Supplementary Information for Global predictions of short- to medium-term COVID-19 transmission trends : a retrospective assessment

August 10, 2021

Contents

1 Overview

In this supplement, we present additional results on the performance assessment of the model for short- (SI2 Sec. 2) and medium-term forecasts (SI2 Sec. 3).

2 Model performance assessment

The following metrics were used to assess the model performance:

- **Mean relative error** The mean relative error (MRE) is a widely used measure of model accuracy [1]. The mean relative error for the forecasts $\hat{D}_t$ at time t is defined as:

$$MRE_t(D_t, \hat{D}_t) = \frac{\sum_{s=1}^{N} |D_t - \hat{D}_{t,s}|}{N * (D_t + 1)},$$

  where D_t denotes the observed deaths at time t, N is the number of simulated trajectories and $\hat{D}_{t,s}$ denotes the s^{th} simulation at time t [2]. That is the mean relative error at time t is averaged across all simulated trajectories and normalised by the observed incidence. We add 1 to the observed value to prevent division by 0. A MRE value of k means that the average error is k times the observed value.

- **Comparison with null model** Compare the absolute error made by the model with the absolute error made by a null model that uses the average of the last 10 observations as the forecast for the week ahead. We also compared the model error with the error made by a linear model (forecasts from a line fitted to the last 10 observations).

- **Coverage probability** Coverage probability refers to the proportion of observations that are contained in given credible interval (CrI) of the distribution of forecasts. For a well-calibrated model, 50% of the observations should be contained in the 50% CrI [3]. For a X% CrI, coverage probability higher than X% indicates that the model is under-confident while a value less than X% suggests that the model is over-confident with narrow CrIs.

For each country and for each week, the time series of observed deaths was first smoothed by taking a 3-day rolling mean. The average of the daily MRE was used as the weekly MRE.

2.1 Mean relative error by epidemic phase

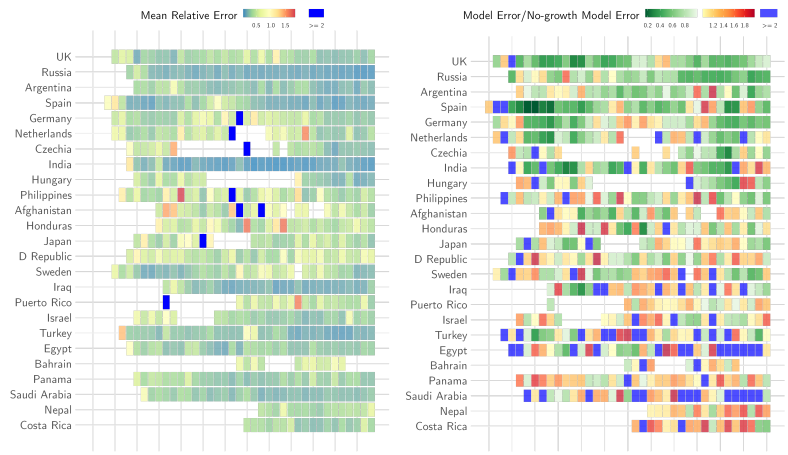
Epidemic phase	Proportion in 50% CrI	Proportion in 95% CrI	MRE
Definitely decreasing	48.5% (29.9%)	83.1% (24.3%)	0.4 (0.3)
Likely decreasing	60.7% (33.1%)	91.2% (20.2%)	0.4 (0.3)
Definitely growing	48.6% (31.9%)	84.7% (24.7%)	0.4 (0.7)
Likely growing	62.2% (31.0%)	92.4% (19.4%)	0.5 (0.5)
Indeterminate	66.1% (31.2%)	92.4% (18.9%)	0.5 (0.6)

Table 1. Coverage probability and mean relative error of short-term forecasts in each epidemic phase. The values show the average of the metric across countries and weeks of forecast. The standard deviation is shown in parentheses.

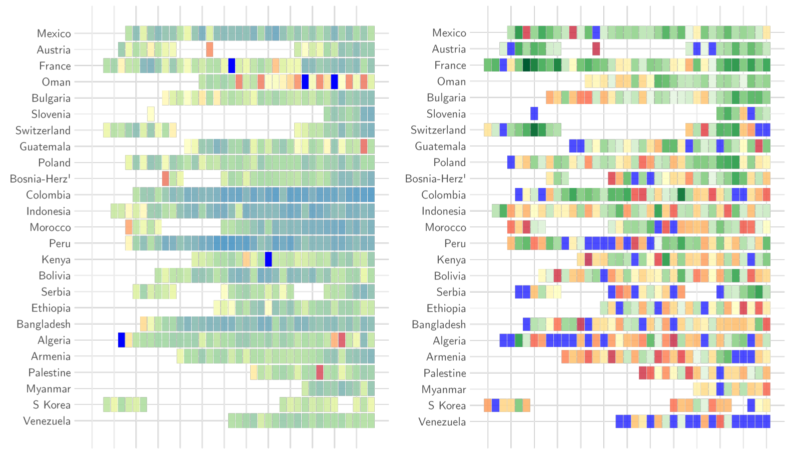
2.2 Relative error and comparison with no-growth model

This section presents the mean relative error of the model and comparison of the model error with the error made by a model that uses the average of the past 10 days as the forecast for the week ahead.

(a)


Mean Relative Error
0.5 1.0 1.5
>= 2
Model Error/No-growth Model Error
0.2 0.4 0.6 0.8
1.2 1.4 1.6 1.8
>= 2
UK
Russia
Argentina
Spain
Germany
Netherlands
Czechia
India
Hungary
Philippines
Afghanistan
Honduras
Japan
D Republic
Sweden
Iraq
Puerto Rico
Israel
Turkey
Egypt
Bahrain
Panama
Saudi Arabia
Nepal
Costa Rica

(b)



(c)

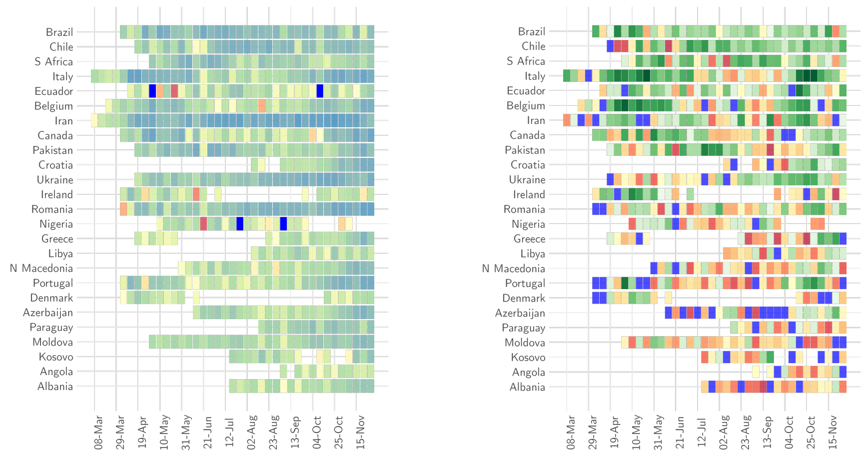


Figure 1. Mean relative error and comparison with null model In each panel, the left graph shows the relative error of the ensemble model for each week of forecast (x-axis) and for each country (y-axis). Dark blue cells indicate weeks where the relative error of the model was greater than 2. The right panel shows the ratio of the absolute error of the model to the absolute error of the no-growth null model. Shades of green show weeks for a given country where the ratio was smaller than 1 i.e., the model error was smaller, and weeks where the ratio was greater than 1 i.e. the model error was bigger than the null model error are shown in shades of red (yellow to red). Dark blue cells indicate weeks where the ratio was greater than 2. Panels (a) - (c) show results for all countries included in the analysis.

2.2.1 Comparison with no-growth and linear models by phase

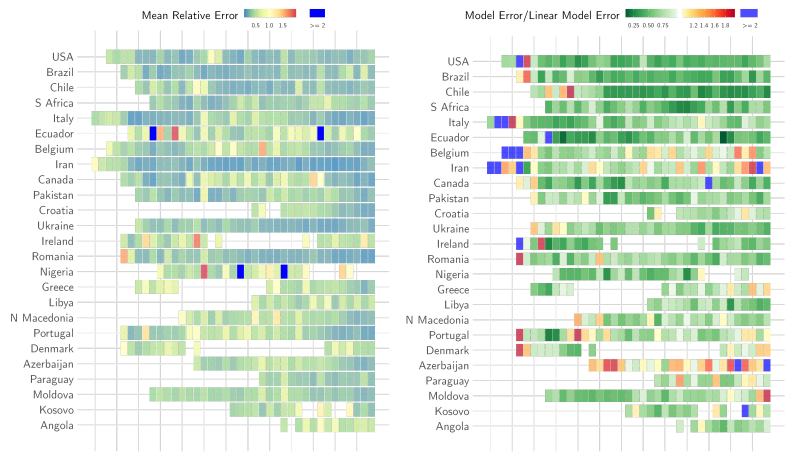
Phase	Ensemble model error <No-growth model error	Ensemble model error <Linear model error	Weeks
Likely decreasing	54.5%	88.4%	224
Definitely decreasing	80.9%	96.4%	251
Likely growing	31.9%	74.8%	301
Definitely growing	61.4%	70.3%	542
Indeterminate	32.9%	80.7%	887

Table 2. Comparison of the absolute error of the ensemble model with that made by a null no-growth model or a predictions from a linear model as forecast for the week ahead for each phase of the pandemic. The rightmost column (Weeks) shows the total number of weeks in a given phase.

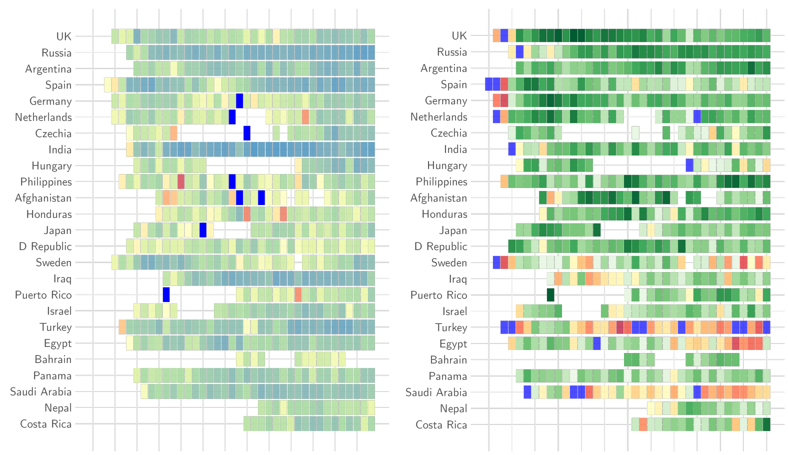
2.3 Relative error and comparison with a linear model

This section presents the relative error of the ensemble model and comparison of the model error with the error of a linear model (a line fitted to the past 10 observations). The linear model was fitted in rstannarm [4] and the forecasts were sampled from the posterior predictive distribution.

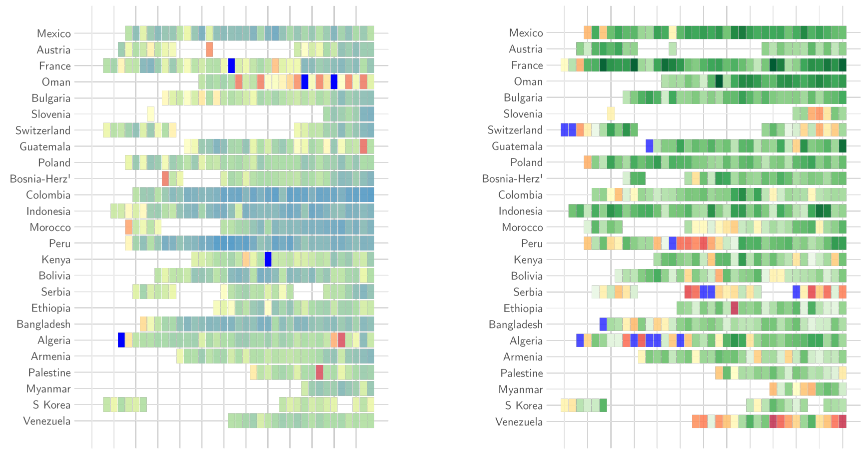
(a)


Mean Relative Error
0.5 1.0 1.5
>= 2
Model Error/Linear Model Error
0.25 0.50 0.75
1.2 1.4 1.6 1.8
>= 2
USA
Brazil
Chile
S Africa
Italy
Ecuador
Belgium
Iran
Canada
Pakistan
Croatia
Ukraine
Ireland
Romania
Nigeria
Greece
Libya
N Macedonia
Portugal
Denmark
Azerbaijan
Paraguay
Moldova
Kosovo
Angola

(b)


UK
Russia
Argentina
Spain
Germany
Netherlands
Czechia
India
Hungary
Philippines
Afghanistan
Honduras
Japan
D Republic
Sweden
Iraq
Puerto Rico
Israel
Turkey
Egypt
Bahrain
Panama
Saudi Arabia
Nepal
Costa Rica

(c)



(d)

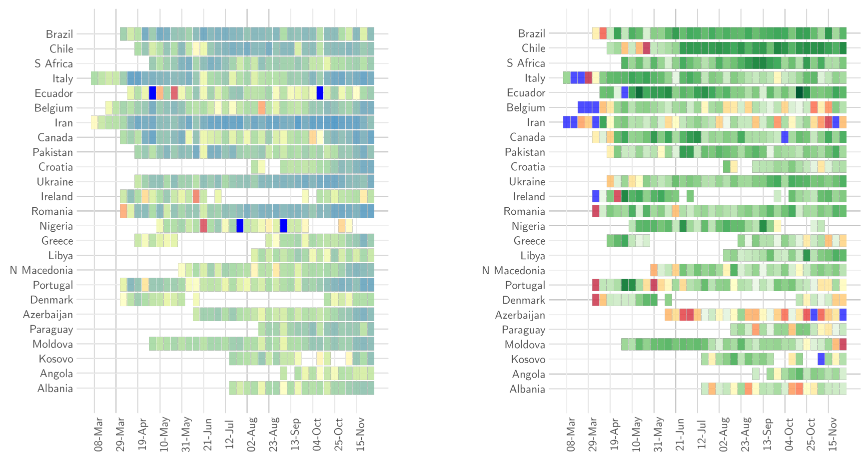


Figure 2. Relative error and comparison with a linear model In each panel, the left graph shows the mean relative error of the model for each week of forecast (x-axis) and for each country (y-axis). Dark blue cells indicate weeks where the relative error of the model was greater than 2. The right panel shows the ratio of the absolute error of the model to the absolute error of forecasts made using a linear model. Shades of green show weeks for a given country where the ratio was smaller than 1 i.e., the model error was smaller, and weeks where the ratio was greater than 1 i.e. the model error was bigger than the null model error are shown in shades of red (yellow to red). Dark blue cells indicate weeks where the ratio was bigger than 2. Panels (a)-(d) show results for all countries included in the analysis.

2.4 Mean relative error compared with the weekly CV

The relative error of the model was proportional to the CV of the number of deaths reported each week and inversely proportional to the weekly incidence.

(a) **(b)**

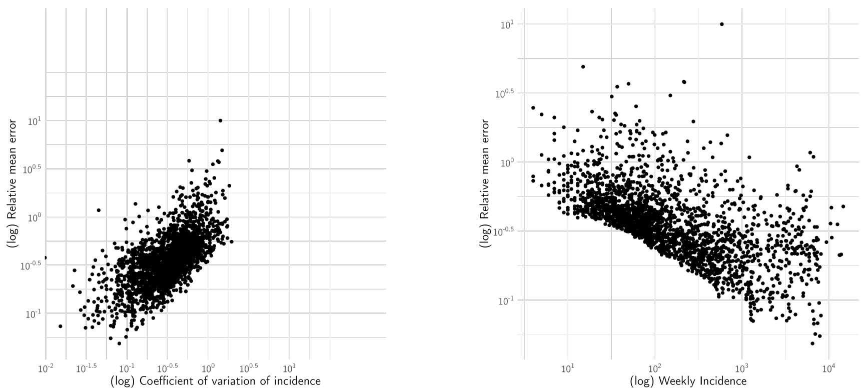


Figure 3. The log MRE scales linearly with the log weekly CV (a) and inversely with the log weekly incidence (b).

2.5 Coverage Probability

This section presents the proportion of observations in 50% CrI and 95% CrI for each country and each week of forecast.

Proportion of observations in 50% CrI

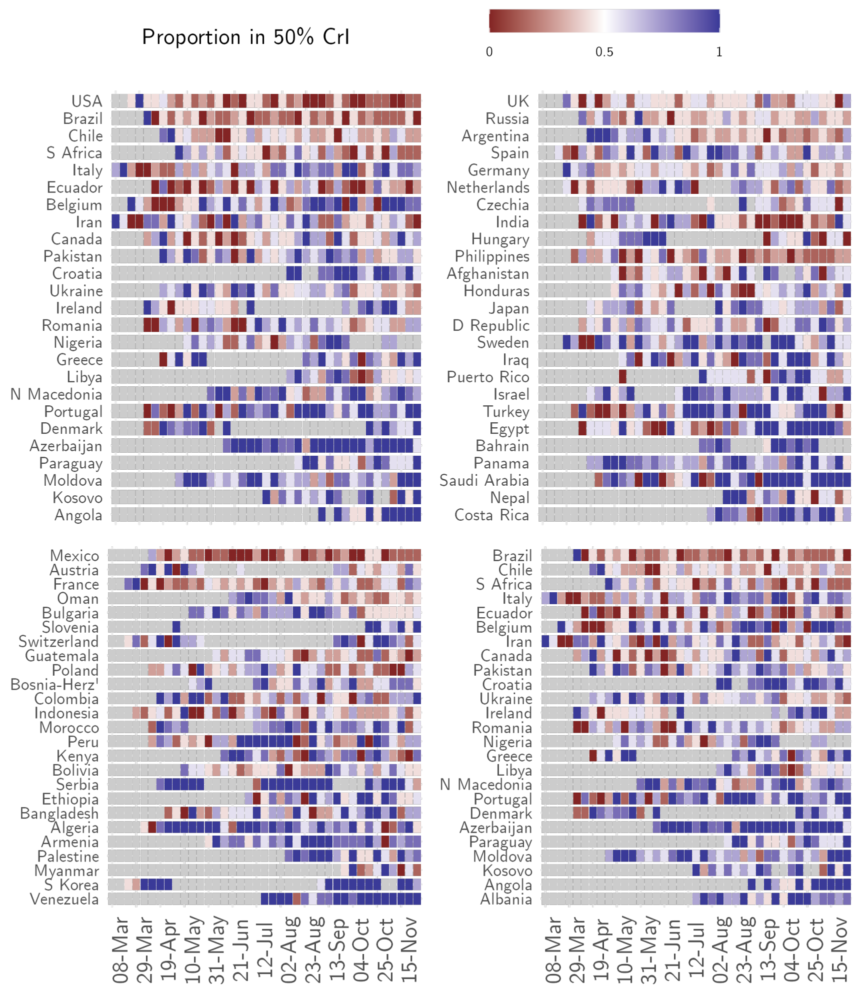


Figure 4. For each week of forecast (x-axis) and each country (y-axis), the proportion of observations in the 50% CrI of the forecasts. Gray cells indicate weeks where a country was not included in the analysis because the number of deaths did not meet the threshold (SI Sec. 5).

Proportion of observations in 95% CrI

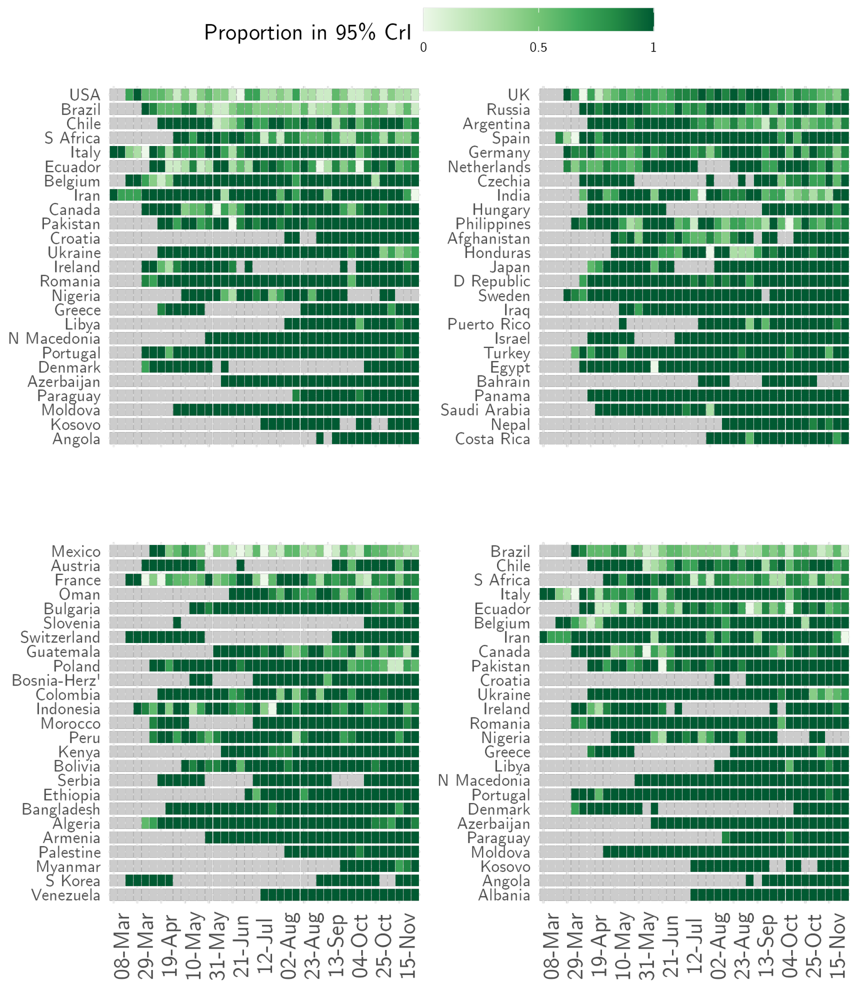


Figure 5. For each week of forecast (x-axis) and each country (y-axis), the proportion of observations in 95% CrI of the forecasts. Gray cells indicate weeks where a country was not included in the analysis because the number of deaths did not meet the threshold (SI Sec. 5).

3 Medium-term forecasts

This section presents the performance assessment results for medium-term forecasts. The relative error for each country and week of forecast are presented in (SI2 Sec. 3.1) and coverage probability are shown in (SI2 Sec. 3.2).

3.1 Relative error

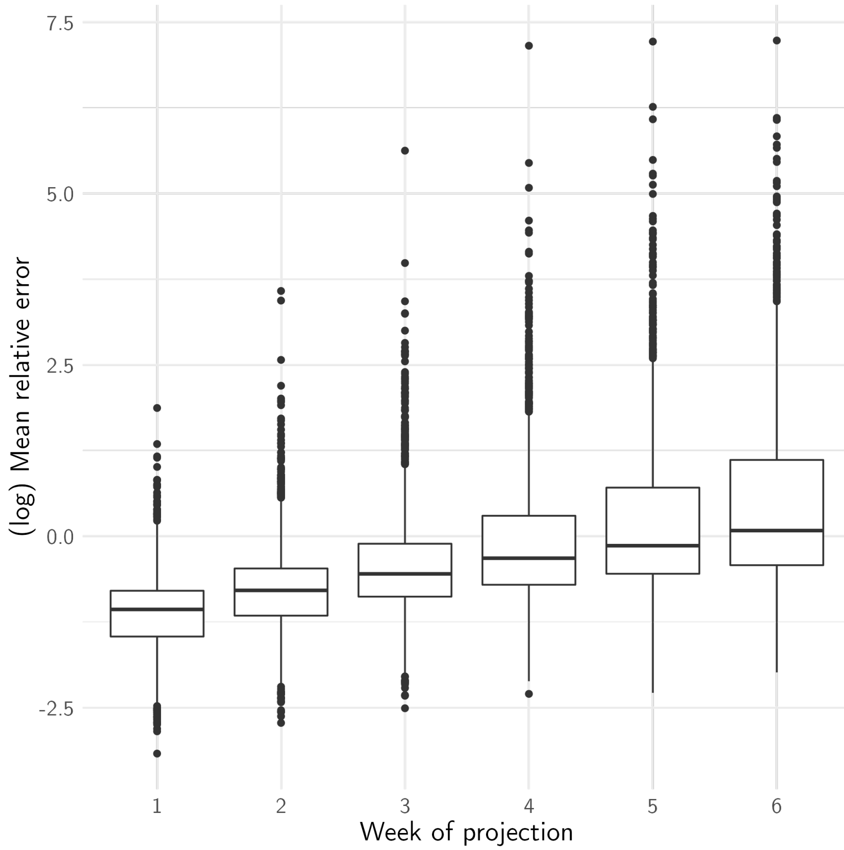


Figure 6. The mean relative error grew over the projection horizon becoming unaccpetably high beyond a 4-week horizon.

(a)


Mean Relative Error
0.5 1.0 1.5
>= 2
1-week ahead
2-weeks ahead
3-weeks ahead
4-weeks ahead
UK
Russia
Argentina
Spain
Germany
Netherlands
Czechia
India
Hungary
Philippines
Afghanistan
Honduras
Japan
D Republic
Sweden
Iraq
Puerto Rico
Israel
Turkey
Egypt
Bahrain
Panama
Saudi Arabia
Nepal
Costa Rica

(b)


1-week ahead
2-weeks ahead
3-weeks ahead
4-weeks ahead
Mexico
Austria
France
Oman
Bulgaria
Slovenia
Switzerland
Guatemala
Poland
Bosnia-Herz'
Colombia
Indonesia
Morocco
Peru
Kenya
Bolivia
Serbia
Ethiopia
Bangladesh
Algeria
Armenia
Palestine
Myanmar
S Korea
Venezuela

(c)

Figure 7. Mean relative error of medium-term forecasts. The relative error of the model in 1-week, 2-week , 3-week, and 4-week ahead forecasts for each week of forecast (x-axis) and for each country (y-axis). Dark blue cells indicate weeks where the relative error of the model was greater than 2. Panels (a)-(c) present results for all countries included in the analysis.

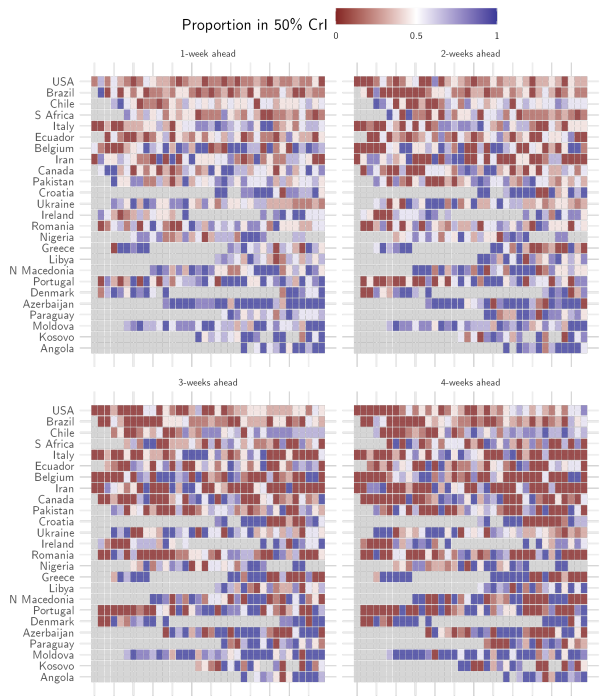
Week of forecast	MRE <0.5	MRE <1
1	80.8%	91.1%
2	58.3%	89.5 %
3	33.2%	78.3%
4	25.6%	66.0%

Table 3. The MRE of medium-term forecasts remained relatively small over a 4-week forecast horizon. The MRE was less than 1 in 66.0% and less than 0.5 in 25.6% of weeks in 4-week ahead forecasts.

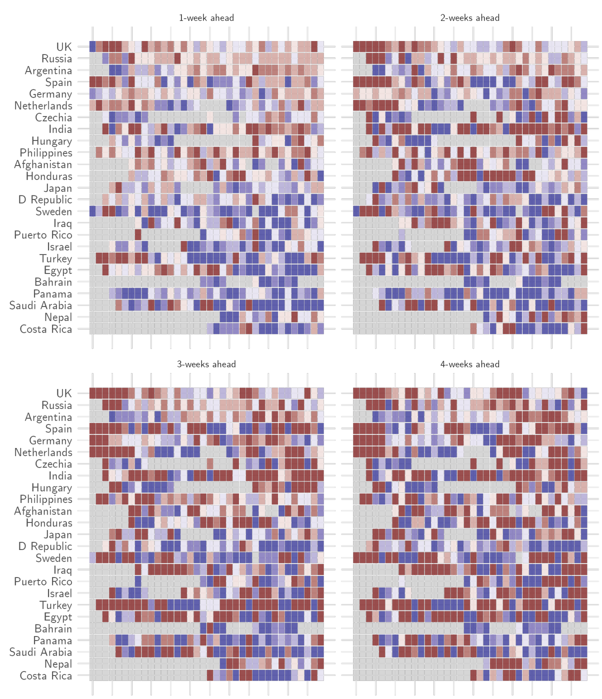
3.2 Coverage Probability

Proportion of observations in 50% CrI

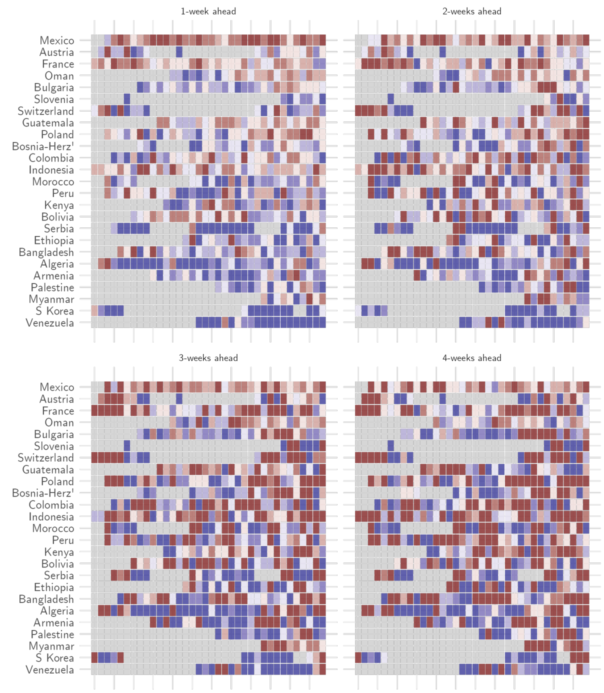
(a)


Proportion in 50% CrI
0
0.5
1
1-week ahead
2-weeks ahead
3-weeks ahead
4-weeks ahead
USA
Brazil
Chile
S Africa
Italy
Ecuador
Belgium
Iran
Canada
Pakistan
Croatia
Ukraine
Ireland
Romania
Nigeria
Greece
Libya
N Macedonia
Portugal
Denmark
Azerbaijan
Paraguay
Moldova
Kosovo
Angola

(b)


1-week ahead
2-weeks ahead
3-weeks ahead
4-weeks ahead
UK
Russia
Argentina
Spain
Germany
Netherlands
Czechia
India
Hungary
Philippines
Afghanistan
Honduras
Japan
D Republic
Sweden
Iraq
Puerto Rico
Israel
Turkey
Egypt
Bahrain
Panama
Saudi Arabia
Nepal
Costa Rica

(c)



(d)

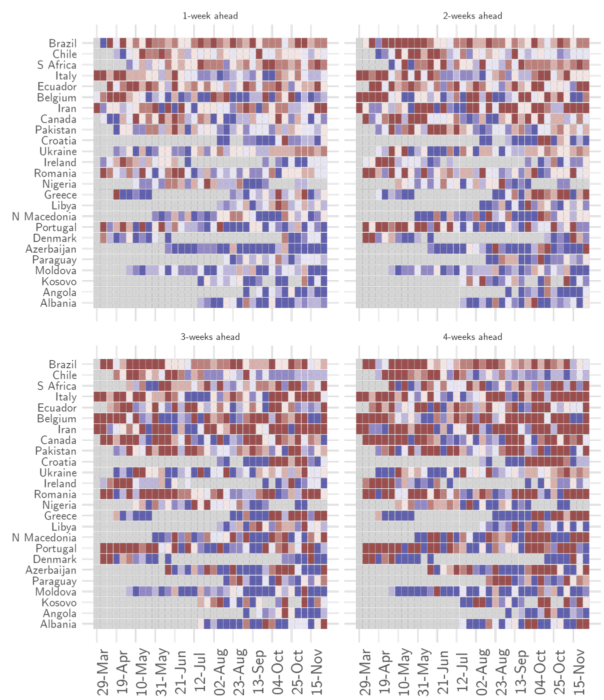
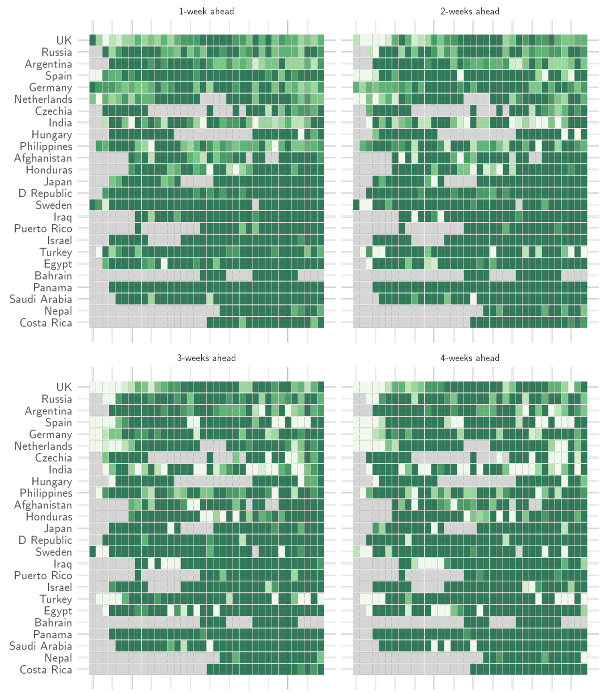


Figure 8. The proportion of observations in the 50% CrI of the forecasts for 1-week, 2-week , 3-week, and 4-week (clockwise from top left) ahead for each week of forecast (x-axis) and for each country (y-axis). Panels (a)-(d) present results for all countries included in the analysis. Gray cells indicate weeks where a country was not included in the analysis because the number of deaths did not meet the threshold (SI Sec. 5).

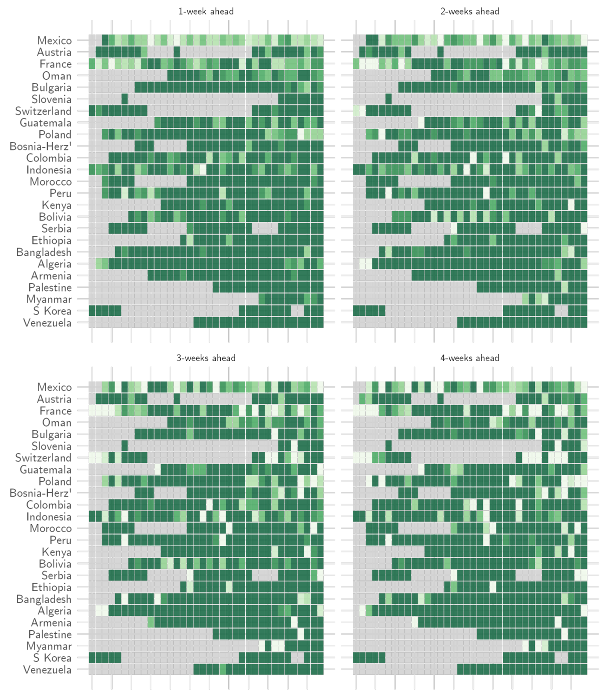
Proportion of observations in 95% CrI

(a)

(b)


1-week ahead
2-weeks ahead
3-weeks ahead
4-weeks ahead
UK
Russia
Argentina
Spain
Germany
Netherlands
Czechia
India
Hungary
Philippines
Afghanistan
Honduras
Japan
D Republic
Sweden
Iraq
Puerto Rico
Israel
Turkey
Egypt
Bahrain
Panama
Saudi Arabia
Nepal
Costa Rica

(c)



(d)

Figure 9. The proportion of observations in the 95% CrI of the forecasts for 1-week, 2-week, 3-week, and 4-week ahead for each week of forecast (x-axis) and for each country (y-axis). Panels (a)-(d) present results for all countries included in the analysis. Gray cells indicate weeks where a country was not included in the analysis because the number of deaths did not meet the threshold (SI Sec. 5).

Medium-term phase

Misclassified epidemic phase

	Phase using R^S				
Phase using R^{curr}	Definitely decreasing	Likely decreasing	Definitely growing	Likely growing	Indeterminate
Definitely decreasing	0.00% (0)	100.00% (253)	0.00% (0)	0.00% (0)	0.00% (0)
Likely decreasing	72.73% (328)	0.00% (0)	0.00% (0)	0.00% (0)	27.27% (123)
Definitely growing	0.00% (0)	0.00% (0)	0.00% (0)	56.29% (1513)	43.71% (1175)
Likely growing	0.00% (0)	0.00% (0)	0.92% (30)	0.00% (0)	99.08% (3239)
Indeterminate	1.68% (31)	79.35% (1460)	0.00% (0)	18.97% (349)	0.00% (0)

Table 4. In country-days where the phase definitions using R_t^{curr} (shown along rows) and R_t^S (shown along columns) were different, R_t^S most frequently mis-classified the phase as a phase with greater uncertainty. The numbers in parenthesis indicate the number of country-days for a given combination of phase in row and column.

References

[1] Tofallis C. A better measure of relative prediction accuracy for model selection and model estimation. Journal of the Operational Research Society. 2015;66(8):1352–1362. doi:10.1057/jors.2014.124.

[2] Bhatia S, Lassmann B, Cohn E, Desai AN, Carrion M, Kraemer MU, et al. Using digital surveillance tools for near real-time mapping of the risk of infectious disease spread. NPJ Digital Medicine. 2021;4(1):1–10. doi:10.1038/s41746-021-00442-3.

[3] Funk S, Camacho A, Kucharski AJ, Lowe R, Eggo RM, Edmunds WJ. Assessing the performance of real-time epidemic fore-

casts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. PLoS Computational Biology. 2019;15(2):e1006785. doi:10.1371/journal.pcbi.1006785.

[4] Goodrich B, Gabry J, Ali I, Brilleman S. rstanarm: Bayesian applied regression modeling via Stan.; 2020. Available from: https://mc-stan.org/rstanarm.