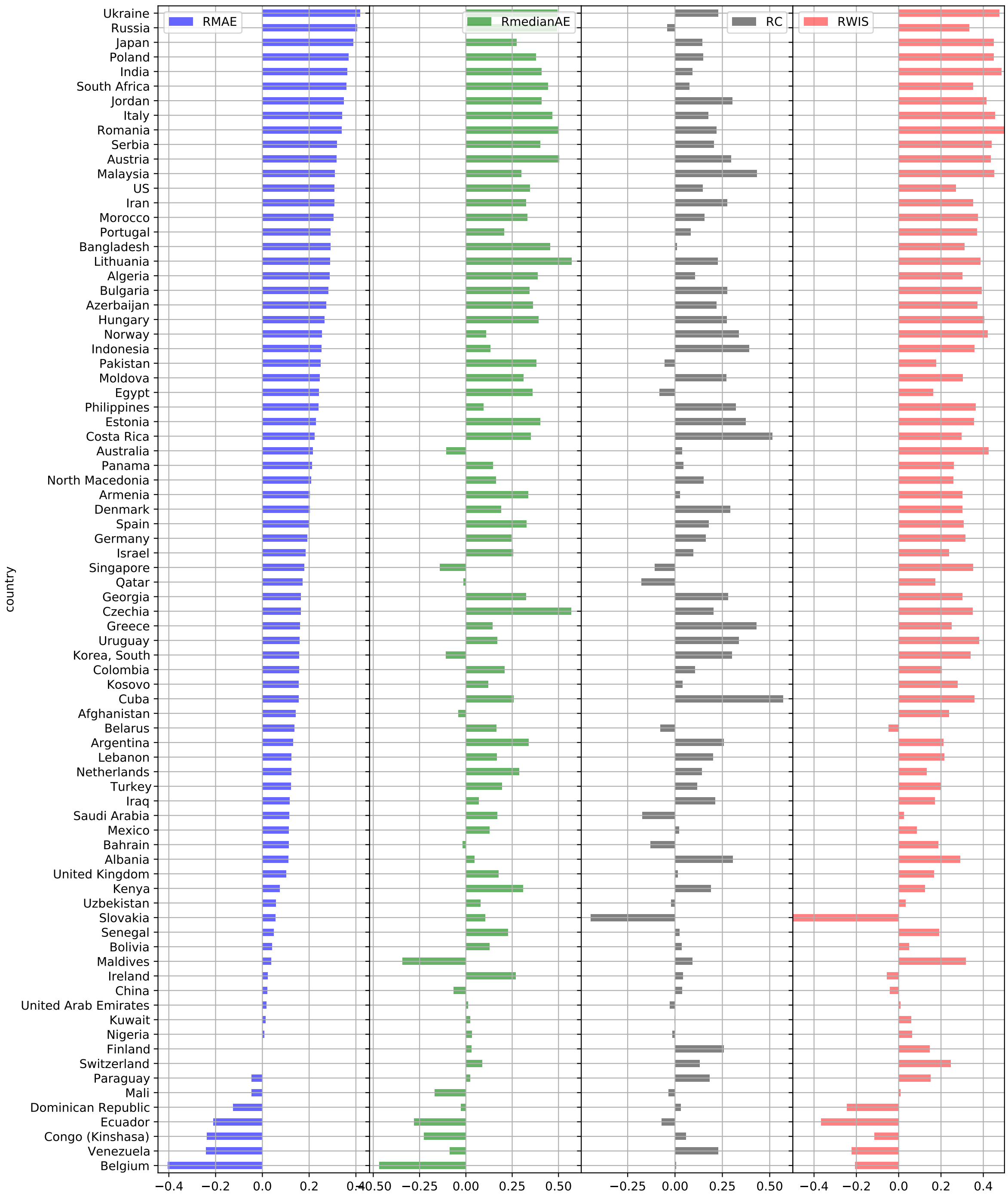

Fig. S5. RMAE: For 90 % of countries of the selected set our method outperforms the baseline in MAE for average weekly cases. (b) RmedianAE: For 82.5% of countries of the selected set our method outperforms the baseline in median AE for average weekly cases. (c) R0: For 80% of the countries of the selected set, our method performs better than the baseline in total coverage (d) RWIS; For 88.75% of countries of the selected set our method outperforms the baseline in average WIS for average weekly cases

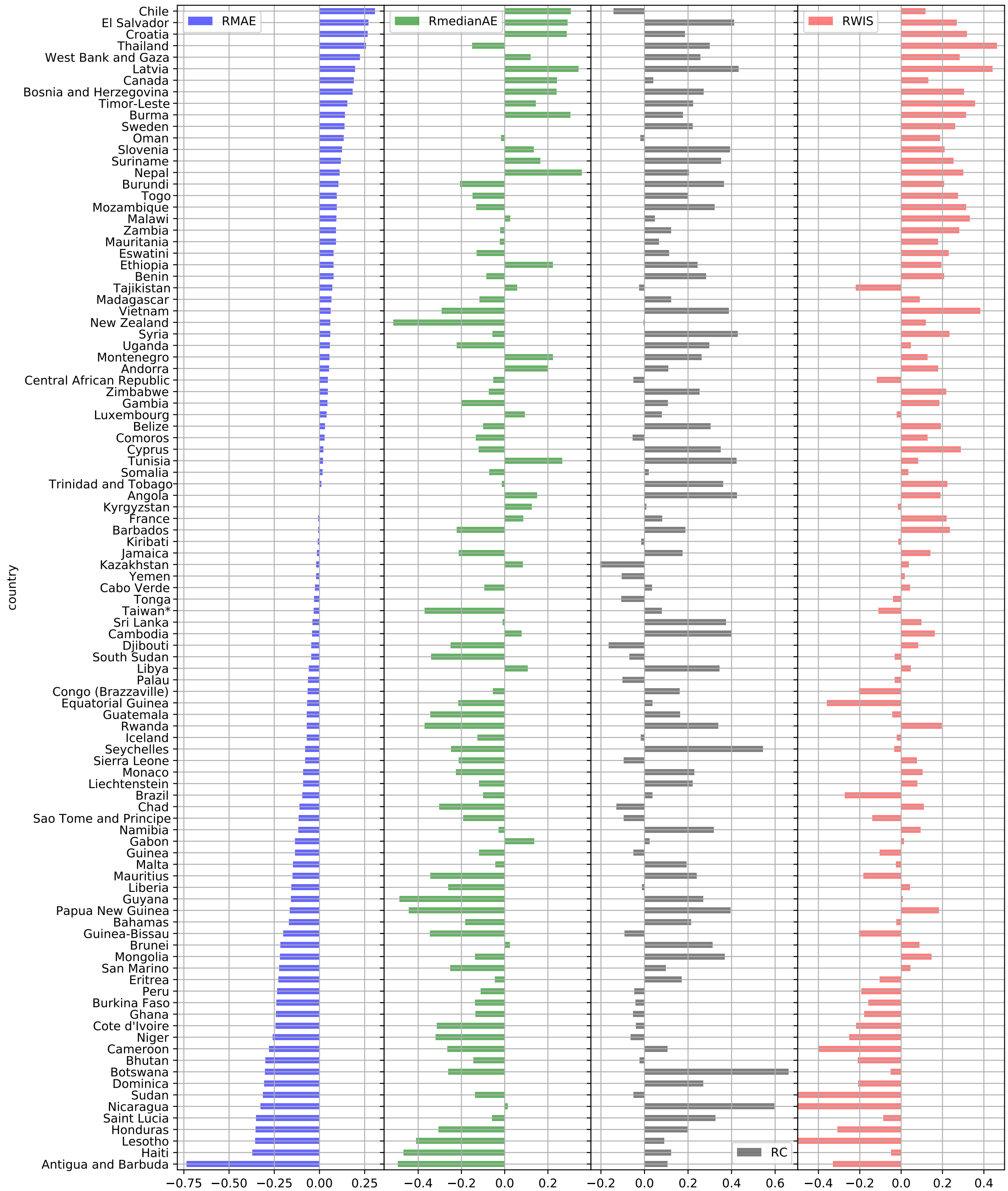

Fig. S6. Evaluation scores for the 101 countries, corresponding to the countries not included in the list of 80 countries considered in Section that have more than 10% of non-zero daily observations (this excludes countries with more than 90% of zero daily observations: Marshall Islands, Grenada, Vanuatu, Tanzania, Fiji, Saint Kitts and Nevis, Micronesia, Samoa, the Holy See, Solomon Islands, Laos, Saint Vincent and the Grenadines) (a) RMAE. (b) RmedianAE. (c) RC (improvement in total coverage). (d) RWIS.

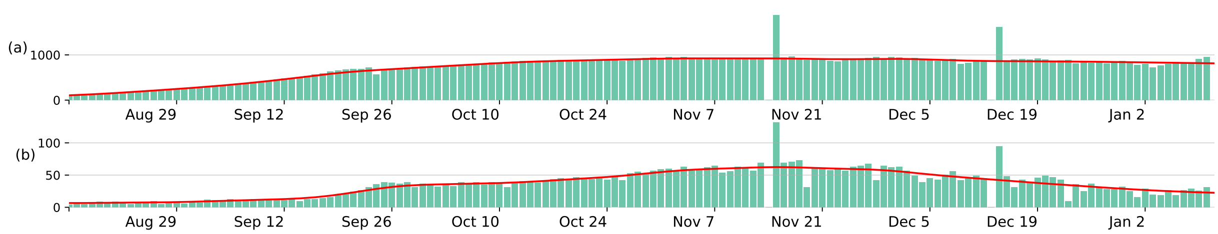

Fig. S7. Cases in (a) and deaths numbers in (b) in Egypt in the end of 2021: the growth and decrease of death numbers is not preceeded by the similar behavior in the cases.

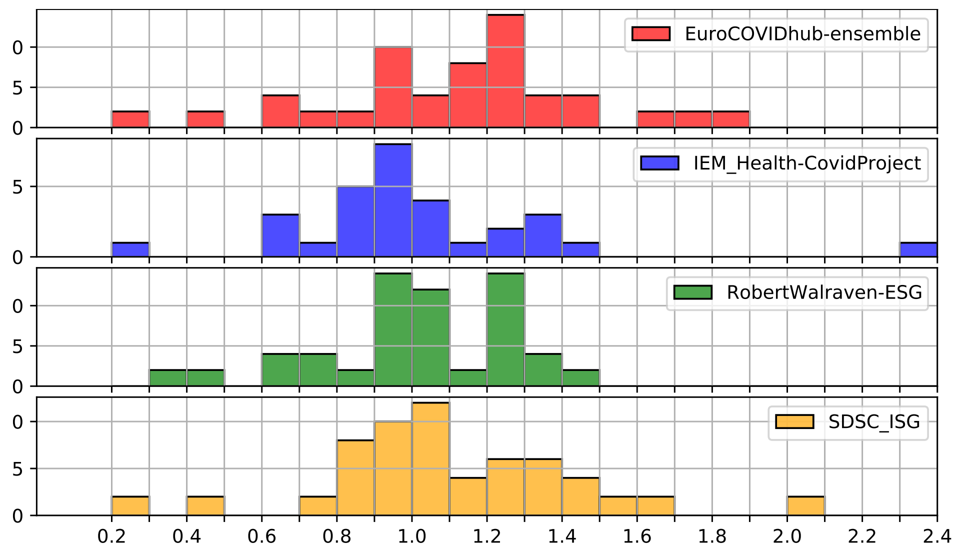

Fig. S8. Histograms for the MAE (in x-axis) based on 1 week ahead deaths forecasts for 31 European countries

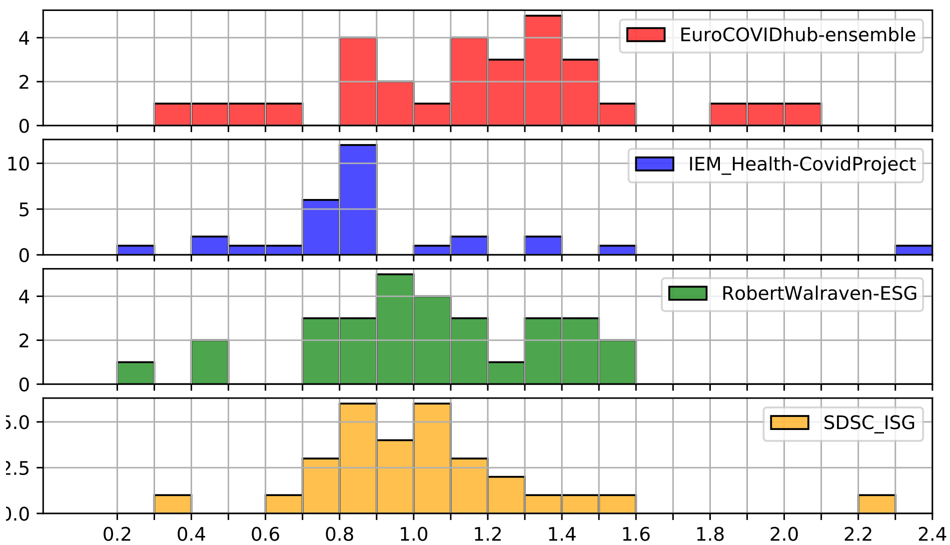

Fig. S9. Histograms for the average WIS (in x-axis) based on 1 week ahead deaths forecasts for 31 European countries

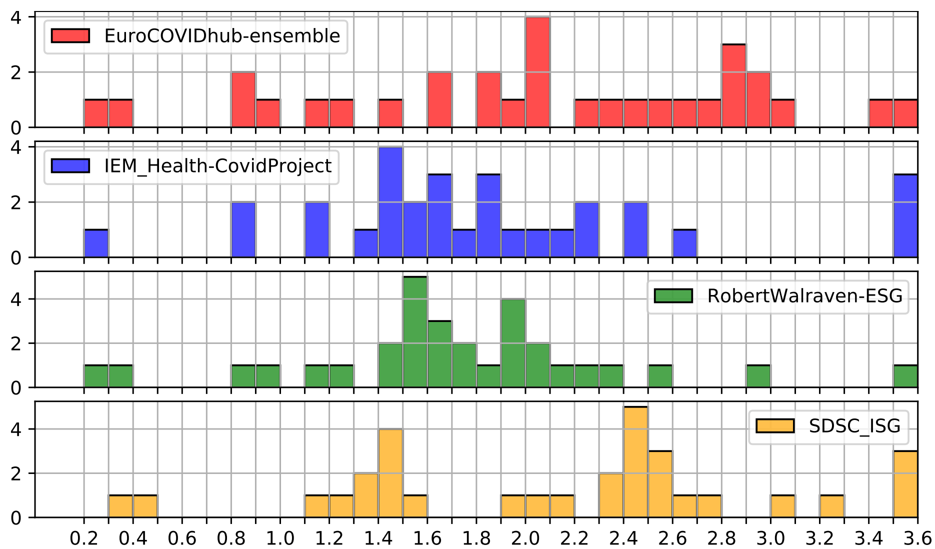

Fig. S10. Histograms for the MAE (in x-axis) based on 2 week ahead deaths forecasts for 31 European countries

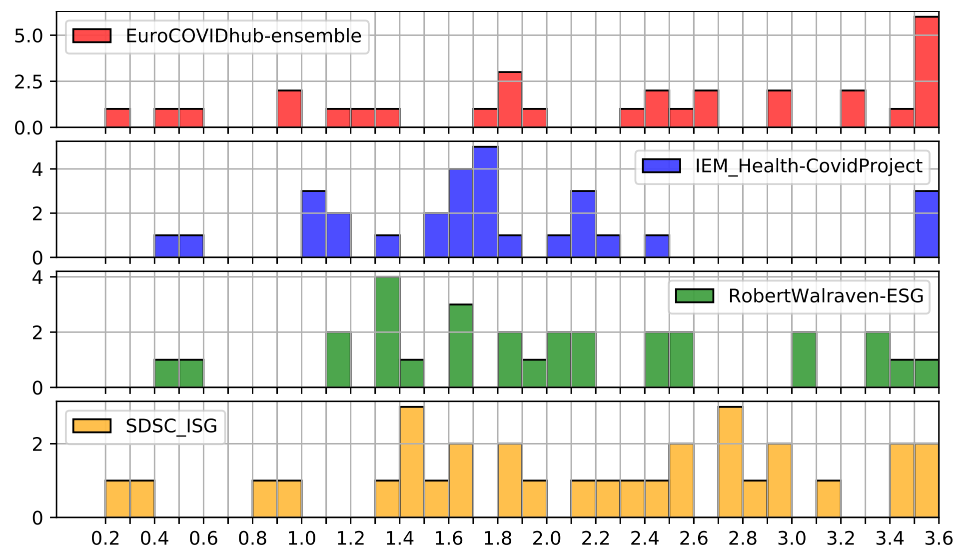

Fig. S11. Histograms for the average WIS (in x-axis) based on 2 week ahead deaths forecasts for 31 European countries

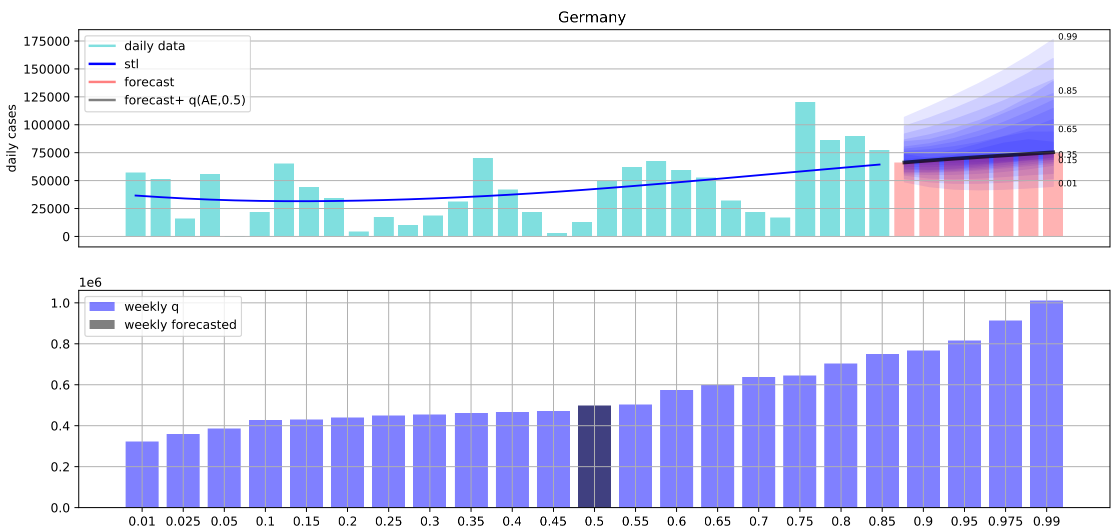

Fig. S12. Confidence intervals for Germany for one week ahead prediction obtained on January 14, 2022. The upper subplot demonstrates the forecast together with the predictive intervals; additional spline smoothing is applied for each confidence level to smooth the quantiles in time. Lower plot shows the estimated quantiles for the weekly number.

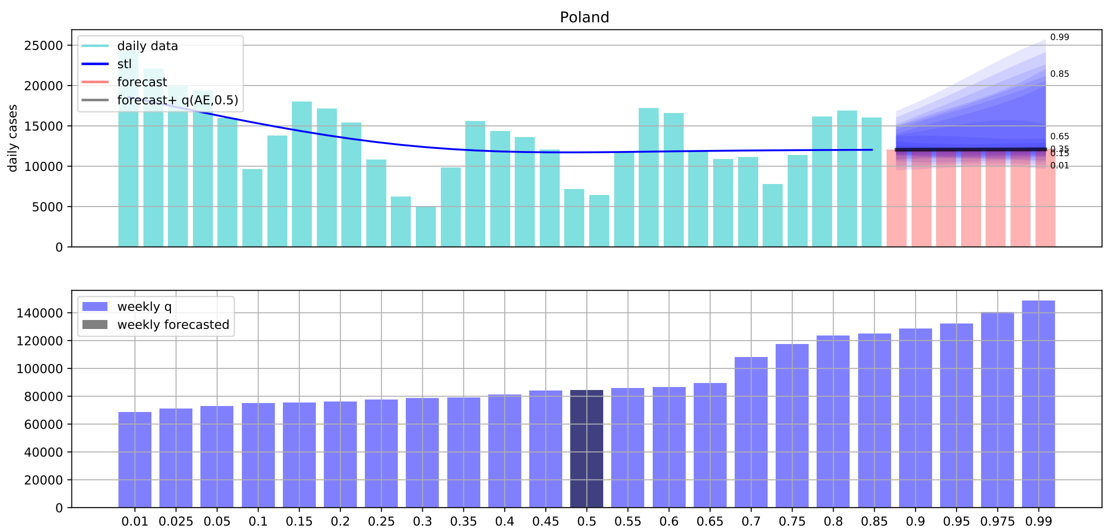

Fig. S13. Confidence intervals for Poland for one week ahead prediction obtained on January 14, 2022. The upper subplot demonstrates the forecast together with the predictive intervals; additional spline smoothing is applied for each confidence level to smooth the quantiles in time. Lower plot shows the estimated quantiles for the weekly number.

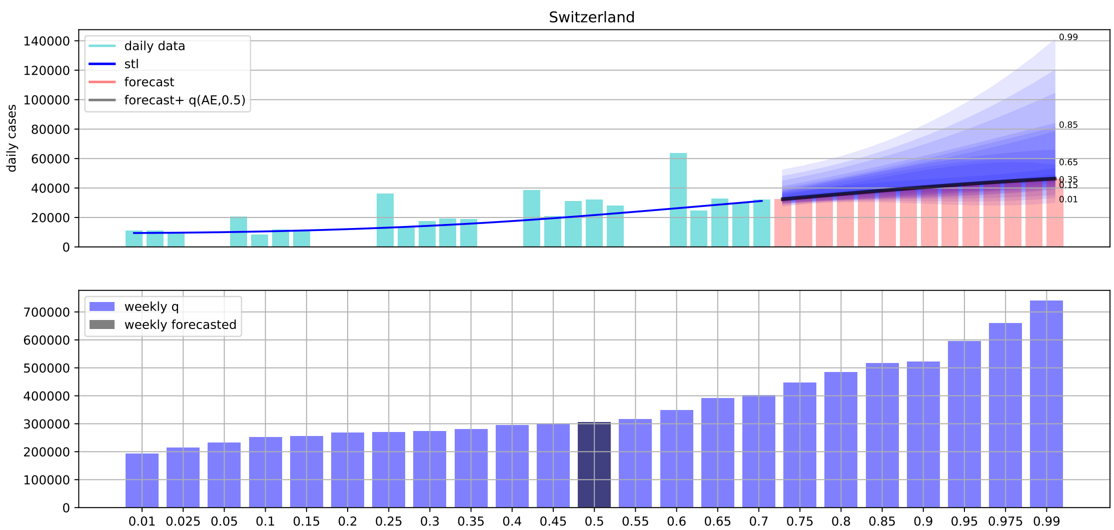

Fig. S14. Confidence intervals for Switzerland for two weeks ahead prediction obtained on January 14, 2022. The upper subplot demonstrates the forecast together with the predictive intervals; additional spline smoothing is applied for each confidence level to smooth the quantiles in time. Lower plot shows the estimated quantiles for weekly numbers of the second forecasted week.

country	EUHub-ens	IEM_Health	ILM	MUNI	RW	USC	SDSC_ISG
Austria	0.71	1.05	0.79	0.82	1.05	**0.64**	0.70
Belgium	0.81	1.11	0.85	1.09	1.33	**0.77**	1.49
Bulgaria	0.77	1.34	1.14	**0.53**	1.13	0.80	0.79
Croatia	0.70	1.16	0.94	0.72	0.94	0.83	**0.61**
Cyprus	0.85	**0.77**	2.28	0.93	1.13	0.99	0.81
Czechia	0.65	0.89	0.74	0.75	1.02	**0.61**	0.62
Denmark	0.74	0.76	0.98	0.94	1.00	0.89	**0.64**
Estonia	0.88	0.85	1.16	1.02	1.06	1.24	**0.81**
Finland	0.84	0.82	1.02	1.10	1.18	1.17	**0.80**
France	0.61	0.70	0.73	0.78	0.93	0.66	**0.54**
Germany	**0.60**	1.19	0.65	0.78	0.89	0.74	0.69
Greece	0.96	1.44	1.46	**0.84**	1.34	1.47	0.94
Hungary	0.66	1.06	0.79	**0.58**	1.10	0.65	0.64
Iceland	0.65	**0.43**	2.18	0.90	0.57	0.72	0.79
Ireland	1.25	1.45	1.98	1.27	1.40	**1.20**	1.31
Italy	0.52	0.63	0.60	**0.38**	1.08	0.69	0.47
Latvia	0.86	0.99	1.04	1.00	1.07	**0.68**	0.73
Liechtenstein	**0.71**	1.12	1.52	0.86	0.83	1.24	0.95
Lithuania	0.64	0.84	1.13	**0.48**	0.92	0.98	0.64
Luxembourg	**0.89**	1.14	1.24	1.06	1.16	1.37	1.07
Malta	1.10	1.52	5.11	**0.74**	1.21	1.45	1.20
Netherlands	0.69	0.81	1.09	**0.56**	1.07	1.00	0.84
Norway	0.70	0.71	0.86	0.62	1.06	0.94	**0.54**
Poland	0.49	0.92	0.64	**0.29**	1.00	0.73	0.53
Portugal	0.70	0.79	0.97	0.93	1.04	1.01	**0.63**
Romania	0.55	0.97	0.57	**0.33**	0.93	0.44	0.52
Slovakia	0.70	0.81	**0.67**	0.67	0.91	0.79	0.77
Slovenia	0.79	1.25	1.18	**0.48**	1.13	1.06	0.59
Spain	0.79	0.96	0.98	0.98	0.90	0.99	**0.60**
Sweden	0.87	1.00	0.90	1.24	0.86	0.69	**0.66**
Switzerland	**0.67**	0.67	0.81	0.89	0.95	0.72	0.70
ranks best in	4	2	1	10	0	5	9
in top 2	15	5	2	13	2	7	18
in top 3	28	8	3	16	4	9	25
in top 4	31	14	8	20	6	16	29

Table S1. One week ahead forecast AE normalized by the EuroCovidhub_baseline AE. The values, which are highlighted in bold and orange color, correspond to the best performance. The lower part of the table reports for each method the number of countries for which its forecast is best, or in the top 2, 3 or 4 best performing methods.

country	EUHub-ens	IEM_Health	ILM	MUNI	RW	USC	SDSC_ISG
Austria	0.55	0.99	0.54	0.78	1.06	0.61	**0.54**
Belgium	**0.61**	0.92	0.64	0.97	1.30	0.79	1.21
Bulgaria	0.57	1.11	0.88	**0.45**	1.06	0.80	0.66
Croatia	0.58	0.83	0.73	**0.56**	0.81	0.85	0.58
Cyprus	0.66	0.77	2.19	0.83	1.11	1.04	**0.59**
Czechia	0.49	0.67	0.58	0.68	1.11	0.46	**0.45**
Denmark	0.59	0.69	0.77	0.84	1.08	0.99	**0.50**
Estonia	0.77	**0.73**	0.88	1.03	1.07	1.03	0.77
Finland	0.64	**0.59**	0.73	1.01	1.22	1.32	0.64
France	**0.45**	0.60	0.65	0.65	1.00	0.76	0.47
Germany	0.46	0.86	**0.38**	0.57	0.77	0.68	0.47
Greece	0.76	1.18	1.15	**0.69**	1.39	1.38	0.86
Hungary	0.50	0.90	0.54	**0.47**	1.14	0.61	0.54
Iceland	0.73	1.30	1.42	0.57	**0.38**	1.41	1.22
Ireland	**1.03**	1.17	1.64	1.10	1.38	1.37	1.25
Italy	0.41	0.64	0.51	**0.31**	1.09	0.98	0.39
Latvia	0.62	0.73	0.69	0.96	0.94	0.58	**0.54**
Liechtenstein	0.68	0.81	1.09	**0.65**	0.83	1.24	0.73
Lithuania	0.49	0.72	0.95	**0.43**	0.98	0.98	0.61
Luxembourg	**0.73**	1.27	1.21	0.96	1.39	1.58	1.12
Malta	0.82	1.75	5.07	**0.58**	1.37	1.75	1.31
Netherlands	**0.55**	0.85	1.11	0.55	1.15	0.82	0.83
Norway	0.55	0.57	0.64	0.47	0.99	0.84	**0.35**
Poland	0.35	0.69	0.40	**0.23**	0.91	0.57	0.42
Portugal	**0.57**	0.64	0.73	1.05	1.08	1.23	0.62
Romania	0.36	0.81	0.36	**0.25**	0.86	0.38	0.34
Slovakia	0.54	0.57	**0.47**	0.60	0.87	0.75	0.61
Slovenia	0.64	0.95	0.91	**0.37**	1.13	0.88	0.54
Spain	0.70	0.97	1.27	1.15	1.13	1.22	**0.52**
Sweden	0.78	1.01	0.80	1.39	0.90	0.90	**0.62**
Switzerland	**0.53**	0.66	0.65	0.81	1.09	0.78	0.55
ranks best in	7	2	2	11	1	0	8
in top 2	21	2	4	16	1	2	16
in top 3	31	9	7	16	1	4	25
in top 4	31	14	19	20	3	8	29

Table S2. One week ahead forecast WIS normalized by the EuroCovidhub_baseline WIS. The values, which are highlighted in bold and orange color, correspond to the best performance. The lower part of the table reports for each method the number of countries for which its forecast is best, or in the top 2, 3 or 4 best performing methods.

country	EUHub-ens	IEM_Health	ILM	MUNI	RW	USC	SDSC_ISG
Austria	0.81	1.49	1.43	0.60	1.00	**0.56**	0.93
Belgium	**0.74**	1.45	1.31	1.21	1.21	0.91	1.23
Bulgaria	0.95	1.59	1.68	**0.62**	1.05	0.87	1.04
Croatia	0.88	1.66	1.56	**0.74**	0.98	0.94	0.92
Cyprus	1.04	0.88	3.69	**0.78**	1.13	0.91	0.94
Czechia	0.99	1.00	1.65	0.71	0.89	1.24	**0.68**
Denmark	0.69	0.86	1.30	1.04	1.00	0.97	**0.65**
Estonia	0.82	**0.82**	1.32	0.92	0.97	1.33	0.88
Finland	**0.88**	0.91	1.23	1.11	1.35	1.20	0.90
France	**0.64**	0.86	1.36	0.78	0.96	0.78	0.64
Germany	0.78	1.49	1.16	0.79	0.87	**0.56**	0.76
Greece	1.34	1.71	1.96	**0.90**	1.43	1.44	1.15
Hungary	0.99	1.57	1.52	0.74	1.03	**0.69**	0.71
Iceland	0.47	**0.41**	2.73	1.00	0.95	0.92	1.00
Ireland	1.51	1.78	3.32	**1.35**	1.48	1.43	1.58
Italy	**0.54**	0.77	0.99	0.56	1.07	0.77	0.56
Latvia	0.92	1.25	1.32	0.99	0.93	**0.62**	0.79
Liechtenstein	**0.85**	1.75	1.38	0.96	0.91	1.77	1.01
Lithuania	0.61	1.03	1.51	**0.52**	0.91	0.92	0.78
Luxembourg	**0.85**	1.07	1.03	1.05	1.10	2.09	1.04
Malta	1.67	1.90	9.98	**0.77**	1.13	1.69	1.40
Netherlands	1.28	0.89	3.01	**0.86**	1.04	1.05	0.89
Norway	0.85	0.93	1.12	0.69	1.09	1.43	**0.66**
Poland	0.79	1.29	1.37	**0.43**	1.01	1.03	0.49
Portugal	0.82	0.91	1.31	0.97	1.08	1.07	**0.74**
Romania	0.73	1.28	1.11	0.52	0.86	**0.46**	0.71
Slovakia	**0.54**	0.63	0.56	0.62	0.88	0.66	0.71
Slovenia	1.15	1.56	1.70	**0.58**	1.08	1.66	0.92
Spain	1.00	1.11	1.70	0.96	0.94	1.10	**0.82**
Sweden	0.77	1.04	0.85	1.19	0.84	**0.60**	0.61
Switzerland	**0.67**	1.02	1.19	0.91	1.01	0.82	0.71
ranks best in	8	2	0	10	0	6	5
in top 2	14	3	2	14	3	9	17
in top 3	22	7	2	19	7	13	23
in top 4	29	10	2	27	14	15	27

Table S3. Two week ahead forecast AE normalized by the EuroCovidhub_baseline AE. The values, which are highlighted in bold and orange color, correspond to the best performance. The lower part of the table reports for each method the number of countries for which its forecast is best, or in the top 2, 3 or 4 best performing methods.

country	EUHub-ens	IEM_Health	ILM	MUNI	RW	USC	SDSC_ISG
Austria	0.61	1.44	1.03	**0.45**	1.10	0.53	0.69
Belgium	**0.61**	1.27	1.03	1.10	1.33	1.01	1.09
Bulgaria	0.73	1.49	1.46	**0.50**	1.12	0.92	1.02
Croatia	0.67	1.22	1.15	**0.53**	0.89	0.92	0.88
Cyprus	0.84	1.03	3.83	**0.69**	1.21	1.10	0.81
Czechia	0.81	0.83	1.41	0.67	1.04	0.84	**0.66**
Denmark	**0.58**	0.84	1.07	0.94	1.20	1.06	0.63
Estonia	**0.75**	0.80	1.10	0.94	1.08	1.02	0.98
Finland	**0.75**	0.76	1.07	1.04	1.56	1.50	0.84
France	**0.54**	0.83	1.37	0.73	1.20	1.01	0.59
Germany	0.57	1.20	0.73	0.59	0.85	**0.55**	0.59
Greece	0.98	1.53	1.70	**0.73**	1.63	1.49	1.14
Hungary	0.77	1.51	1.13	**0.54**	1.18	0.68	0.58
Iceland	0.81	1.52	1.94	0.55	0.52	1.63	1.03
Ireland	1.37	1.81	3.44	**1.17**	1.58	1.97	1.40
Italy	0.51	0.85	0.90	**0.44**	1.23	1.15	0.61
Latvia	0.73	1.07	1.00	0.98	0.85	**0.59**	0.60
Liechtenstein	0.78	1.49	1.06	**0.72**	0.99	1.93	0.77
Lithuania	0.52	0.97	1.36	**0.52**	1.06	0.98	0.91
Luxembourg	**0.84**	1.30	1.19	1.01	1.42	2.28	1.41
Malta	1.25	2.33	10.50	**0.69**	1.37	2.36	1.33
Netherlands	1.24	1.00	3.10	**0.80**	1.20	0.93	1.02
Norway	0.69	0.83	0.87	0.52	1.11	1.31	**0.48**
Poland	0.59	1.13	1.04	0.31	1.10	0.74	0.41
Portugal	**0.69**	0.83	1.14	1.11	1.28	1.47	0.73
Romania	0.52	1.21	0.79	0.39	0.82	0.40	0.50
Slovakia	0.37	0.46	**0.36**	0.45	0.88	0.61	0.49
Slovenia	0.94	1.30	1.34	**0.42**	1.16	1.35	0.83
Spain	0.93	1.19	2.34	1.09	1.28	1.41	**0.89**
Sweden	0.69	1.06	0.70	1.31	0.80	0.77	**0.67**
Switzerland	**0.55**	1.08	0.98	0.80	1.27	1.05	0.57
ranks best in	8	0	1	15	1	2	4
in top 2	19	2	1	19	1	6	14
in top 3	28	5	4	25	1	8	22
in top 4	30	14	5	28	7	11	29

Table S4. Two week ahead forecast WIS normalized by the EuroCovidhub_baseline WIS. The values, which are highlighted in bold and orange color, correspond to the best performance. The lower part of the table reports for each method the number of countries for which its forecast is best, or in the top 2, 3 or 4 best performing methods.

References

1. EY Cramer, et al., Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *medRxiv* (2021).