The influence of model structure and geographic specificity on predictive accuracy among European COVID-19 forecasts

Supplementary information

Contents

1 Code and data availability

1.1 Code

The codebase for this paper is publicly available at:

- Github: https://github.com/epiforecasts/eval-by-method
- Zenodo with DOI: https://doi.org/10.5281/zenodo.14903162

Comments and code contributions are welcome - please use Github Issues.

Please cite code using:

- Katharine Sherratt & Sebastian Funk. (2025). epiforecasts/eval-by-method: Zenodo. https://doi.org/10.5281/zenodo.14903162

1.2 Source data

Forecast and observed data were sourced from the European COVID-19 Forecast Hub, available to view at https://covid19forecasthub.eu/ . All Hub data are now archived at:

- Github: https://github.com/european-modelling-hubs/covid19-forecast-hub-europe_archive
- Zenodo with DOI: https://doi.org/10.5281/zenodo.13986751

Data for this work were downloaded on 30th May 2023. These data are available in the Github repository for this paper at: https://github.com/epiforecasts/eval-by-method/tree/main/data

2 Model characteristics

2.1 Eligibility criteria

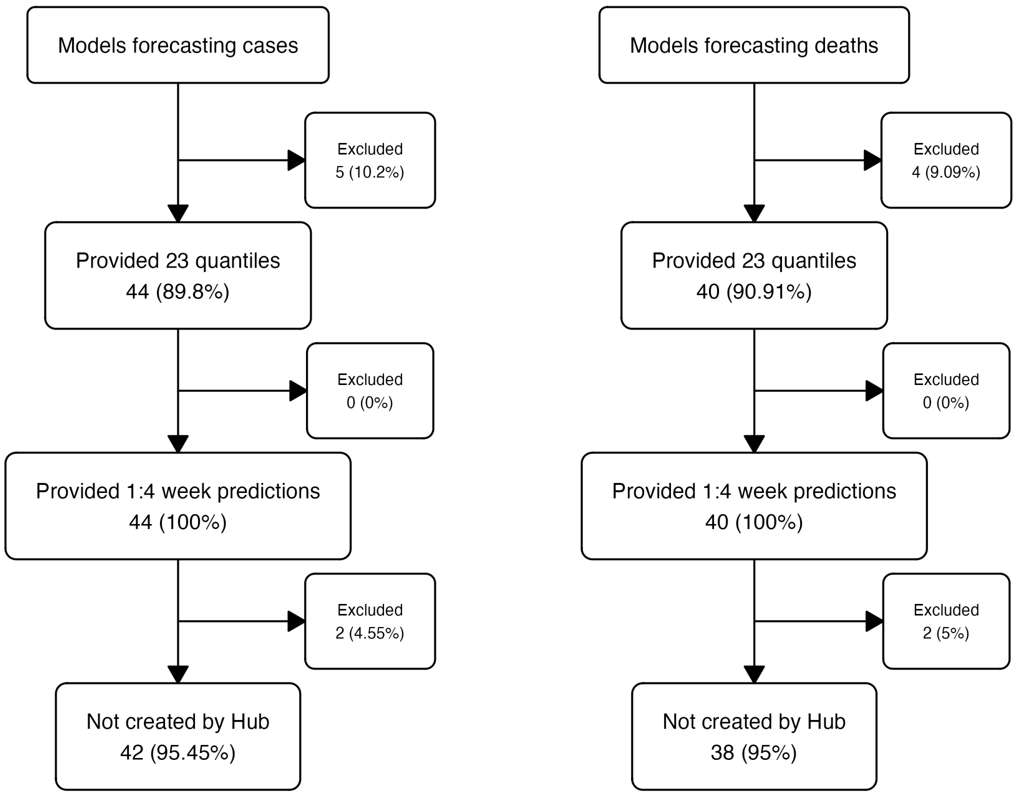


Figure 1: Eligibility criteria for models contributing case (left) and death (right) forecasts to the European COVID-19 Forecast Hub, March 2021 - March 2023

2.2 Model characteristics

Table 1: Model characteristics contributing to the European COVID-19 Forecast Hub, by method used, number of countries targeted, and number of forecasts contributed.

Model	Method	Country Targets	Case forecasts	Death forecasts
AMM-EpiInvert	Statistical	Multi-country	2,788 (1.3%)	
CovidMetrics-epiBATS	Statistical	Single-country	343 (0.2%)	
DSMPG-bayes	Semi-mechanistic	Multi-country	760 (0.4%)	
EuroCOVIDhub-baseline	Statistical	Multi-country	13,082 (6.3%)	13,040 (6.3%)
FIAS_FZJ-Epi1Ger	Mechanistic	Single-country	264 (0.1%)	264 (0.1%)
GoeWroc-BaseBayes	Semi-mechanistic	Single-country	12 (0%)	
HZI-AgeExtendedSEIR	Mechanistic	Single-country	382 (0.2%)	382 (0.2%)
ICM-agentModel	Agent-based	Single-country	334 (0.2%)	334 (0.2%)
IEM_Health-CovidProject	Mechanistic	Multi-country	7,710 (3.7%)	7,708 (3.7%)
ILM-EKF	Semi-mechanistic	Multi-country	11,998 (5.8%)	11,961 (5.8%)
ITWW-county_repro	Semi-mechanistic	Single-country	650 (0.3%)	600 (0.3%)
Imperial-DeCa	Semi-mechanistic	Multi-country		571 (0.3%)
Imperial-RtI0	Semi-mechanistic	Multi-country		571 (0.3%)
Imperial-sbkp	Semi-mechanistic	Multi-country		571 (0.3%)
JBUD-HMXK	Mechanistic	Multi-country	1,324 (0.6%)	1,324 (0.6%)
KITmetricslab-bivar_branching	Statistical	Single-country	8 (0%)	
Karlen-pypm	Mechanistic	Multi-country	3,208 (1.5%)	3,186 (1.5%)
LANL-GrowthRate	Semi-mechanistic	Multi-country	3,692 (1.8%)	3,696 (1.8%)
LeipzigIMISE-SECIR	Mechanistic	Single-country	16 (0%)	16 (0%)
MIMUW-StochSEIR	Mechanistic	Single-country	76 (0%)	76 (0%)
MIT_CovidAnalytics-DELPHI	Mechanistic	Single-country	348 (0.2%)	500 (0.2%)
MOCOS-agent1	Agent-based	Single-country	386 (0.2%)	386 (0.2%)
MUNI-ARIMA	Statistical	Multi-country	10,979 (5.3%)	11,314 (5.4%)
MUNI-LaggedRegARIMA	Statistical	Multi-country		736 (0.4%)
MUNI-VAR	Statistical	Multi-country	976 (0.5%)	976 (0.5%)
MUNI_DMS-SEIAR	Mechanistic	Single-country	224 (0.1%)	200 (0.1%)
PL_GRedlarski-DistrictsSum	Mechanistic	Single-country	378 (0.2%)	
RobertWalraven-ESG	Statistical	Multi-country	9,190 (4.4%)	10,465 (5%)
SDSC_ISG-TrendModel	Statistical	Multi-country	1,756 (0.8%)	1,744 (0.8%)
UB-BSLCoV	Statistical	Single-country	96 (0%)	96 (0%)
UC3M-EpiGraph	Agent-based	Single-country	94 (0%)	
ULZF-SEIRC19SI	Mechanistic	Single-country	249 (0.1%)	249 (0.1%)
UMass-MechBayes	Mechanistic	Multi-country		5,948 (2.9%)
UMass-SemiMech	Semi-mechanistic	Multi-country	1,888 (0.9%)	1,904 (0.9%)
UNED-PreCoV2	Statistical	Single-country	147 (0.1%)	147 (0.1%)
UNIPV-BayesINGARCHX	Statistical	Multi-country	426 (0.2%)	
USC-SIkJalpha	Mechanistic	Multi-country	12,900 (6.2%)	12,688 (6.1%)
UpgUmibUsi-MultiBayes	Semi-mechanistic	Single-country	99 (0%)	99 (0%)
bisop-seirfilter	Mechanistic	Single-country	32 (0%)	32 (0%)
bisop-seirfilterlite	Mechanistic	Multi-country	336 (0.2%)	336 (0.2%)
epiMOX-SUIHTER	Mechanistic	Single-country	134 (0.1%)	134 (0.1%)
epiforecasts-EpiExpert	Judgement	Multi-country	945 (0.5%)	948 (0.5%)
epiforecasts-EpiExpert_Rt	Judgement	Multi-country	404 (0.2%)	404 (0.2%)
epiforecasts-EpiExpert_direct	Judgement	Multi-country	394 (0.2%)	392 (0.2%)
epiforecasts-EpiNow2	Semi-mechanistic	Multi-country	8,843 (4.3%)	7,721 (3.7%)
epiforecasts-weeklygrowth	Statistical	Multi-country	5,971 (2.9%)	
itwm-dSEIR	Mechanistic	Single-country	406 (0.2%)	406 (0.2%)
prolix-euclidean	Semi-mechanistic	Multi-country	800 (0.4%)	800 (0.4%)

Table 2: Classification of models by number of raters in total and agreement on model structure

Model	Final classification	Agreement	Total raters	Semi-mechanistic	Mechanistic	Age
AMM-EpiInvert	Statistical	FALSE	3	0	0	
DSMPG-bayes	Semi-mechanistic	FALSE	3	2	0	
ILM-EKF	Semi-mechanistic	FALSE	3	2	0	
ITWW-county_repro	Semi-mechanistic	FALSE	3	2	0	
Imperial-DeCa	Semi-mechanistic	FALSE	3	2	0	
Imperial-RtI0	Semi-mechanistic	FALSE	3	2	0	
Imperial-sbkp	Semi-mechanistic	FALSE	3	2	0	
KITmetricslab-bivar_branching	Statistical	FALSE	3	1	0	
Karlen-pypm	Mechanistic	FALSE	3	0	2	
LANL-GrowthRate	Semi-mechanistic	FALSE	3	2	0	
SDSC_ISG-TrendModel	Statistical	FALSE	3	0	0	
UMass-SemiMech	Semi-mechanistic	FALSE	4	3	1	
USC-SIkJalpha	Mechanistic	FALSE	4	1	3	
UpgUmibUsi-MultiBayes	Semi-mechanistic	FALSE	3	2	0	
bisop-seirfilter	Mechanistic	FALSE	4	1	3	
bisop-seirfilterlite	Mechanistic	FALSE	4	1	3	
prolix-euclidean	Semi-mechanistic	FALSE	4	3	0	
CovidMetrics-epiBATS	Statistical	TRUE	3	0	0	
EuroCOVIDhub-baseline	Statistical	TRUE	4	0	0	
FIAS_FZJ-Epi1Ger	Mechanistic	TRUE	3	0	3	
GoeWroc-BaseBayes	Semi-mechanistic	TRUE	3	3	0	
HZI-AgeExtendedSEIR	Mechanistic	TRUE	3	0	3	
ICM-agentModel	Agent-based	TRUE	3	0	0	
IEM_Health-CovidProject	Mechanistic	TRUE	4	0	4	
JBUD-HMXK	Mechanistic	TRUE	3	0	3	
LeipzigIMISE-SECIR	Mechanistic	TRUE	3	0	3	
MIMUW-StochSEIR	Mechanistic	TRUE	3	0	3	
MIT_CovidAnalytics-DELPHI	Mechanistic	TRUE	3	0	3	
MOCOS-agent1	Agent-based	TRUE	3	0	0	
MUNI-ARIMA	Statistical	TRUE	3	0	0	
MUNI-LaggedRegARIMA	Statistical	TRUE	3	0	0	
MUNI-VAR	Statistical	TRUE	3	0	0	
MUNI_DMS-SEIAR	Mechanistic	TRUE	3	0	3	
PL_GRedlarski-DistrictsSum	Mechanistic	TRUE	3	0	3	
RobertWalraven-ESG	Statistical	TRUE	3	0	0	
UB-BSLCoV	Statistical	TRUE	3	0	0	
UC3M-EpiGraph	Agent-based	TRUE	3	0	0	
ULZF-SEIRC19SI	Mechanistic	TRUE	3	0	3	
UMass-MechBayes	Mechanistic	TRUE	3	0	3	
UNED-PreCoV2	Statistical	TRUE	3	0	0	
UNIPV-BayesINGARCHX	Statistical	TRUE	3	0	0	
epiMOX-SUIHTER	Mechanistic	TRUE	3	0	3	
epiforecasts-EpiExpert	Judgement	TRUE	3	0	0	
epiforecasts-EpiExpert_Rt	Judgement	TRUE	3	0	0	
epiforecasts-EpiExpert_direct	Judgement	TRUE	3	0	0	
epiforecasts-EpiNow2	Semi-mechanistic	TRUE	3	3	0	
epiforecasts-weeklygrowth	Statistical	TRUE	2	0	0	
itwm-dSEIR	Mechanistic	TRUE	3	0	3	

3 Statistical methods

3.1 Epidemic trend identification

We retrospectively categorised each week as “Stable”, “Decreasing”, or “Increasing”, based on the difference over a three-week moving average of incidence (with a change of +/-5% as “Stable”).

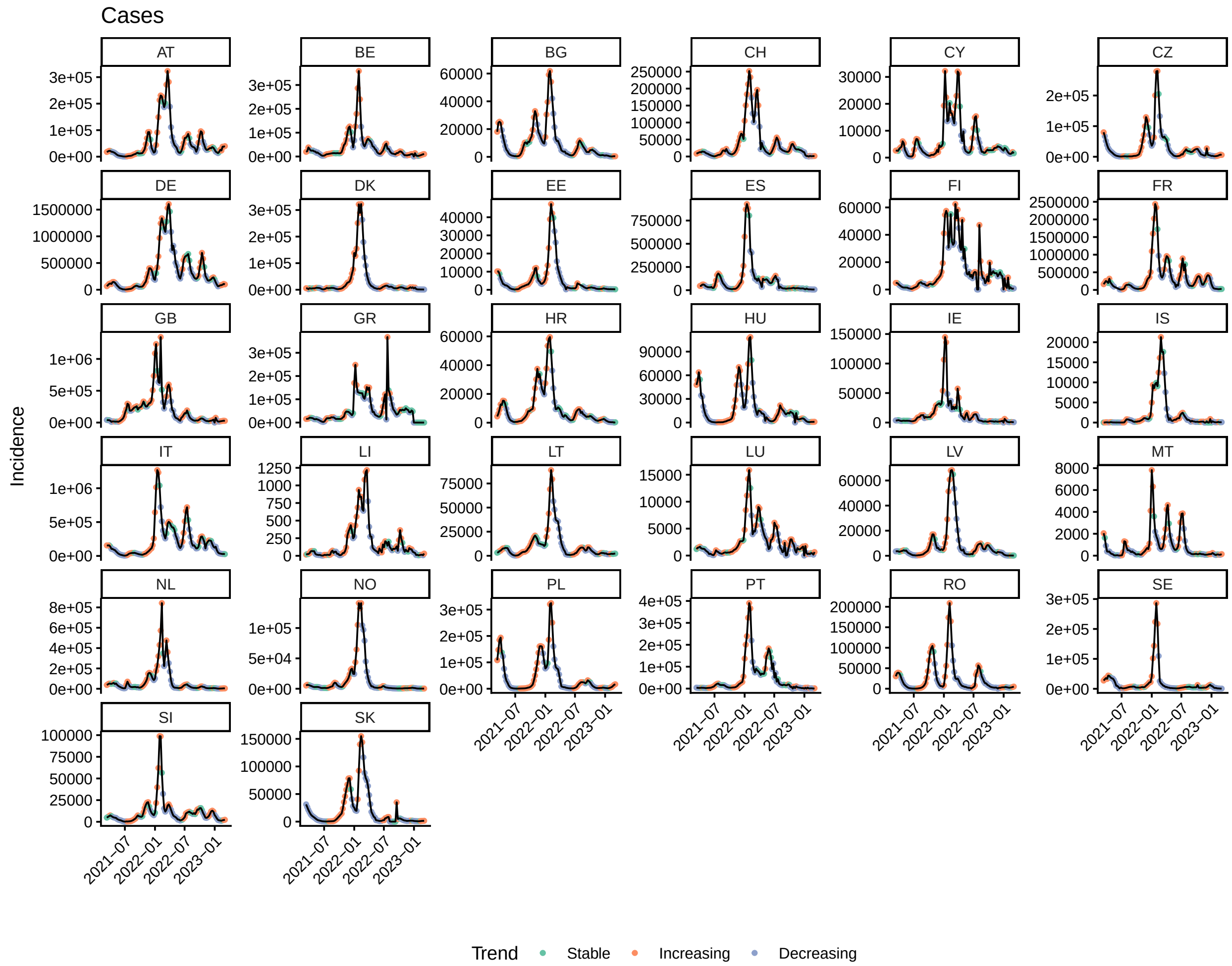


Figure 2: Trends (cases)

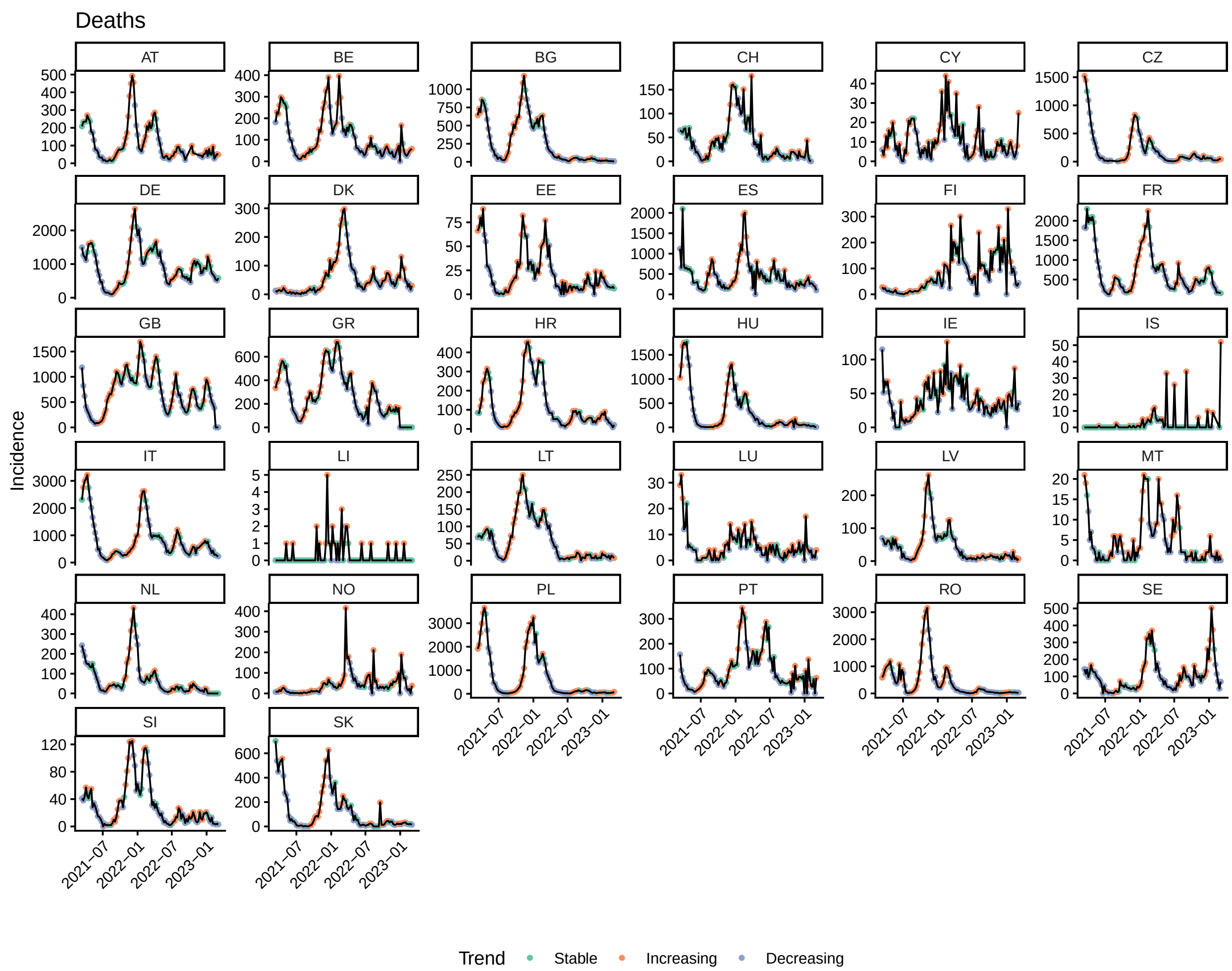


Figure 3: Trends (deaths)

3.2 Variant phase identification

Figure 4: Variant phases identified by dominant variant in each location and week

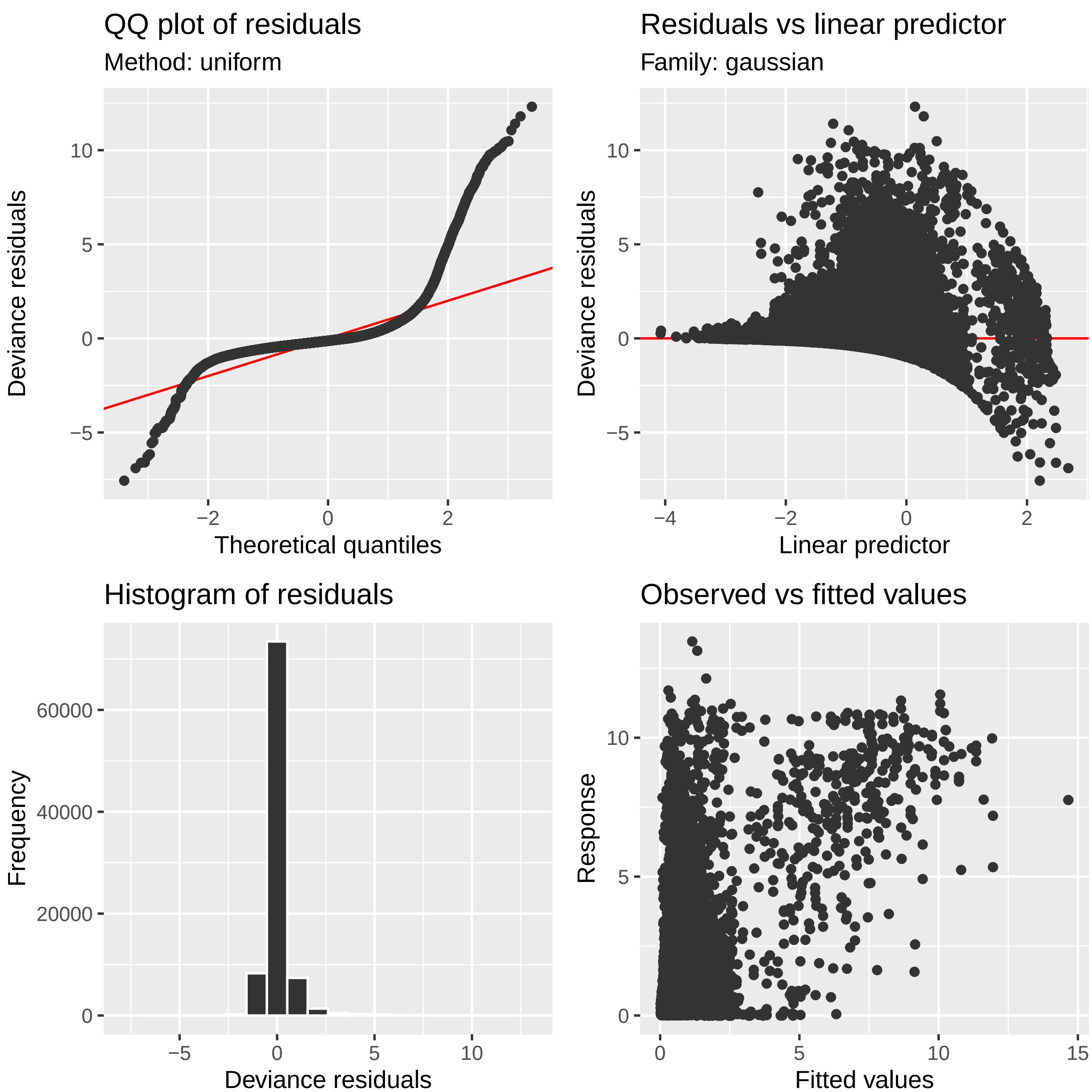
Genomic surveillance data were obtained from three sources: ECDC (covering 30 European countries), UKHSA (Great Britain), and the Swiss Federal Office of Public Health (Switzerland). Variant lineages were mapped to six named phases in expected chronological order: Alpha, Delta, Omicron-BA.1, Omicron-BA.2, Omicron-BA.4/5, and Omicron-BQ/XBB. For each country, we identified the first week in which each named variant exceeded 50% of sequenced samples. We enforced chronological ordering by removing any out-of-sequence phases, then expanded phase assignments to all weeks by filling forward and backward from observed transition dates. This per-location approach accounts for the fact that variant dominance dates differed substantially across European countries. Where genomic surveillance data were too sparse to identify a transition (Hungary), we supplemented with epidemiological reports to set the Alpha-to-Delta transition date.

3.3 Model fitting

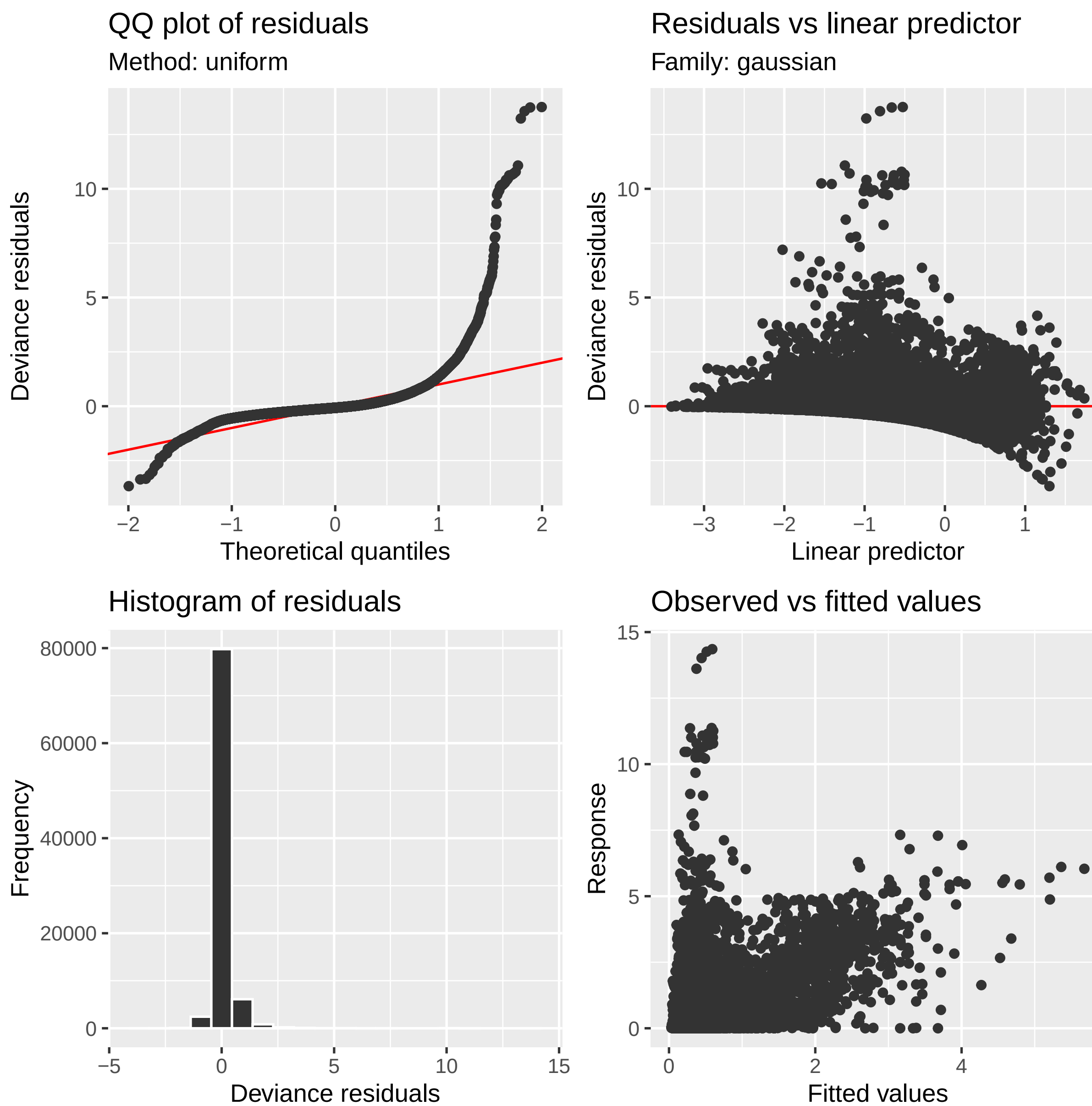
3.3.1 Model formula

~, wis, s(Method, bs = "re") + s(CountryTargets, bs = "re") + s(Incidence) + s(Trend, bs = "re") + s(Location, bs = "re") + s(VariantPhase, bs = "re") + s(Horizon, by = Model, k = 3, bs = "sz") + s(Model, bs = "re")

3.3.2 Model diagnostics



3.3.2.1 Cases



3.3.2.2 Deaths

3.4 Model results

3.4.1 Difference in effects for each modelled variable

3.4.2 Results on the natural scale

We present results from scoring forecast error on the natural scale (difference between observation and prediction in count of case or death incidence), from which the WIS is calculated.

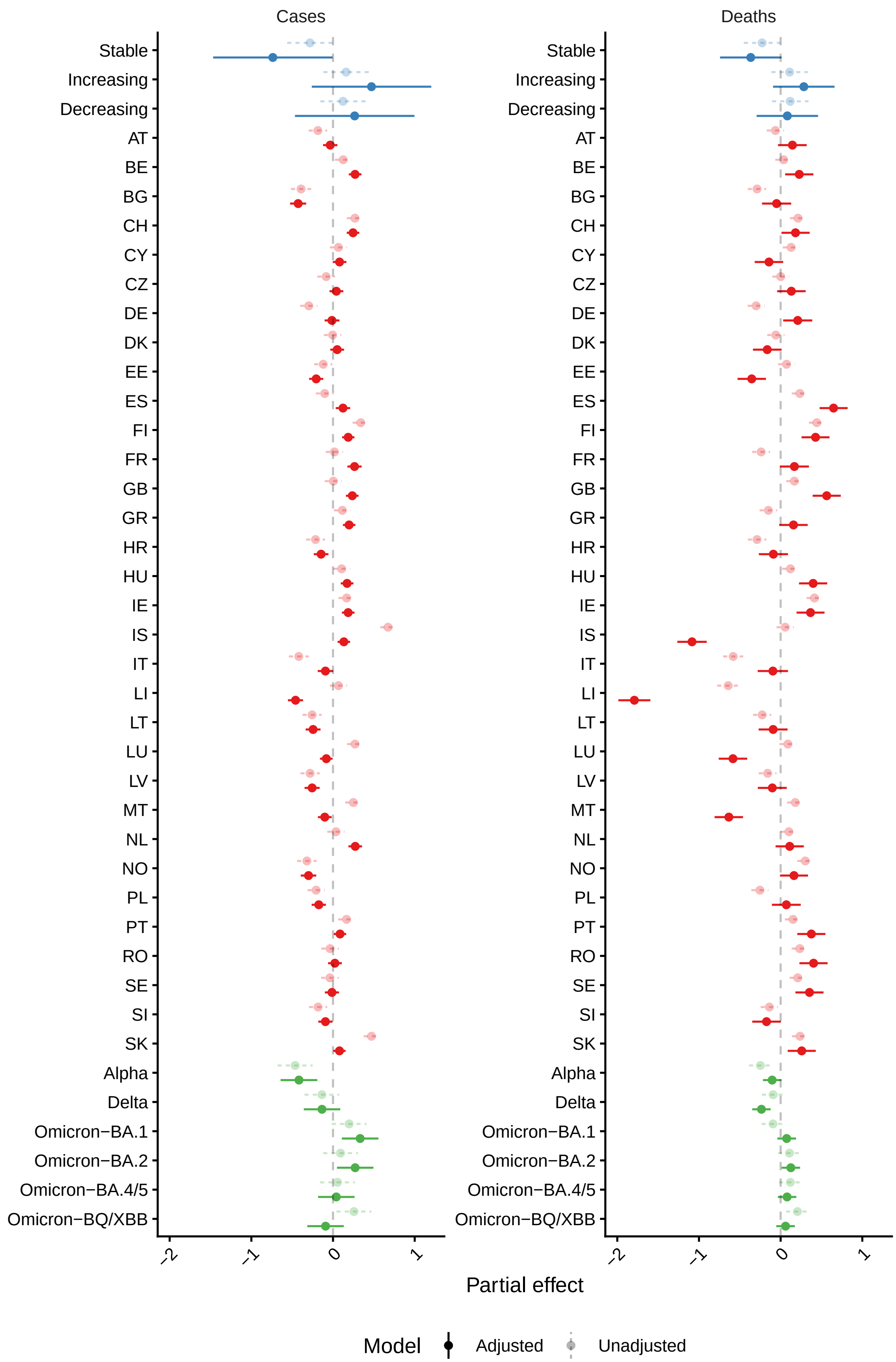


Figure 5: Effect of counfounding factors on forecast performance (WIS), showing epidemic trend, country location, and dominant variant phase. An effect <0 indicates improved forecast performance relative to other forecast targets. Effects of each variable shown from a univariate model (unadjusted), and in a joint (adjusted) model accounting for all variables.

Table 3: Characteristics of forecasts sampled from the European COVID-19 Forecast Hub, March 2021-2023. Forecast performance was measured using the weighted interval score (WIS), with a lower score indicating a more accurate forecast.

	Cases			Deaths		
Variable	Models	Forecasts	Mean WIS (SD)	Models	Forecasts	Mean WIS
Overall	42 (100%)	91,966 (100%)	34611.95 (269907.54)	38 (100%)	89,885 (100%)	2833.23 (1709
Method						
Semi-mechanistic	9 (21.4%)	28,742 (31.3%)	44715.97 (439142.11)	10 (26.3%)	28,494 (31.7%)	8749 (303
Statistical	11 (26.2%)	32,680 (35.5%)	28387.12 (140683.69)	7 (18.4%)	25,478 (28.3%)	61.87 (1
Mechanistic	16 (38.1%)	27,987 (30.4%)	33108.23 (133479.4)	16 (42.1%)	33,449 (37.2%)	84.2 (2
Agent-based	3 (7.1%)	814 (0.9%)	21585.86 (59798.66)	2 (5.3%)	720 (0.8%)	158.05 (
Judgement	3 (7.1%)	1,743 (1.9%)	14971.31 (50352.63)	3 (7.9%)	1,744 (1.9%)	95.18 (1
Number of country targets						
Single-country	19 (45.2%)	3,680 (4%)	57503.7 (176819.05)	14 (36.8%)	2,821 (3.1%)	170.13 (4
Multi-country	23 (54.8%)	88,286 (96%)	33657.7 (273058.96)	24 (63.2%)	87,064 (96.9%)	2919.74 (1737