Evaluating multi-model and ensemble forecasts in the context of the Mpox outbreak in multiple countries, July 28th, 2022, through January 19th, 2023

Amanda Bleichrodt*[1], Ruiyan Luo[1], Alexander Kirpich[1], Gerardo Chowell[1]

[1] Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA.

* Corresponding author

Email: ableichrodt1@student.gsu.edu

Abstract

In May 2022, public health officials noted an unprecedented surge in Mpox cases in non-endemic countries worldwide. As the epidemic accelerated, multi-model forecasts of the epidemic's trajectory were critical in guiding the implementation of public health interventions, and in determining policy. As the case levels have significantly decreased as of mid-August, evaluating model performance is essential to advance the growing field of epidemic forecasting. Using laboratory-confirmed Mpox case data from the CDC and OWID teams through the week of January 26th, 2023, we generated retrospective sequential weekly forecasts (e.g., 1-week, 2-week, 3-week, and 4-week) for Brazil, Canada, England, France, Germany, Spain, the USA, and at the global scale using models that require minimal input data including the auto-regressive integrated moving average (ARIMA), general additive model (GAM), simple linear regression (SLR) as well as the spatial-wave and n-sub-epidemic modeling frameworks. Forecast performance was assessed using MSE, MAE, WIS, 95% PI coverage metrics as well as skill scores. Overall, the spatial-wave sub-epidemic modeling framework outcompeted other models across most locations and forecasting horizons in terms of the average MSE, MAE, and WIS. It was followed closely in success by the n-sub-epidemic top-ranked and weighted ensemble (2) models. Regarding average 95% PI coverage, the n-sub-epidemic unweighted (equally weighted top-models) ensemble (2) model performed best across all forecasting horizons for most locations. However, many locations had multiple models performing equally well in terms of the average 95% PI coverage. Both the n-sub-epidemic and spatial-wave frameworks improved considerably in average MSE, MAE, and WIS, and minimally (<10%) in average 95% PI coverage relative to the ARIMA model. Findings lend further support to sub-epidemic models for short-term forecasting epidemics of emerging and re-emerging infectious diseases.

Summary

In the face of many unknowns (i.e., transmission, symptomology) posed by the unprecedented 22-23 Mpox epidemic, near real-time short-term forecasts of the epidemic's trajectory were essential in intervention implementation and guiding policy. As case levels continue to dissipate (as of mid-August), evaluating the modeling strategies used in producing forecasts is critical to refine and grow the field of epidemiological forecasting. Here, we systematically evaluate the performance of an ensemble n-sub-epidemic and related spatial-wave modeling frameworks against ARIMA, GAM, and SLR models in producing sequential retrospective weekly (1-4 week) forecasts of Mpox cases for the highest burdened countries (i.e., Brazil, Canada, England, France, Germany, Spain, the United States) and on a global scale. Overall, the spatial-wave framework outperformed all other models most frequently, followed closely in success by the ensemble n-sub-epidemic framework for the average MSE, MAE, and WIS metrics. However, the n-sub-epidemic unweighted ensemble (2) model produced the highest average 95% PI coverage for most locations and forecasting horizons. The considerable success seen with both frameworks highlights the continued utility of sub-epidemic methodologies in producing short-term forecasts and their potential application to other epidemiologically different diseases.

Introduction

In May 2022, an unprecedented surge in Mpox cases affecting multiple countries previously free of the disease began worldwide [1,2]. At the time, little was known regarding key transmission and control factors – vaccination, the role of asymptomatic individuals, and

transmission pathways – which shaped the overall trajectory of the epidemic [2,3]. Fortunately, as of mid-August cases have steadily declined worldwide [1,4,5-6], with non-endemic countries reporting a total of 85,738 cases and 112 deaths as of May 9, 2023 [6]. Most cases have occurred via close sexual contact, mainly in the gay, bisexual, and men who have sex with men (MSM) communities [2,3,7–9]. When an emerging pathogen takes off in a population, short-term forecasts of the epidemic's trajectory at different spatial scales help guide policy and the implementation of intervention strategies [10–14].

As cases increased in late July 2022, we started to produce weekly short-term forecasts of new cases for the highest-affected countries (e.g., Brazil, Canada, England, France, Germany, Spain, and the United States) and at the global level in real-time. For this purpose we primarily used a sub-epidemic modeling framework, which has performed competitively in short-term forecasts of emerging infectious diseases [10,11,13,14]. Over the past few months, each country saw variable epidemic trajectories, with the United States (OWID) and USA (CDC) seeing the highest burden of cases (Fig 1) [1,15]. Brazil experienced the most prolonged peak (8-weeks), and Canada, England, and France saw multiple subsequent smaller peaks during the decline phase (Fig 1) [1]. Figure 1 illustrates the trajectories in each of the study areas. Given the heterogeneous impact of the epidemic at different spatial scales, a retrospective evaluation of multi-model forecasts is key for better understanding model performance in different contexts, which can contribute to advancing the science of epidemic forecasting.

Fig 1. **The epidemiological trajectories of new weekly cases for each location (April 28th, 2022-January 26th, 2023).** We obtained case information for Canada, England, Germany, France, Spain, Brazil, the United States, and the World from the OWID GitHub page [1],

whereas the case counts for the USA-CDC came from the CDC webpage [15]. The solid black line shows the reported cases each week as of January 26th, 2023. The red-dashed lines indicate the forecasting periods used as part of the analysis.

Several methodologies have been employed to forecast the case trajectory of the 2022 Mpox epidemic in various geographical regions [10,16–22], including models focused on human judgment [22], deep learning and artificial intelligence models (AI) [16,19,20], machine learning models [16-18,20,21], statistical models such as auto-regressive integrated moving average (ARIMA) models [16–18,20,21], and semi-mechanistic sub-epidemic models [10]. Of the implemented strategies, complex deep learning, artificial intelligence (AI), and machine learning models were most commonly employed and evaluated [16–21]. However, performance metrics employed have mainly included variations of MAE and MSE [16–21], with only one study also including probabilistic measures of performance such as the weighted interval score [10]. Of the studies, three studies only evaluated forecasts of the ascending phase of the epidemic (May-Early August) [17,18,22], one focused on the descending phase (Mid-August - Late September) [21], and four studies analyzed the ascending, peak, and decline phases of the epidemic (May-Present) [10,16,19,20].

In this paper, we carried out a comprehensive retrospective assessment of forecasting performance (1-4 weeks ahead) of sub-epidemic models compared against statistical models, including ARIMA, General Additive Models (GAM), and simple linear regression (SLR) for the highest impacted countries throughout the Mpox epidemic (weeks of July 28th, 2022 through January 19th, 2023). We took into account a comprehensive set of performance metrics in the

field of epidemic forecasting, namely the mean absolute error (MAE), mean square error (MSE), 95% prediction interval coverage (95% PI), and weighted interval score (WIS).

Results

Overall forecast performance

Overall, the spatial-wave and ensemble n-sub-epidemic frameworks outperformed the established statistical models (e.g., ARIMA, SLR, and GAM), except for in Brazil, were the ARIMA model saw the most success. Regarding specific study areas, the spatial-wave framework performed best in Canada, England, France, the United States, and the World. In contrast, the n-sub-epidemic model outperformed other models in Germany, Spain, and the USA (CDC). However, the ensemble n-sub-epidemic unweighted ensemble (2) model saw the highest 95% PI coverage in every study area, albeit in France, compared to all other models included in the analysis. Supplementary Figs 1 through 36 visualize the metric values for each location and horizon, along with the best-performing and following best-performing models (S1 Fig). Tables A through R include average forecasting performance metrics for each study area, horizon, and model (S1 Table). Table A in S2 Table contains individual forecast performance metrics for each study area, forecasting horizon, model, and forecasting period (S2 Table).

Country-level performance

Brazil

Unlike the other study areas, the ARIMA model outperformed other models 68.8% of the time, followed by the ensemble n-sub-epidemic framework (25.0%), spatial-wave framework

(6.3%), and SLR model (6.3%) (Table A in S1 Table). The ARIMA model produced the lowest average MSE (Range: 28659.5-104554.4) and MAE (Range: 123.1-183.7) in most forecasting horizons and across all forecasting horizons for average WIS (Range: 77.5-125.5) (Figs 2-4; Table A in S1 Table). Similar to most study areas, the unweighted ensemble (2) n-sub-epidemic model produced the highest 95% PI coverage at all forecasting horizons (Range: 92.3-94.6%) (Fig 5; Table A in S1 Table), albeit improving minimally over the ARIMA model (<10%)(Fig 6; Table B in S1 Table). The ARIMA model performed equally well at the 1-week horizon regarding average 95% PI coverage (Fig 5; Table A in S1 Table).

Fig 2. Average Mean Square Errors (MSE) for each location, model, and forecasting horizon (weeks of July 28th, 2022-January 19th, 2023). A total of 882 forecasts were produced for each model. The graph is shown on the log10 scale. The red bars indicate the best-performing model at a given location and horizon, and multiple red bars indicate ties between models. Overall, the spatial-wave framework performed best most frequently across forecasting horizons and locations regarding average MSE.

Fig 3. Average Mean Absolute Errors (MAE) for each location, model, and forecasting horizon (weeks of July 28th, 2022-January 19th, 2023). A total of 882 forecasts were produced for each model. The graph is shown on the log10 scale. The red bars indicate the best-performing model at a given location and horizon, and multiple red bars indicate ties between models. Overall, the spatial-wave framework performed best most frequently across forecasting horizons and locations regarding average MAE.

Fig 4. Average Weighted Interval Scores (WIS) for each location, model, and forecasting horizon (weeks of July 28th, 2022-January 19th, 2023). A total of 882 forecasts were

produced for each model. The graph is shown on the log10 scale. The red bars indicate the best-performing model at a given location and horizon, and multiple red bars indicate ties between models. Overall, the spatial-wave framework performed best most frequently across forecasting horizons and locations regarding average WIS.

Fig 5. Average 95% Prediction Interval Coverage (95% PI) for each location, model, and forecasting horizon (weeks of July 28th, 2022-January 19th, 2023). A total of 882 forecasts were produced for each model. The graph is shown on the log10 scale. The red bars indicate the best-performing model at a given location and horizon, and multiple red bars indicate ties between models. Overall, the unweighted ensemble (2) n-sub-epidemic model performed best most frequently across forecasting horizons and locations regarding average 95% PI coverage.

Fig 6. The skill scores for the 95% Prediction Interval Coverage (95% PI) comparing the ARIMA model to the models included as part of the n-sub-epidemic and spatial-wave frameworks for each location and horizon. Negative skill scores indicate that the comparison models improved over the ARIMA model.

Canada

The spatial-wave framework outperformed the other models 62.5% of the time, followed by the n-sub-epidemic framework (31.3%) and SLR model (6.3%) (Table C in S1 Table). The second-ranked and top-ranked spatial-wave models produced the lowest average MSE (Range: 508.7-820.4) and MAE (Range: 10.2-12.9), respectively, both seeing over a 30% improvement compared to the ARIMA model (Figs 2-3, 7-8; Tables C and D S1 Table). The second-ranked spatial-wave model also performed best in all forecasting horizons regarding average WIS (Range: 7.9-9.4), seeing a 25.0-39.7% improvement over the ARIMA model. (Figs 4 and 9;

Tables C and D in S1 Table). Similar to many other locations, the second-ranked and unweighted ensemble (2) n-sub-epidemic models produced the highest (100%) coverage across all forecasting horizons; however, it improved minimally over the ARIMA model (<10%) (Fig 5 and 6; Tables C and D in S1 Table).

Fig 7. The skill scores for the Mean Square Error (MSE) comparing the ARIMA model to the models included as part of the n-sub-epidemic and spatial-wave frameworks for each location and horizon. Positive skill scores indicate that the comparison models improved over the ARIMA model.

Fig 8. The skill scores for the Mean Absolute Error (MAE) comparing the ARIMA model to the models included as part of the n-sub-epidemic and spatial-wave frameworks for each location and horizon. Positive skill scores indicate that the comparison models improved over the ARIMA model.

Fig 9. The skill scores for the Weighted Interval Score (WIS) comparing the ARIMA model to the models included as part of the n-sub-epidemic and spatial-wave frameworks for each location and horizon. Positive skill scores indicate that the comparison models improved over the ARIMA model.

England

The spatial-wave framework outperformed the other models 75.0% of the time, followed by the n-sub-epidemic framework (25.0%) and ARIMA model (6.3%) (Table E in S1 Table). Both ensemble (2) spatial-wave models produced the lowest average MSE and MAE values across forecasting horizons and improved over 24% on the ARIMA model (Figs 2-3, 7-8; Tables

E and F in S1 Table). The spatial-wave framework continued to outperform other models regarding average WIS but with more success seen across the top-ranked (Range: 14.8-17.9) and second-ranked (Range: 14.6-17.6) models (Fig 4; Table E in S1 Table). Similar to most other study areas, the second-ranked and unweighted ensemble (2) n-sub-epidemic models consistently produced the highest 95% PI coverage across all forecasting horizons (Range: 92.3-96.7%) (Fig 5; Table E in S1 Table). However, both models improved minimally over the ARIMA model (<10%) (Fig 6; Table F in S1 Table).

France

The spatial-wave framework outperformed the other models 81.3% of the time, followed by the n-sub-epidemic framework (43.8%) and the ARIMA model (25.0%) (Table G in S1 Table). The spatial-wave framework produced the lowest average MSE and MAE values for most forecasting horizons, with the most success noted for the weighted ensemble (2) model (Figs 2-3; Table G in S1 Table). The model improved 12.5-37.8% in average MSE and 30.5-45.1% in average MAE over the ARIMA model (Figs 7-8; Table H in S1 Table). The weighted ensemble (2) n-sub-epidemic produced the lowest average WIS most frequently (Range: 38.9-40.0), improving between 14.1-34.3% over the ARIMA model (Figs 4 and 9; Tables G and H in S1 Table). Unlike most other study areas, the top-ranked n-sub-epidemic model, all spatial-wave models, and the ARIMA model produced the highest 95% PI coverage across all forecast horizons (Fig 5; Table G in S1 Table).

Germany

The n-sub-epidemic framework outperformed the other models 62.5% of the time, followed by the spatial-wave framework (50.0%) (Table I in S1 Table). The top-ranked and weighted ensemble (2) n-sub-epidemic models produced the lowest average MSE most frequently, improving between 80.0-94.7% over the ARIMA model (Figs 2 and 7; Tables I and J in S1 Table). However, the second-ranked spatial-wave model consistently produced the lowest average MAE across all horizons (Range: 9.3-11.3) and improved over the ARIMA model by 59.2-75.7% (Figs 3 and 8; Tables I and J in S1 Table). The weighted ensemble (2) n-sub-epidemic model performed best most frequently regarding average WIS (Range: 6.9-8.3), improving 52.6-71.1% over the ARIMA model. (Figs 4 and 9; Tables I and J in S1 Table). Similar to most study areas, the unweighted ensemble (2) n-sub-epidemic model consistently saw the highest 95% PI coverage across all forecasting horizons (Range: 91.7-94.6%), albeit improving minimally over the ARIMA model (<10%) (Figs 5 and 6; Tables I and J in S1 Table).

Spain

The n-sub-epidemic framework outperformed the other models 81.3% of the time, followed by the spatial-wave framework (25.0%) and the ARIMA model (6.3%) (Table K in S1 Table). The unweighted ensemble (2) n-sub-epidemic model produced the lowest MSE (Range: 2758.7-3226.3) and MAE (Range: 32.8-34.8) across all forecasting horizons and improved significantly over the ARIMA model for both metrics (Figs 2-3, 7-8; Tables K and L in S1 Table). The second-ranked spatial-wave model had the lowest average WIS at the 2- through 4-week forecasting horizons (Range: 25.0-26.5), improving over the ARIMA model 43.2-61.1% (Figs 4 and 9; Tables K and L in S1 Table). Similar to most other study areas, the second-ranked (Ranked: 92.3-95.8%) and unweighted ensemble (2) (Range: 92.3-95.7%) n-sub-

epidemic models produced the best 95% PI coverage across most forecasting horizons, improving minimally on the ARIMA model (<10%)(Figs 5 and 6; Tables K and L in S1 Table).

United States (OWID)

The spatial-wave framework outperformed the other models 62.5% of the time, followed by the n-sub-epidemic framework (37.5%) (Table M in S1 Table). The unweighted ensemble (2) spatial-wave model produced the lowest average MSE most frequently (Range: 124619.8-193440.6), improving over the ARIMA model between 58.6-84.4% (Figs 2 and 7; Tables M and N in S1 Table). The spatial-wave framework continued to succeed in MAE and WIS, with the second-ranked and weighted ensemble (2) models performing best in average MAE and WIS, respectively (Figs 3-4; Table M in S1 Table). Here, the second-ranked model improved on the ARIMA model between 36.4-60.5% in average MAE, and the weighted ensemble improved in average WIS 35.9-61.4% (Figs 8-9; Table N in S1 Table). Like most other study areas, the unweighted ensemble (2) n-sub-epidemic model produced the highest 95% PI coverage across all forecasting horizons (Range: 95.7-96.2%), outperforming the ARIMA model between 8.7-17.3% (Figs 5 and 6; Tables M and N in S1 Table).

USA (CDC)

The n-sub-epidemic framework outperformed the other models 56.3% of the time, followed by the spatial wave framework (43.8%) (Table O in S1 Table). The top-ranked and weighted ensemble (2) n-sub-epidemic models outperformed other models most frequently regarding average MSE and MAE (Figs 2-3; Table O in S1 Table). Both top-performing models saw considerable improvements over the ARIMA model at all horizons (Figs 7-8; Table P in S1

Table). The top-ranked spatial-wave model produced the lowest WIS at the 2- through 4-week horizons (Range: 35.5-51.0), improving between 75.0-79.9% over the ARIMA model (Figs 4 and 9; Tables O and P in S1 Table). Similar to most study areas, the unweighted ensemble (2) n-sub-epidemic model saw the highest 95% PI coverage most frequently (Range: 96.2-98.9%) and improved 8.7-26.4% over the ARIMA model (Figs 5 and 6; Tables O and P in S1 Table).

World

The spatial-wave framework outperformed the other models 81.3% of the time, followed by the n-sub-epidemic framework (25.0%) (Table Q in S1 Table). Regarding average MSE and MAE, the weighted and unweighted ensemble (2) spatial-wave models performed best in the 1- and 4-week and 2-3 weeks horizons, respectively (Figs 2-3; Table Q in S1 Table). However, both models improved over the ARIMA model between 49-56% regarding average MSE and 30-48% for average MAE (Figs 7 and 9; Table R in S1 Table). The top-and second-ranked spatial-wave models produced the lowest average WIS compared to other models, improving between 30-46% over the ARIMA model (Fig 4 and 9; Tables Q and R in S1 Table). Similar to other study areas, albeit France, the unweighted ensemble (2) n-sub-epidemic model produced the highest 95% PI coverage most frequently across forecasting horizons (Range: 88.5-90.3%) and saw between a 9.5-20.6% improvement over the ARIMA model (Fig 5 and 6; Tables Q and R in S1 Table).

Discussion

In the context of the global emergency posed by Mpox in 2022-23, we have systematically investigated short-term forecasting performance of multiple models that have

shown competitive performance during epidemics and pandemics and rely on minimal data of the epidemic's trajectory [10,13,14]. Overall, the spatial-wave framework, followed closely by the ensemble n-sub-epidemic framework, performed best most frequently across study areas and forecasting horizons compared to established modeling techniques (e.g., ARIMA, SLR, GAM). The spatial-wave framework outperformed other models in Canada, England, France, the United States, and the World. In contrast, the ensemble n-sub-epidemic framework performed best in Germany, Spain, and USA (CDC). Brazil, however, behaved uniquely compared to the other countries, with the ARIMA model performing best. Nevertheless, the unweighted ensemble (2) n-sub-epidemic model continuously outperformed other included models regarding average 95% PI coverage in every country, albeit France, where the top-ranked ensemble n-sub-epidemic, all spatial-wave models, and the ARIMA model performed best. Compared to the ARIMA model, the spatial-wave and ensemble n-sub-epidemic models frequently improved considerably in each metric, except for 95% PI coverage, where minimal improvement (<10%) over the ARIMA was noted for most countries. Among the models investigated in our study, our findings further support the sub-epidemics frameworks followed by the ARIMA model for forecasting emerging infectious diseases, especially in the face of limited epidemiological data [10,13,14].

Each included study area experienced unique epidemic trajectories throughout the Mpox epidemic, which may explain differences in model performance. In fitting the model, both the ensemble n-sub-epidemic and spatial-wave frameworks make assumptions about the timing of the aggregated sub-epidemics. For example, the ensemble n-sub-epidemic framework allows for asynchronous sub-epidemics, thereby capturing more complicated epidemic trajectories [10,13]. However, the spatial-wave framework assumes the presence of sequential sub-epidemics whose sizes follow a declining pattern dictated by an exponential or power-law function, as described in

the methods [14].

When looking at the epidemic trajectories for Canada, England, France, the United States, and the World, they have multiple outbreak peaks throughout the declining phase of the epidemic [1]. As the spatial-wave framework specifically accounts for multi-peak epidemics, the epidemic trajectory within those areas may explain the success of the spatial-wave framework in producing short-term forecasts [14]. Germany, Spain, and the USA (CDC) exhibited uni-model peaks without the subsequent peaks noted in the decline phase of the epidemic [1,15]. Here, the ensemble n-sub-epidemic framework performed best overall, possibly because of the model's flexibility in estimating the underlying sub-epidemics within the single peak [10,13]. Overall, while both the spatial-wave and ensemble n-sub-epidemic frameworks did well, the differences in success in models between locations may be due to the underlying assumptions made by the frameworks about the trajectory of the epidemic in each study area.

Of all the included study areas, Brazil was the only one where the ARIMA model performed best overall. Previous studies have confirmed the effectiveness of ARIMA models in producing short-term Mpox forecasts [20]. However, in the other study areas, the ensemble n-sub-epidemic and spatial-wave frameworks outperformed the ARIMA model. Compared to the other included areas, Brazil saw a similar ascension phase of the outbreak. However, unlike any other area, Brazil experienced a prolonged peak in cases (8-weeks) which may be explained by multiple structural factors that complicated the management of the outbreak [23]. The peak itself appeared linear. ARIMA models are a subset of linear models [24], whereas the ensemble n-sub-epidemic and spatial-wave models are not [10,13,14]. Therefore, the extended linear trend noted in Brazil’s epidemic trajectory may explain the significant success seen with the ARIMA model.

The significant success of the unweighted ensemble (2) n-sub-epidemic model in 95% prediction interval coverage also highlights the utility of ensemble modeling in short-term forecasting [10-11,13, 25-35] . Ensemble modeling has also shown considerable success against individual top-ranking sub-epidemic and other statistical models in past Mpox short-term forecasting efforts [10,11], along with that of COVID-19 [13,30,35], seasonal influenza[28,33], and Zika [34]. Each ensemble (2) model (e.g., weighted and unweighted) included within the analysis performed competitively at various forecasting horizons and for different areas. However, the unweighted ensemble (2) n-sub-epidemic outperformed all other included models for each area, albeit France, in average 95% PI coverage, demonstrating the continued competitiveness of ensemble modeling in short-term Mpox forecasting.

Our study is not exempt from limitations. We used two sources for weekly time-series laboratory-confirmed Mpox cases, each using varying approaches to compile and disseminate case information. For example, the CDC defines case *date* as the earliest date regarding a positive case (e.g., positive lab-test reporting date, symptom onset date, diagnosis date, date reported to CDC Call Center) [15]. The OWID data defines case *date* as either the date or symptom onset, followed by the date of diagnosis and then the date of reporting depending on the availability of information [36]. Due to the un-precedented nature of the epidemic and limited epidemic data available, we examined forecasts derived from the second-ranked spatial wave and n-sub-epidemic models. Frequently, there was little statistical support for the second-ranked models relative to the top-ranked model for both frameworks. The spatial-wave and ensemble n-sub-epidemic frameworks are semi-mechanistic and provide insight into the natural processes that formed the observed epidemic trends from the aggregated sub-epidemics. Nevertheless, both frameworks do not account for explicit mechanisms of reactive behavioral

modification and interventions, which likely played a significant role in controlling the epidemic at local and global levels [9, 37–39]. Finally, current models are not sensitive to seasonal or event-specific risk behaviors, which limit their application to short-term forecasts.

In conclusion, our findings further support the competitive performance of the aggregated sub-epidemic methodologies in producing short-term forecasts within the context of Mpox, along with the utility of ensemble modeling therein producing short-term forecasts. However, the frameworks could be expanded to other disease types as they have shown utility in short-term forecasting for epidemiologically different diseases [10,13-14]. Finally, short-term forecasting tools are essential in the face of unprecedented pandemics and epidemics. Therefore, further evaluation and refinements of the spatial-wave and ensemble n-sub-epidemic frameworks could be achieved via comparison to other models since all the forecasting results are publicly available to the community for transparency and further investigation.

Materials and Methods

Data

Our team utilized weekly confirmed Mpox cases through the week of January 26th, 2023, from the CDC and Our World in Data (OWID) publicly available data sources [1,15] to produce and evaluate sequential retrospective weekly forecasts for the weeks of July 28th, 2022, through the week of January 19th, 2023. We retrieved data for the United States from the CDC on February 1st, 2023 [15]. Data on the global level and for the countries that saw the highest levels of cases throughout the outbreak (e.g., Brazil, Canada, England, France, Germany, Spain, and the

United States) were downloaded from the OWID GitHub repository on February 3rd, 2023 [1]. We utilized the United States Mpox case data from *both* the OWID and CDC sources as they collected and reported case information using different methodologies [1,15,40-41]. Here, both the CDC and OWID teams define a *confirmed case* as a laboratory verified Mpox infection [40,41].

Forecasting strategy

For each model, we conducted a total of 882 retrospective weekly sequential 1- to 4-week ahead forecasts (weeks of July 28th, 2022, through the week of January 19th, 2023) using 10-week calibration periods. For the week of July 28th, 2022, only 9-weeks of data were used to calibrate the n-sub-epidemic and spatial-wave models for the United States (OWID) and Canada, and 8-weeks for Brazil. In the following week (August 4th, 2022), only 9-weeks of data were used in calibrating the models for Brazil. The shortened calibration periods are due to initial zeros seen at the start of earlier forecasting periods. Descriptions of the n-sub-epidemic, sub-epidemic wave, Auto-regressive Integrated Moving Average (ARIMA), and Generalized Additive Models (GAM) methodologies can be found below. We also conducted forecasts using a simple linear regression model (SLR), with *time* being the only predictor. Data posted the week of January 26th, 2023, was used to produce all forecasts.

Model Evaluation strategy

Forecast performance was evaluated for the retrospectively generated 1-4 weeks ahead forecasts (through the week of January 19th, 2023) utilizing data downloaded the week of January 26th, 2023. Descriptions of the mean absolute error (MAE), mean square error (MSE),

95% prediction interval coverage (95% PI) and weighted interval score (WIS) metrics used in the evaluation can be found below. We also calculated skill scores to compare the proportion of improvement over the established ARIMA modeling framework (baseline model) for average MSE, MAE, WIS, and 95% PI coverage, using the n-sub-epidemic and spatial-wave frameworks as comparison. Here, *average* refers to the average metric over all forecasting periods, for each model, location, and forecasting horizon. The formula is as follows [42]:

$$\frac{Baseline\ Model\ -\ Comparison\ Model}{Baseline\ Model} * 100 \tag{1}$$

We use the ARIMA model as the baseline model, and compare its average metric to that for one of the models included as part of the n-sub-epidemic and spatial-wave frameworks.

The n-sub-epidemic modeling framework

In this n-sub-epidemic modeling framework, epidemic trajectories are modeled as the aggregation of overlapping and asynchronous sub-epidemics. A sub-epidemic follows the 3-parameter generalized-logistic growth model (GLM), which has displayed competitive performance [43–45]. This model is given by the following differential equation:

$$\frac{dC(t)}{dt} = C'(t) = rC^p(t)\left(1 - \frac{C(t)}{K_0}\right), \tag{2}$$

where $C(t)$ denotes the cumulative number of cases at time t and $\frac{dC(t)}{dt}$ describes the incidence curve over time t. The parameter r is positive, denoting the growth rate per unit of time, K_0 is the final outbreak size, and $p \in [0,1]$ is the "scaling of growth" parameter which allows the model to capture early sub-exponential and exponential growth patterns. If $p = 0$, this equation describes a constant number of new cases over time, while $p = 1$ indicates that the early growth phase is exponential. Intermediate values of p $(0 < p < 1)$ describe early sub-exponential (e.g., polynomial) growth dynamics.

An n-sub-epidemic trajectory comprises n overlapping sub-epidemics and is given by the following system of coupled differential equations:

$$\frac{dC_i(t)}{dt} = C_i{}'(t) = A_i(t) r_i C_i{}^{p_i}(t)\left(1 - \frac{C_i(t)}{K_{0_i}}\right), \tag{3}$$

Where $C_i(t)$ tracks the cumulative number of cases for sub-epidemic i, and the parameters that characterize the shape of the i_{th} sub-epidemic are given by (r_i, p_i, K_{0_i}), for $i = 1, \dots, n$. Thus, the 1-sub-epidemic model is equivalent to the generalized growth model described above (2). When $n > 1$, we model the onset timing of the $(i+1)_{th}$ sub-epidemic, where $(i+1) \leq n$, by employing an indicator variable given by $A_i(t)$ so that the $(i+1)_{th}$ sub-epidemic is triggered when the cumulative curve of the i_{th} sub-epidemic exceeds C_{thr}. The $(i+1)_{th}$ sub-epidemic is only triggered when $C_{thr} \leq K_{0_i}$. Then, we have:

$$A_i(t) = \begin{cases} 1, & C_{i-1}(t) > C_{thr} \\ 0, & Otherwise \end{cases} \qquad for\ i = 2, \dots, n$$

(4)

where $A_1(t)$ =1 for the first sub-epidemic. Hence, the total number of parameters needed to model an n-sub-epidemic trajectory is given by $3n + 1$. The initial number of cases is given by $C_1(0) = I_0$, where I_0 is the initial number of cases in the observed data. The cumulative curve of the n-sub-epidemic trajectory is given by:

$$C_{tot}(t) = \sum_{i=1}^{n} C_i(t).$$

(5)

Hence, this modeling framework is suitable for diverse epidemic patterns including those characterized by multiple peaks.

Parameter estimation for the n-sub-epidemic model

Denote the time series of new incident Mpox cases as

$$y_{t_j} = y_{t_1}, y_{t_2}, \dots, y_{t_{n_d}} \text{ where } j = 1,2, \dots, n_d$$

(6)

Here, t_j are the time points for the time series data, n_d is the number of observations. Using these case series, we estimate a total of $3n + 1$ model parameters, namely $\Theta = \left(C_{thr}, r_1, p_1, K_{0_1}, \dots, r_n, p_n, K_{0_n}\right)$. Let $f(t, \Theta)$ denote the expected curve of new Mpox cases of the epidemic's trajectory. We can estimate model parameters by fitting the model solution to the observed data via nonlinear least squares [46] or via maximum likelihood estimation

assuming a specific error structure [47]. For nonlinear least squares, this is achieved by searching for the set of parameters $\hat{\Theta}$ that minimizes the sum of squared differences between the observed data $y_{t_j=}y_{t_1,}y_{t_2} \ldots\ldots y_{t_{n_d}}$ and the model mean $f(t,\Theta)$. That is, $\Theta = \left(C_{thr}, r_1, p_1, K_{0_1}, \ldots, r_n, p_n, K_{0_n}\right)$ is estimated by $\hat{\Theta} = arg\ min \ \ \sum_{j=1}^{n_d}(f\left(t_j,\Theta\right) - y_{t_j})^2$.

We quantify parameter uncertainty using a bootstrapping approach described in ref [48] which allows the computation of standard errors and related statistics in the absence of closed-form solutions. To that end, we use the best-fit model $f(t,\hat{\Theta})$ to generate *B*-times replicated simulated datasets of size n_d, where the observation at time t_j is sampled from a normal distribution with mean $f(t_j,\hat{\Theta})$ and variance $\frac{\sum_{j=1}^{n_d}(f(t_j,\hat{\Theta})-y_{t_j})^2}{n_d-(3n+1)}$. Then, we refit the model to each *B* simulated dataset to re-estimate each parameter. The new parameter estimates for each realization are denoted by $\hat{\Theta}_b$ where $b = 1,2,\ldots,B$. Using the sets of re-estimated parameters $\left(\hat{\Theta}_b\right)$, the empirical distribution of each estimate can be characterized, and the resulting uncertainty around the model fit can similarly be obtained from $f\left(t,\hat{\Theta}_1\right), f\left(t,\hat{\Theta}_2\right), \ldots, f(t,\hat{\Theta}_B)$. We run the calibrated model forward in time to generate short-term forecasts with quantified uncertainty.

Selecting the top-ranked sub-epidemic models

We used the AIC_c values of the set of best fit models based on one and two sub-epidemics to select the top-ranked sub-epidemic models. We ranked the models from best to worst according to their AIC_c values, which is given by [49,50]:

$$AIC_c = n_d log(SSE) + 2m + \frac{2m(m+1)}{n_d - m - 1} \tag{7}$$

where $SSE = \sum_{j=1}^{n_d}(f(t_j, \hat{\Theta}) - y_{t_j})^2$, $m = 3n + 1$ is the number of model parameters, and n_d is the number of data points. Parameters from the above formula for AIC_c are estimated from the nonlinear least-squares fit, which implicitly assumes normal distribution for error.

Constructing ensemble n-sub-epidemic models

We generated ensemble models from both the unweighted (equally weighted top-ranking models) and weighted combination of the highest-ranking sub-epidemic models as deemed by the AIC_{c_i} for the i-th ranked model where $AIC_{c_1} \leq \cdots \leq AIC_{c_I}$ and i = 1, ..., I. We employed the first- and second-ranked models to derive the ensemble forecasts. An ensemble derived from the top-ranking I models are denoted by Ensemble(I). Thus, Ensemble(2) refers to the ensemble model generated from the weighted combination of the top-ranking 2 sub-epidemic models. We compute the weight w_i for the i-th model, i = 1, ... , I, where $\sum w_i = 1$ is as follows:

$$w_i = \frac{l_i}{l_1 + l_2 + \cdots + l_I} \quad \text{for all } i = 1,2, \ldots, I, \tag{8}$$

where l_i is the relative likelihood of model i, which is given by $l_i = e^{((AIC_{min} - AIC_i)/2)}$ [51] and hence $w_I \leq \cdots \leq w_1$. For the unweighted model, we use the same weight $w_i = 1/I$ for all models, i = 1, ... , I.

The prediction intervals based on the ensemble model can be obtained using a bootstrap approach similar as before.

Sub-epidemic wave model (Spatial wave) model

We used a five-parameter epidemic wave model that aggregates linked overlapping sub-epidemics [14]. This sub-epidemic framework can characterize diverse epidemic patterns, including epidemic plateaus where the epidemic stabilizes at a high level for an extended period, and epidemic waves with multiple peaks. The strength (e.g., weak vs. strong) of their overlap determines when the next sub-epidemic is triggered and is controlled by the onset threshold parameter, C_{thr}. The mathematical equation for the sub-epidemic building block is the 3-parameter generalized-logistic growth model (GLM) discussed above (2).

The sub-epidemic wave model consists of a system of coupled differential equations:

$$\frac{dC_i(t)}{dt} = rA_{i-1}(t)C_i(t)^p\left(1 - \frac{C_i(t)}{K_i}\right), \tag{9}$$

Here $C_i(t)$ is the cumulative number of cases for sub-epidemic i, $A_i(t)$ is the indicator function given in (4), and K_i is the size of the i_{th} sub-epidemic where $i = 1, \dots, n$. Starting from an initial sub-epidemic size K_0, the size of consecutive sub-epidemics K_i decline at the rate q following an exponential or power-law function as described below. Hence, a total of 5 parameters (r, p, C_{thr}, q, K_0) for $i = 1, \dots, n$ are needed to characterize a sub-epidemic wave composed of

two more sub-epidemics. The threshold parameters are defined so that $1 \leq C_{thr} < K_0$ and $A_0(t) = 1$ for the first sub-epidemic. We considered a maximum number of three sub-epidemics in the epidemic wave trajectory. The initial number of cases is given by $C_1(0) = I_0$, where I_0 is the initial number of cases in the observed data.

In this framework, the size of the subsequent i_{th} sub-epidemic (K_i) remains steady or declines as a result of the effects of behavior changes or interventions. We considered both exponential and inverse decline functions to model the size of consecutive sub-epidemics as described below.

Exponential decline of sub-epidemic sizes

If consecutive sub-epidemics follow exponential decline, then K_i is given by:

$$K_i = K_0 e^{-q(i-1)} \tag{10}$$

where K_0 is the size of the initial sub-epidemic ($K_1 = K_0$). If $q = 0$, then the model predicts an epidemic wave composed of sub-epidemics of the same size.

Power-law decline of sub-epidemic sizes

If consecutive sub-epidemics decline according to the inverse function, we have:

$$K_i = K_0 \left(\frac{1}{i}\right)^q. \tag{11}$$

Fixed sub-epidemic onset

We utilized sub-epidemic wave models with a fixed onset time at 0, Therefore, all sub-epidemics started at time 0, and the threshold parameter C_{thr} was dropped from the model.

Top-ranked sub-epidemic models

To select the top-ranked sub-epidemic models, we analyzed the AIC_c values of the set of best-fit sub-epidemic wave models with different values of C_{thr}. The AIC_c is given by (7), $m = 5$ is the number of model parameters including parameter C_{thr} and n_d is the number of data points. The AIC_c for the parameter estimation from the nonlinear least-squares fit, which implicitly assumes normal distribution for error.

Auto-Regressive Integrated Moving Average Models (ARIMA)

ARIMA models are commonly employed when forecasting financial [52–54] and weather [55–57] trends, and have become a standard benchmark in disease forecasting [13,16-18, 20-21, 58-59]. The ARIMA (p, d, q) process is given by:

$$\phi(B)(1-B)^d y_t = c + \theta(B)\epsilon_t \tag{12}$$

where B denotes the backshift operator which implies $By_t = y_{t-1}$, $B(By_t) = B^2 y_t = y_{t-2}$, d is the degree of differencing, $\phi(\mathrm{B}) = 1 - \phi_1 \mathrm{B} - \cdots - \phi_p B^p$ and p is the order of the AR model, $\phi(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ and the q is the order of the MA model, and $\{\epsilon_t\}$ is a white noise process with mean 0 and variance σ^2. The constant $c = \mu(1 - \phi_1 - \cdots - \phi_p)$, and μ is the mean of $(1 - B)^d y_t$. When d = 0, μ is the mean of y_t. The *auto.arima* function in the R package "forecast" was used to select orders and build the model [60].

Generalized Additive Models (GAM)

Generalized Additive Models (GAM) are an extension of Generalized Linear Models (GLM) models by including a sum of unknown smooth functions of some covariates [61]. Specific to our study with time as the only covariate, assuming that y_t has normal distribution, the GAM is given as [62]:

$$y_t = \beta_0 + s(t) + \epsilon_t, \tag{13}$$

where s() is an unknown smooth function of time, and $\epsilon_t \sim N(0, \sigma^2)$. We used the *gam* function in the “mgcv” package to fit this model, where the smooth function s() is represented using penalized regression splines. Specifically,

$$s(t) = \sum_{k}^{K} \beta_k b_k(t),$$

(14)

where$\{b_k(.)\}$ represent the B-spline basis functions, $\{\beta_k\}$ are the expansion coefficients to be estimated, and *K* is the number of basis functions [62]. Here, the number of basis functions varied depending on the length of calibration data available for a given forecasting period. A discrete penalty was imposed on the basis coefficients to control the degree of smoothness and the model was fitted by solving a penalized least square problem. The smoothness tuning parameter was selected by the GCV criterion. A more detailed description of the model fitting methodology can be found in [63]. The associated *predict* function was used for forecasting [64].

Performance metrics

To assess performance, we used the mean absolute error (MAE), the mean squared error (MSE), the coverage of the 95% prediction interval (95% PI), and the weighted interval score (WIS). A more detailed description of the performance metrics strategy can be found in [13]. MAE and MSE were used in assessing deviations of the mean model fit to the observed case data [65,66]. See below:

$$MAE = \frac{1}{N} \sum_{h=1}^{N} |f(t_h, \hat{\Theta}) - \tilde{y}_{t_h}|$$

(15)

$$MSE = \frac{1}{N}\sum_{h=1}^{N}\left(f\left(t_h, \hat{\Theta}\right) - \tilde{y}_{t_h}\right)^2$$

(16)

$\tilde{y}_{t_h}$ represents the originally observed case counts across each study location of the *h*-time units ahead forecasts, and t_h are the times for the sample data [65,66].

Model uncertainty was assessed via the 95% prediction interval coverage and WIS. Here, the 95% PI represents the proportion of observed cases that fall within the 95% prediction interval. Through the combination of a set of interval scores (IS) from probabilistic forecasts, the WIS proves quantiles of the predictive forecast distribution. Only a central (1−α)×100% PI is needed to calculate an IS [65]. See below:

$$IS_\alpha(F, y) = (i - l) + \frac{2}{\alpha} * (l - y) * 1(y < l) + \frac{2}{\alpha} * (y - u) * 1(y > u)$$

(17)

The indicator function, represented above by **1**, refers to the fact that **1**($y < l$) = 1 if $y < l$ and 0 otherwise. The $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the forecast *F* are represented by *l* and *u*. The IS consists of three distinct quantities:

1. The sharpness of *F*, given by the width $u - l$ of the central $(1 - \alpha) \times 100\%$ PI.
2. A penalty term $\frac{2}{\alpha} * (l - y) * 1(y < l)$ for the observations that fall below the lower end point *l* of the $(1 - \alpha) \times 100\%$ PI. This penalty term is directly proportional to the

distance between *y* and the lower end *l* of the PI. The strength of the penalty depends on the level.

3. An analogous penalty term $\frac{2}{\alpha} * (y - u) * 1(y > u)$ for the observations falling above the upper limit *u* of the PI.

We report several central PIs at different levels $(1 - \alpha_1) < (1 - \alpha_2) < \cdots < (1 - \alpha_K)$ along with the predictive median, *m*, which can be seen as a central prediction interval at level $1 - \alpha_0 \longrightarrow 0$. This is referred to as the WIS, and it can be evaluated as follows:

$$WIS_{\alpha_{0:K}}(F, y) = \frac{1}{K + \frac{1}{2}} * (w_0 * |y - m| + \sum_{k=1}^{K} w_k * IS_{\alpha_k}(F, y)) \tag{18}$$

where, $w_k = \frac{\alpha_k}{2}$ for $k = 1,2, \ldots K$ and $w_0 = \frac{1}{2}$. Hence, WIS can be interpreted as a measure of how close the entire distribution is to the observation in units on the scale of the observed data [35,67].

Acknowledgements

We would like to acknowledge the GSU Advanced Research Computing Technology & Innovation Core (ARCTIC) team for their help with the development of the Mpox forecast website.

References

1. Mathieu E, Spooner F, Dattani S, Ritchie H, Roser M. Mpox (monkeypox). 2022 [cited 30 Mar 2023]. In: Our World in Data [Internet]. Available from: https://ourworldindata.org/monkeypox

2. Monkeypox declared a global health emergency by the World Health Organization. 23 Jul 2022 [cited 30 Mar 2023]. In: UN News [Internet]. Available from: https://news.un.org/en/story/2022/07/1123152

3. CDC. Technical report 1: Multi-national mpox outbreak, United States, 2022. 27 Feb 2023 [cited 30 Mar 2023]. In: Centers for Disease Control and Prevention [Internet]. Available from: https://www.cdc.gov/poxvirus/mpox/cases-data/technical-report/report-1.html

4. Khan MR, Hossain MJ, Roy A, Islam MR. Decreasing trend of monkeypox cases in Europe and America shows hope for the world: Evidence from the latest epidemiological data. Health Sci Rep. 2023;6: e1030.

5. Fadulu L. Monkeypox Cases Are Declining in New York City and Globally. The New York Times. 26 Aug 2022 [Cited 2023 Mar 2023]. Available from: https://www.nytimes.com/2022/08/26/nyregion/monkeypox-cases-nyc-worldwide.html

6. CDC. 2022 mpox outbreak global map. 09 May 2023 [cited 12 May 2023]. In: Centers for Disease Control and Prevention [Internet]. Available from: https://www.cdc.gov/poxvirus/mpox/response/2022/world-map.html

7. Mpox (monkeypox) outbreak. 2023 [cited 30 Mar 2023]. In: World Health Organization

[Internet]. Available from: https://www.who.int/emergencies/situations/monkeypox-oubreak-2022

8. Philpott D, Hughes CM, Alroy KA, Kerins JL, Pavlick J, Asbel L, et al. Epidemiologic and Clinical Characteristics of Monkeypox Cases - United States, May 17-July 22, 2022. MMWR Morb Mortal Wkly Rep. 2022;71: 1018–1022.

9. Spicknall IH, Pollock ED, Clay PA, Oster AM, Charniga K, Masters N, et al. Modeling the Impact of Sexual Networks in the Transmission of Monkeypox virus Among Gay, Bisexual, and Other Men Who Have Sex with Men - United States, 2022. MMWR Morb Mortal Wkly Rep. 2022;71: 1131–1135.

10. Bleichrodt A, Dahal S, Maloney K, Casanova L, Luo R, Chowell G. Real-time forecasting the trajectory of monkeypox outbreaks at the national and global levels, July-October 2022. BMC Med. 2023;21: 19.

11. Monkeypox forecasting center. 9 Aug 2022 [cited 22 Apr 2023]. In: School of Public Health [Internet]. Available from: https://publichealth.gsu.edu/research/monkeypox-forecasting-center/

12. CDC. Technical Report 3: Multi-National Mpox Outbreak, United States, 2022. 27 Feb 2023 [cited 22 Apr 2023]. In: Centers for Disease Control and Prevention [Internet]. Available from: https://www.cdc.gov/poxvirus/mpox/cases-data/technical-report/report-3.html

13. Chowell G, Dahal S, Tariq A, Roosa K, Hyman JM, Luo R. An ensemble n-sub-epidemic modeling framework for short-term forecasting epidemic trajectories: Application to the

COVID-19 pandemic in the USA. PLoS Comput Biol. 2022;18: e1010602.

14. Chowell G, Tariq A, Hyman JM. A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. BMC Med. 2019;17: 164.

15. CDC. U.S. mpox case trends reported to CDC. 27 Feb 2023 [cited 30 Mar 2023]. In: Centers for Disease Control and Prevention [Internet]. Available from: https://www.cdc.gov/poxvirus/mpox/response/2022/mpx-trends.html

16. Chadaga K, Prabhu S, Sampathila N, Nireshwalya S, Katta SS, Tan R-S, et al. Application of Artificial Intelligence Techniques for Monkeypox: A Systematic Review. Diagnostics (Basel). 2023;13. doi:10.3390/diagnostics13050824

17. Qureshi M, Khan S, Bantan RAR, Daniyal M, Elgarhy M, Marzo RR, et al. Modeling and Forecasting Monkeypox Cases Using Stochastic Models. J Clin Med Res. 2022;11. doi:10.3390/jcm11216555

18. Yasmin F, Hassan MM, Zaman S, Aung ST, Karim A, Azam S. A Forecasting Prognosis of the Monkeypox Outbreak Based on a Comprehensive Statistical and Regression Analysis. Computation. 2022;10: 177.

19. Eid MM, El-Kenawy E-SM, Khodadadi N, Mirjalili S, Khodadadi E, Abotaleb M, et al. Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases. Sci China Ser A Math. 2022;10: 3845.

20. Khan MI, Qureshi H, Bae SJ, Awan UA, Saadia Z, Khattak AA. Predicting monkeypox incidence: Fear is not over! J Infect. 2023;86: 256–308.

21. Long B, Tan F, Newman M. Forecasting the Monkeypox Outbreak Using ARIMA, Prophet, NeuralProphet, and LSTM Models in the United States. Forecasting. 2023;5: 127–137.

22. McAndrew T, Majumder MS, Lover AA, Venkatramanan S, Bocchini P, Besiroglu T, et al. Early human judgment forecasts of human monkeypox, May 2022. Lancet Digit Health. 2022;4: e569–e571.

23. Scheffer M, Paiva VSF, Barberia LG, Russo G. Monkeypox in Brazil between stigma, politics, and structural shortcomings: have we not been here before? Lancet Reg Health Am. 2023;17: 100394.

24. Fuqua School of Business. Introduction to ARIMA models. [cited 24 Apr 2023]. Available from: https://people.duke.edu/~rnau/411arim.htm

25. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, et al. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics. 2018;22: 13–21.

26. Reich NG, Lessler J, Funk S, Viboud C, Vespignani A, Tibshirani RJ, et al. Collaborative Hubs: Making the Most of Predictive Epidemic Modeling. Am J Public Health. 2022;112: 839–842.

27. Johansson MA, Apfeldorf KM, Dobson S, Devita J, Buczak AL, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc Natl Acad Sci U S A. 2019;116: 24268–24274.

28. Reich NG, Brooks LC, Fox SJ, Kandula S, McGowan CJ, Moore E, et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States.

Proc Natl Acad Sci U S A. 2019;116: 3146–3154.

29. McGowan CJ, Johansson M, Apfeldorf KM, Ben-Nun M, Brooks L, Convertino M, et al. Collaborative efforts to forecast seasonal influenza in the United States, 2015–2016. Sci Rep. 2019;9: 1–13.

30. Ray EL, Wattanachit N, Niemi J, Kanji AH, House K, Cramer EY, et al. Ensemble forecasts of Coronavirus disease 2019 (COVID-19) in the U.s. bioRxiv. medRxiv; 2020. doi:10.1101/2020.08.19.20177493

31. Webb CT, Ferrari M, Lindström T, Carpenter T, Dürr S, Garner G, et al. Ensemble modelling and structured decision-making to support Emergency Disease Management. Prev Vet Med. 2017;138: 124–133.

32. Arieli I, Babichenko Y, Smorodinsky R. Robust forecast aggregation. Proc Natl Acad Sci U S A. 2018;115: E12135–E12143.

33. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLoS Comput Biol. 2019;15: e1007486.

34. Oidtman RJ, Omodei E, Kraemer MUG, Castañeda-Orjuela CA, Cruz-Rivera E, Misnaza-Castrillón S, et al. Trade-offs between individual and ensemble forecasts of an emerging infectious disease. Nat Commun. 2021;12: 5379.

35. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Castro Rivadeneira AJ, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the

United States. Proc Natl Acad Sci U S A. 2022;119: e2113561119.

36. 2022-23 Mpox (Monkeypox) Outbreak: Global Trends. 2023 [cited 24 Apr 2023]. In: World Health Organization [Internet]. Available from: https://worldhealthorg.shinyapps.io/mpx_global/

37. Delaney KP. Strategies adopted by gay, bisexual, and other men who have sex with men to prevent monkeypox virus transmission—United States, August 2022. MMWR Morb Mortal Wkly Rep. 2022;71.

38. Stone W, Huang P. The monkeypox outbreak may be slowing in the U.S., but health officials urge caution. NPR. 26 Aug 2022 [Cited 2023 Apr 23]. Available from: https://www.npr.org/sections/health-shots/2022/08/26/1119659681/early-signs-suggest-monkeypox-may-be-slowing-in-the-u-s.

39. Monkeypox outbreak can be eliminated in Europe, WHO says. Reuters. 30 Aug 2022 [Cited 2023 Apr 24. Available from: https://www.reuters.com/world/europe/monkeypox-outbreak-can-be-eliminated-europe-who-officials-2022-08-30/.

40. CDC. Case Definitions† for Use in the 2022 Mpox Response. 27 Feb 2023 [cited 30 Mar 2023]. In: Centers for Disease Control and Prevention [Internet]. Available from: https://www.cdc.gov/poxvirus/mpox/clinicians/case-definition.html

41. Health Organization W. Surveillance, case investigation and contact tracing for mpox (monkeypox): interim guidance, 22 December 2022. 2022 [cited 30 Mar 2023]. In: World Health Organization [Internet]. Available from: https://apps.who.int/iris/bitstream/handle/10665/365398/WHO-MPX-Surveillance-2022.4-

eng.pdf?sequence=1

42. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018.

43. Chowell G, Hincapie-Palacio D, Ospina J, Pell B, Tariq A, Dahal S, et al. Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. PLoS Curr. 2016;8. doi:10.1371/currents.outbreaks.f14b2217c902f453d9320a43a35b9583

44. Pell B, Kuang Y, Viboud C, Chowell G. Using phenomenological models for forecasting the 2015 Ebola challenge. Epidemics. 2018;22: 62–70.

45. Shanafelt DW, Jones G, Lima M, Perrings C, Chowell G. Forecasting the 2001 Foot-and-Mouth Disease Epidemic in the UK. Ecohealth. 2018;15: 338–347.

46. Banks HT, Hu S, Clayton Thompson W. Modeling and Inverse Problems in the Presence of Uncertainty. CRC Press; 2014.

47. Roosa K, Luo R, Chowell G. Comparative assessment of parameter estimation methods in the presence of overdispersion: a simulation study. Math Biosci Eng. 2019;16: 4299–4313.

48. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2009 Aug.

49. Hurvich CM, Tsai C-L. Regression and time series model selection in small samples. Biometrika. 1989;76: 297–307.

50. Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. Commun Stat Theory Methods. 1978;7: 13–26.

51. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer Science & Business Media; 2003.

52. Almasarweh M, AL Wadi S. ARIMA Model in Predicting Banking Stock Market Data. Modern Applied Science. 2018;12: 309.

53. Ariyo AA, Adewumi AO, Ayo CK. Stock Price Prediction Using the ARIMA Model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. 2014. pp. 106–112.

54. Mondal P, Shit L, Goswami S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. International Journal of Computer Science, Engineering and Applications. 2014;4: 13.

55. Dimri T, Ahmad S, Sharif M. Time series analysis of climate variables using seasonal ARIMA approach. J Earth Syst Sci. 2020;129: 149.

56. Shamsnia SA, Shahidi N, Liaghat A, Sarraf A, Vahdat SF. Modeling of weather parameters using stochastic methods (ARIMA model)(case study: Abadeh Region, Iran). International Conference on Environment and Industrial Innovation, IPCBEE 2011 (Vol. 12). 2011. Available from: https://www.semanticscholar.org/paper/Modeling-of-Weather-Parameters-Using-Stochastic-(-)-Shamsnia-Shahidi/4e1ec09bf667fbf0f815dfc20a5ec16aea4499a8

57. Tektaş M. Weather Forecasting Using ANFIS and ARIMA MODELS. EREM. 2010;51: 5–10.

58. Aditya Satrio CB, Darmawan W, Nadia BU, Hanafiah N. Time series analysis and

forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. Procedia Comput Sci. 2021;179: 524–532.

59. Wang Y-W, Shen Z-Z, Jiang Y. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. PLoS One. 2018;13: e0201987.

60. Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. J Stat Softw. 2008;27: 1–22.

61. Wood SN. Generalized Additive Models: An Introduction with R, Second Edition. CRC Press; 2017.

62. Shafi A. What are Generalised Additive Models? 18 May 2021 [cited 31 Mar 2023]. In: Towards Data Science [Internet]. Available from: https://towardsdatascience.com/generalised-additive-models-6dfbedf1350a

63. gam function - RDocumentation. [cited 8 May 2023]. Available: https://www.rdocumentation.org/packages/mgcv/versions/1.8-42/topics/gam

64. Predict function - RDocumentation. [cited 31 Mar 2023]. Available: https://www.rdocumentation.org/packages/car/versions/3.1-1/topics/Predict

65. Gneiting T, Raftery AE. Strictly Proper Scoring Rules, Prediction, and Estimation. J Am Stat Assoc. 2007;102: 359–378.

66. Kuhn M, Johnson K. Applied Predictive Modeling. Springer New York; 2013 Sep 1.

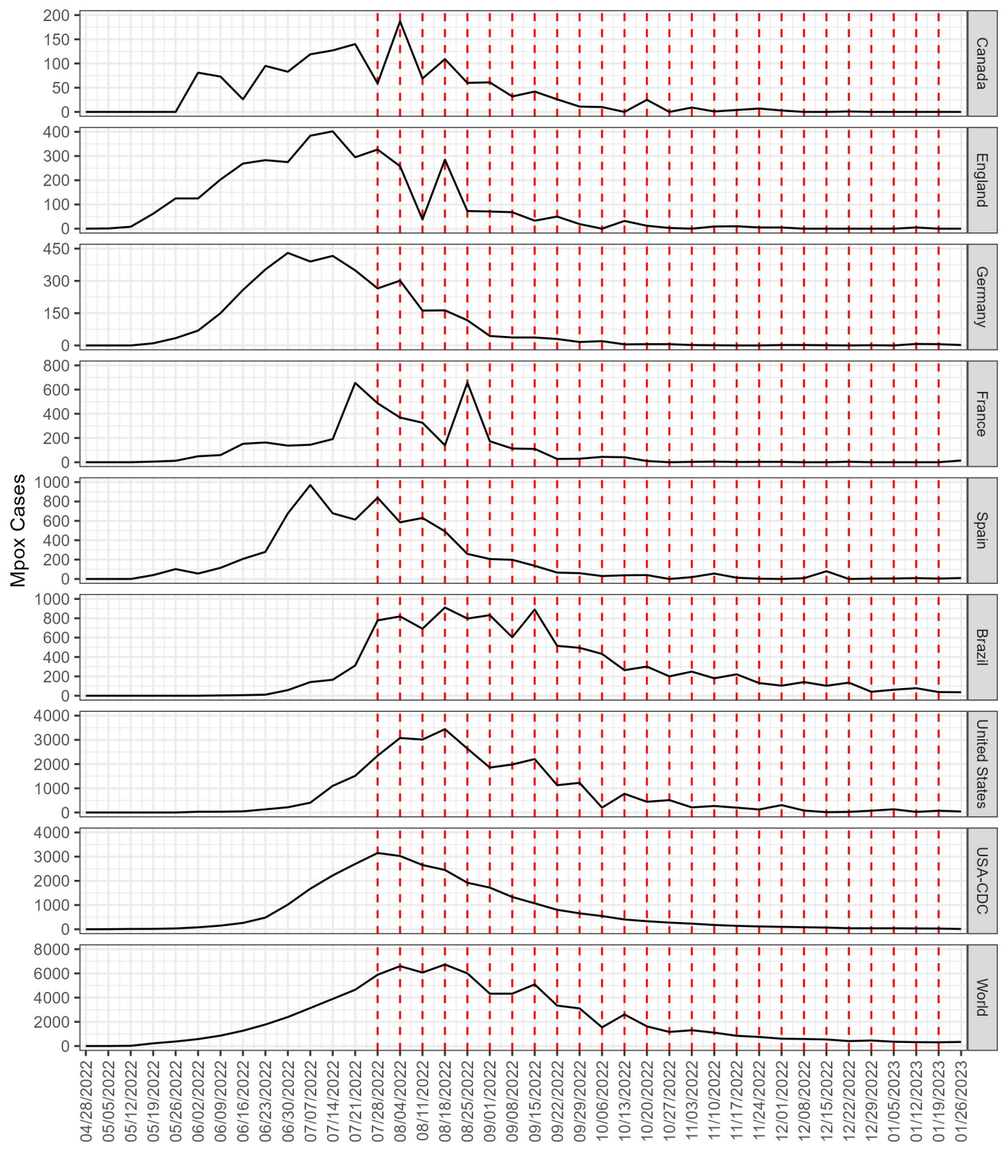
67. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. PLoS Comput Biol. 2021;17: e1008618.

Supporting Information

S1 Table. Average forecast performance metrics and skill scores by location. This file contains individual tables for each location for average metrics, and for skill scores.

S2 Table. Forecast performance metrics for each forecasting period, location, model, and forecasting horizon. This file contains an Excel table with the forecasting performance metrics for each location, forecasting period, forecasting horizon, and model.

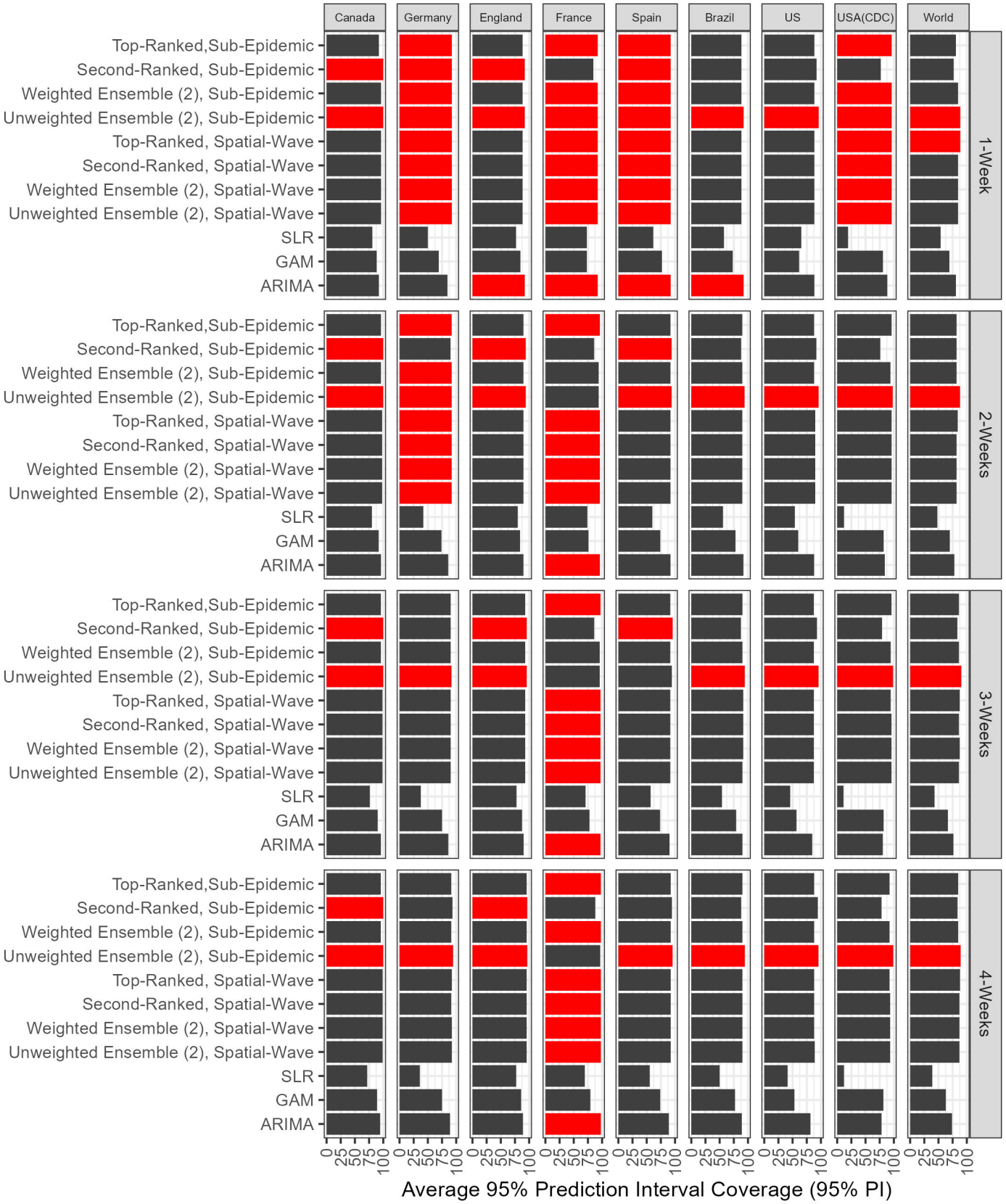
S1 Fig. Visualizations of forecast period specific, and average forecast performance. This file contains multiple figures visualizing the forecast performance metrics for each location, forecasting horizon, model, and forecasting period.

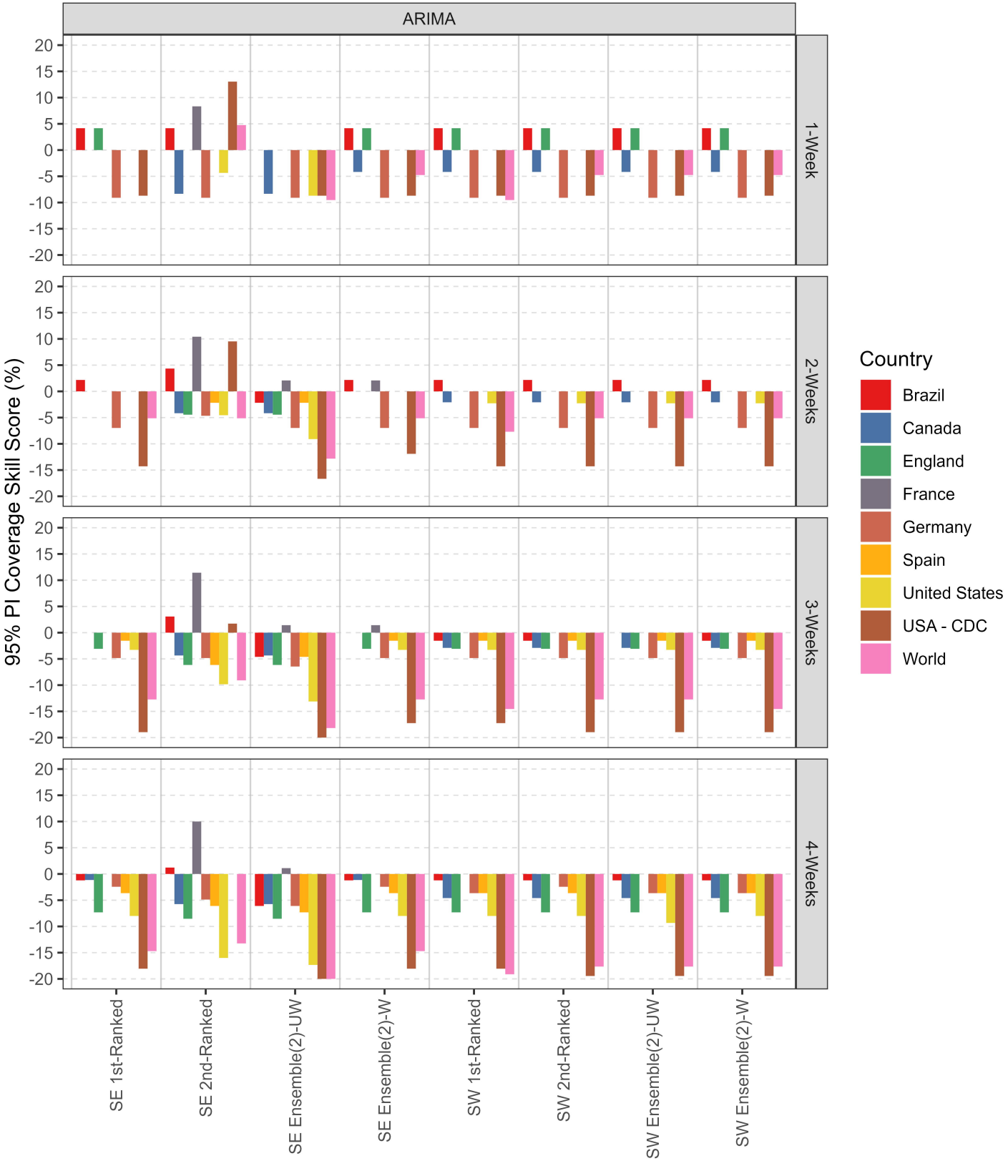

Mpox Cases
Canada
England
Germany
France
Spain
Brazil
United States
USA-CDC
World
04/28/2022
05/05/2022
05/12/2022
05/19/2022
05/26/2022
06/02/2022
06/09/2022
06/16/2022
06/23/2022
06/30/2022
07/07/2022
07/14/2022
07/21/2022
07/28/2022
08/04/2022
08/11/2022
08/18/2022
08/25/2022
09/01/2022
09/08/2022
09/15/2022
09/22/2022
09/29/2022
10/06/2022
10/13/2022
10/20/2022
10/27/2022
11/03/2022
11/10/2022
11/17/2022
11/24/2022
12/01/2022
12/08/2022
12/15/2022
12/22/2022
12/29/2022
01/05/2023
01/12/2023
01/19/2023
01/26/2023

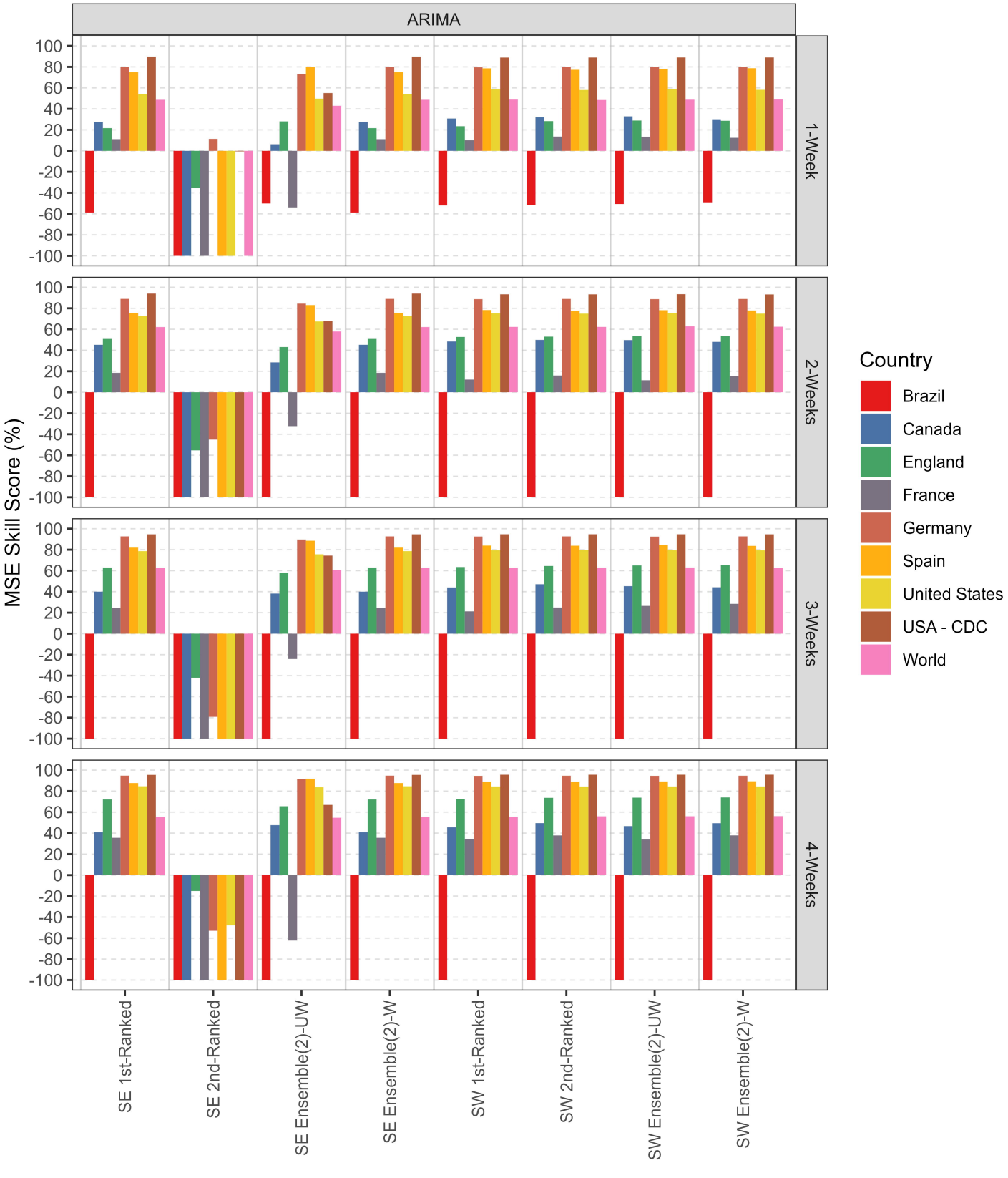
Canada
Germany
England
France
Spain
Brazil
US
USA(CDC)
World
1-Week
2-Weeks
3-Weeks
4-Weeks
Top-Ranked,Sub-Epidemic
Second-Ranked, Sub-Epidemic
Weighted Ensemble (2), Sub-Epidemic
Unweighted Ensemble (2), Sub-Epidemic
Top-Ranked, Spatial-Wave
Second-Ranked, Spatial-Wave
Weighted Ensemble (2), Spatial-Wave
Unweighted Ensemble (2), Spatial-Wave
SLR
GAM
ARIMA
0.00 1.00 2.00 3.00 4.00
0.00 1.17 2.35 3.52 4.70
0.00 1.64 3.29 4.93 6.58
0.00 1.54 3.09 4.63 6.18
0.00 1.71 3.42 5.13 6.85
0.00 1.61 3.22 4.82 6.43
0.00 1.96 3.92 5.88 7.85
0.00 2.50 5.00 7.50 10.00
Average Mean Square Error (MSE) (log10)

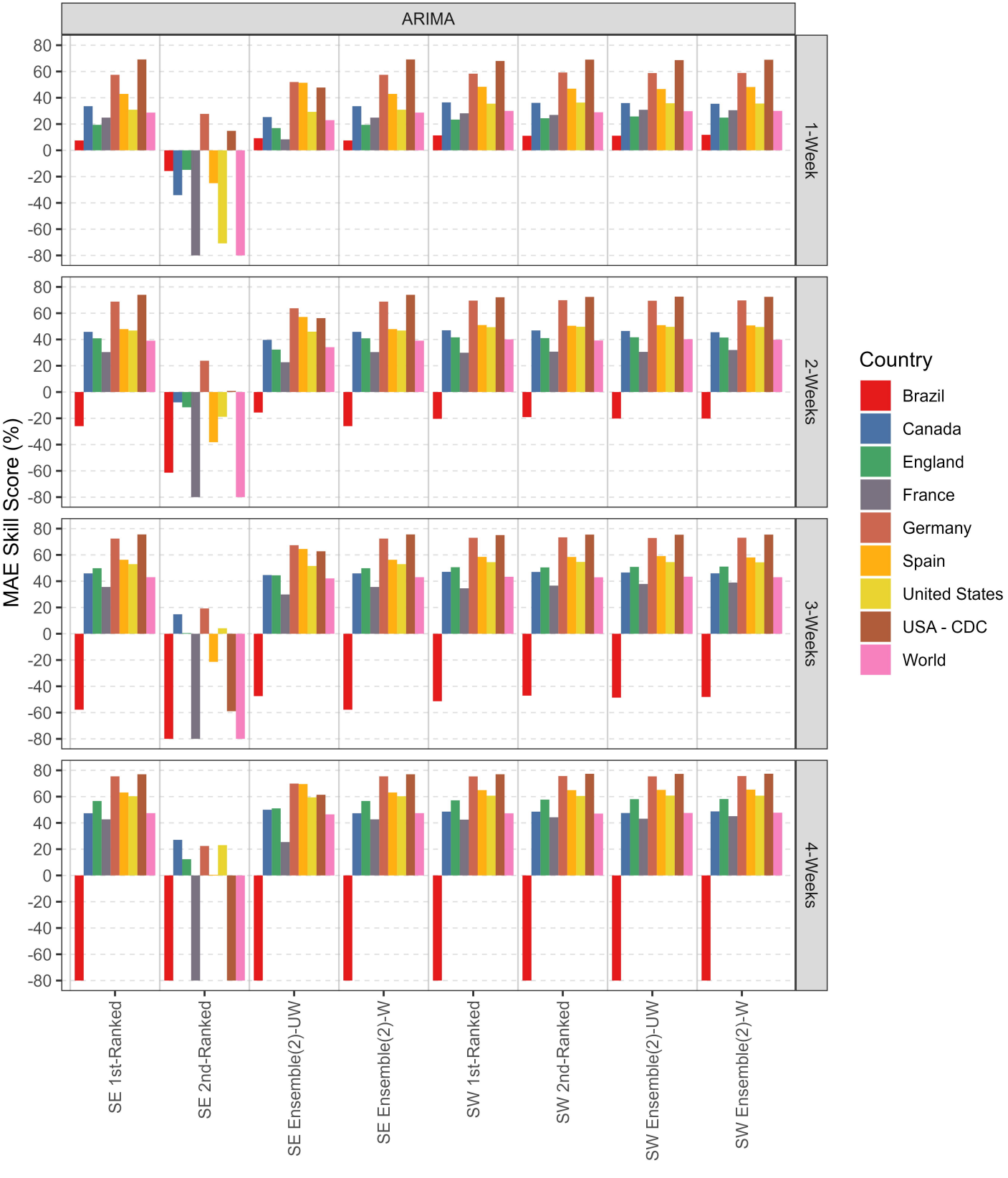
Canada
Germany
England
France
Spain
Brazil
US
USA(CDC)
World
Top-Ranked,Sub-Epidemic
Second-Ranked, Sub-Epidemic
Weighted Ensemble (2), Sub-Epidemic
Unweighted Ensemble (2), Sub-Epidemic
Top-Ranked, Spatial-Wave
Second-Ranked, Spatial-Wave
Weighted Ensemble (2), Spatial-Wave
Unweighted Ensemble (2), Spatial-Wave
SLR
GAM
ARIMA
1-Week
2-Weeks
3-Weeks
4-Weeks
0.00 0.44 0.89 1.33 1.78
0.00 0.54 1.09 1.63 2.18
0.00 0.58 1.15 1.73 2.30
0.00 0.61 1.22 1.83 2.44
0.00 0.64 1.27 1.91 2.54
0.00 0.67 1.35 2.02 2.70
0.00 0.79 1.59 2.38 3.18
0.00 0.79 1.59 2.38 3.18
0.00 1.04 2.09 3.13 4.18
Average Mean Absolute Error (MAE) (log10)

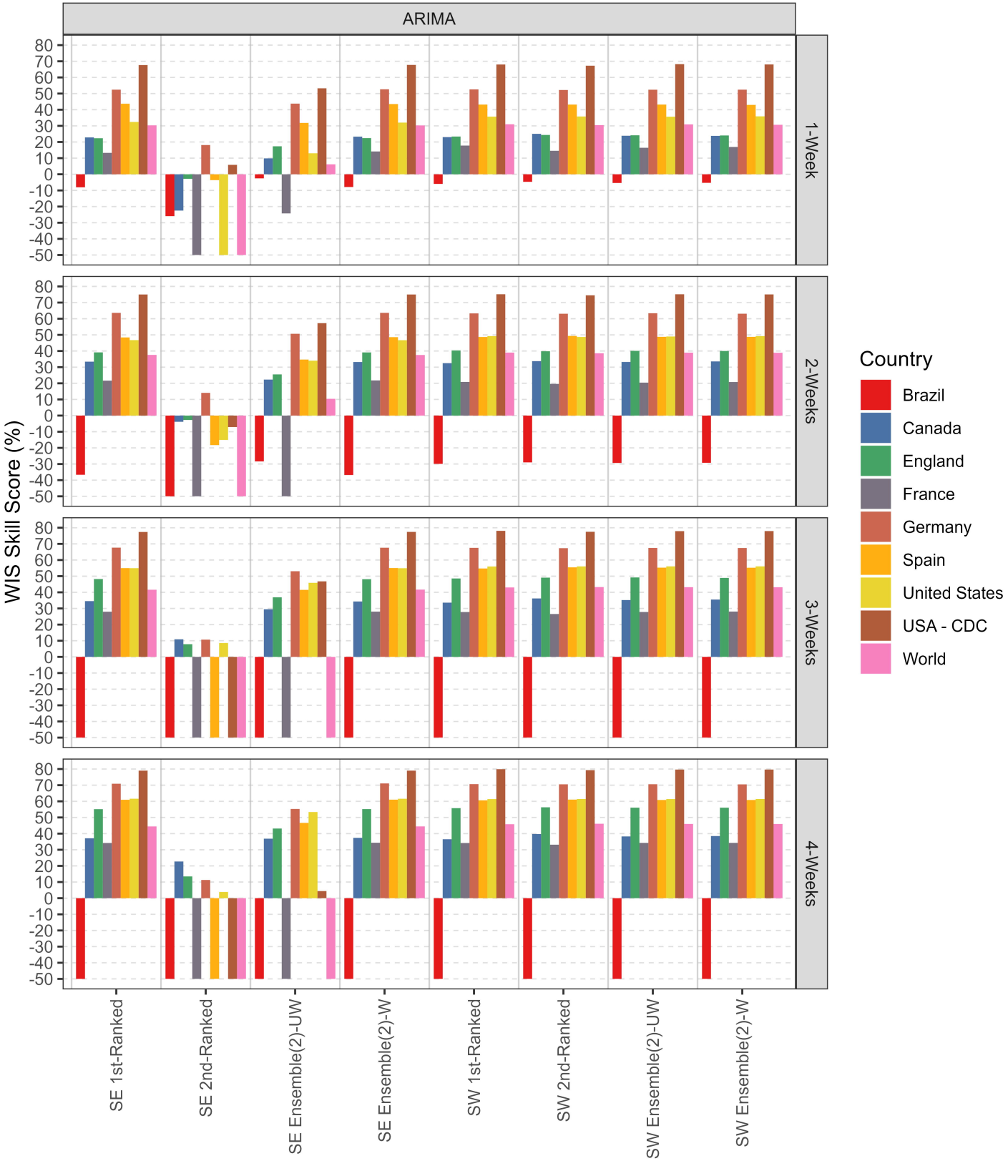
Canada
Germany
England
France
Spain
Brazil
US
USA(CDC)
World
Top-Ranked,Sub-Epidemic
Second-Ranked, Sub-Epidemic
Weighted Ensemble (2), Sub-Epidemic
Unweighted Ensemble (2), Sub-Epidemic
Top-Ranked, Spatial-Wave
Second-Ranked, Spatial-Wave
Weighted Ensemble (2), Spatial-Wave
Unweighted Ensemble (2), Spatial-Wave
SLR
GAM
ARIMA
1-Week
2-Weeks
3-Weeks
4-Weeks
0.00 0.44 0.89 1.33 1.78
0.00 0.54 1.09 1.63 2.18
0.00 0.56 1.13 1.69 2.26
0.00 0.61 1.22 1.83 2.44
0.00 0.65 1.30 1.95 2.60
0.00 0.67 1.35 2.02 2.70
0.00 0.79 1.59 2.38 3.18
0.00 0.79 1.59 2.38 3.18
0.00 1.32 2.64 3.96 5.28
Average Weighted Interval Score (WIS) (log10)


Canada
Germany
England
France
Spain
Brazil
US
USA(CDC)
World
1-Week
2-Weeks
3-Weeks
4-Weeks
Top-Ranked,Sub-Epidemic
Second-Ranked, Sub-Epidemic
Weighted Ensemble (2), Sub-Epidemic
Unweighted Ensemble (2), Sub-Epidemic
Top-Ranked, Spatial-Wave
Second-Ranked, Spatial-Wave
Weighted Ensemble (2), Spatial-Wave
Unweighted Ensemble (2), Spatial-Wave
SLR
GAM
ARIMA
0
25
50
75
100
Average 95% Prediction Interval Coverage (95% PI)


ARIMA
95% PI Coverage Skill Score (%)
1-Week
2-Weeks
3-Weeks
4-Weeks
SE 1st-Ranked
SE 2nd-Ranked
SE Ensemble(2)-UW
SE Ensemble(2)-W
SW 1st-Ranked
SW 2nd-Ranked
SW Ensemble(2)-UW
SW Ensemble(2)-W
Country
Brazil
Canada
England
France
Germany
Spain
United States
USA - CDC
World


ARIMA
MSE Skill Score (%)
1-Week
2-Weeks
3-Weeks
4-Weeks
SE 1st-Ranked
SE 2nd-Ranked
SE Ensemble(2)-UW
SE Ensemble(2)-W
SW 1st-Ranked
SW 2nd-Ranked
SW Ensemble(2)-UW
SW Ensemble(2)-W
Country
Brazil
Canada
England
France
Germany
Spain
United States
USA - CDC
World


ARIMA
MAE Skill Score (%)
1-Week
2-Weeks
3-Weeks
4-Weeks
SE 1st-Ranked
SE 2nd-Ranked
SE Ensemble(2)-UW
SE Ensemble(2)-W
SW 1st-Ranked
SW 2nd-Ranked
SW Ensemble(2)-UW
SW Ensemble(2)-W
Country
Brazil
Canada
England
France
Germany
Spain
United States
USA - CDC
World


ARIMA
WIS Skill Score (%)
1-Week
2-Weeks
3-Weeks
4-Weeks
SE 1st-Ranked
SE 2nd-Ranked
SE Ensemble(2)-UW
SE Ensemble(2)-W
SW 1st-Ranked
SW 2nd-Ranked
SW Ensemble(2)-UW
SW Ensemble(2)-W
Country
Brazil
Canada
England
France
Germany
Spain
United States
USA - CDC
World