

AIM-F Measles Forecasting Final Model

Documentation

Additions

1. Updated Methods

- Changes made from Alpha, to Beta, to Final

2. Discussion & Commentary

- The effect of seasonality and data quality
- Policy/Public Health Implications
- Generalizability
 - i. Additional pathogens/ or health outcomes
 - ii. Variable aggregation for different levels of geographic resolution (provided availability of data)
 - iii. Minimal effort to expand to different time steps

3. Lessons Learned

- What are some lessons from the sweeping explorations of model structure that the team has done?
- Data availability challenges/ data landscape discussion
- Practical uses for limited data

4. Future Projections (update next section)

- Could also do back testing and showcase improvements from Alpha to Final Model (dependent on July data)

5. User Guide- ReadMe (exported as PDF as well)

Aug 29, 2025

Contents

Model Overview and Objective	3
Background	3
Input Datasets	3
Case Data	4
Climate Data	6
Precipitation	6

Temperature	6
Social and Demographic Data	6
Case Data Alignment	6
Seasonal Clustering	7
Prediction Window Cutoff	7
Preprocessor	7
Independent Variable Projection	8
Data Landscape Assessment	9
Limited Automated Data Access	9
Costs Associated With Critical Data Access	9
Inconsistent and Opaque Data Update Schedules	9
Implications for Model Development and Use	10
Country Clustering	10
Outcome Variable	12
Model fitting and selection	13
Country inclusion criteria	14
Missing value handling	14
Model assessment	15
Model performance and validation	18
Prediction Tier validation	18
Comparison to baseline	24
Future Projections	25
Discussion and Commentary	28
The Effect of Seasonality and Data Quality	28
Conclusion	34
Legal Notice & Disclaimers	34
Appendix	35

Model Overview and Objective

The AIM-F Measles Forecasting Model is a flexible modeling framework that predicts country-specific measles incidence (cases per million population) on a monthly timestep. The flexibility of the codebase allows for country-specific model configurations, enabling training and forecasting using the most effective predictor variables and machine learning methods for that specific country.

The model was designed with the goal in mind to help support decision-making on which countries to prioritize for Supplementary Immunization Activities (SIA) by predicting which countries will have large outbreaks in the coming months. Although the forecast window is flexible, a 9-month prediction window was chosen. Countries were prioritized as needing supplementary immunization based upon the predicted number of months with greater than 5 cases per million population in the prediction window.

Background

Measles continues to present a serious challenge to global health, driving significant mortality including an estimated 128,000 deaths in 2021. The greatest burden of disease is observed in children under five.¹ Current vaccine deployment efforts face logistical challenges in the timely allocation of doses, particularly when estimating current and future risks and forecasting where vaccines will be needed within advance timeframes of approximately 3 to 12 months. While some countries have endemic measles patterns that provide sufficient data for reliable seasonal predictions, many other countries with sporadic transmission and singular, black-swan epidemics require more innovative approaches to forecast future vaccine needs.

Input Datasets

All candidate input datasets are provided as a csv file in the github repository ([other deliverables/alpha_model/input_datasets.csv](#)) containing the name and a description of the dataset, the geographic and temporal scope of the data, the file format, and the link to the data download. Some datasets contained several variables assessed by the model. A separate data dictionary is included as a csv file

¹ World Health Organization. Measles Fact Sheet. May 31, 2023.
<https://www.who.int/news-room/fact-sheets/detail/measles#:~:tex>

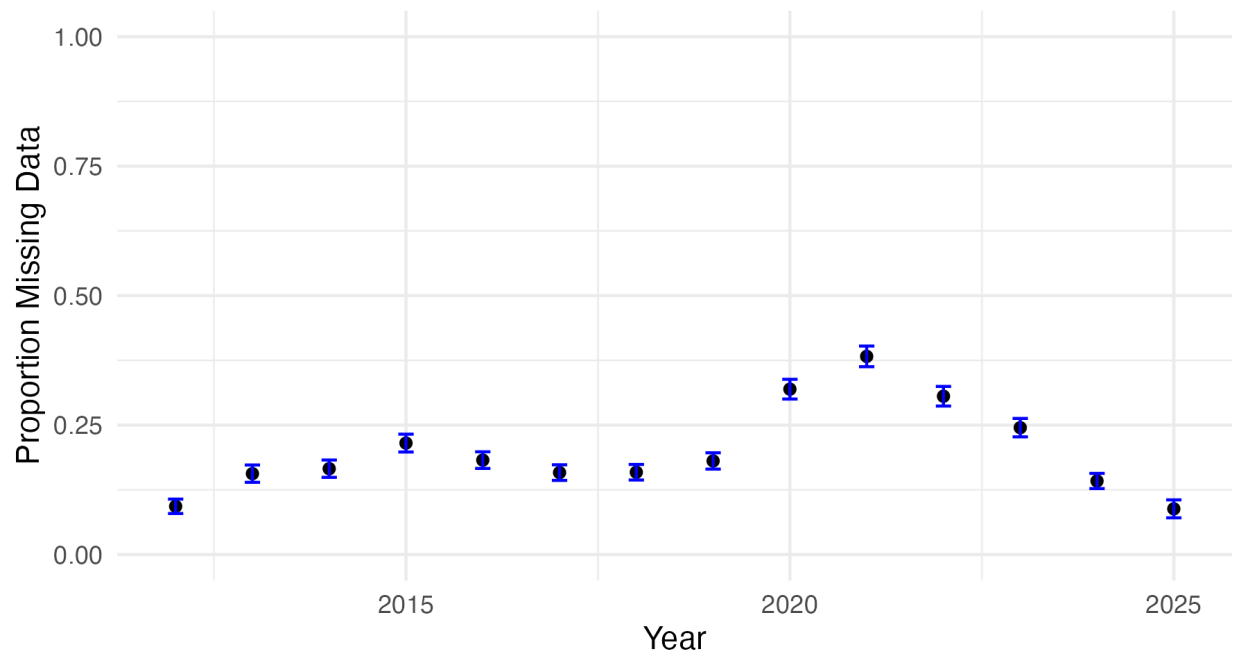
([other_deliverables/alpha_model/input_variable_description.csv](#)) that lists all input variables assessed by the model, their parent dataset, and gives a brief description of the variable.

Case Data

Measles case data from 2011 – present was obtained from the WHO provisional monthly measles and rubella data,² which provides monthly counts of country-level laboratory confirmed, epidemiologically linked, and clinical cases as reported to the World Health Organization. These data were transformed to cases per million population using annual UN population projections.³ The proportion of case data that was missing varied over time (range: 0.08 - 0.4; median 0.17) and between countries (range: 0 - 1; median 0.17) as seen in Figures 1, 2, and A1 in the appendix). Data missingness can cause significant challenges when creating models and may impact model uncertainty.

Figure 1.

Proportion of missing measles case data points by year with 95% confidence interval shown in blue.



²World Health Organization. "Provisional Monthly Measles and Rubella Data." World Health Organization, <https://www.who.int/teams/immunization-vaccines-and-biologicals/immunization-analysis-and-insights/surveillance/monitoring/provisional-monthly-measles-and-rubella-data>. Accessed 03 September 2025..

³United Nations. "World Population Prospects 2024." United Nations, <https://population.un.org/wpp/downloads>. Accessed 25 June 2025.

Figure 2a.

Proportion of missing data points by country with 95% confidence interval shown in blue.

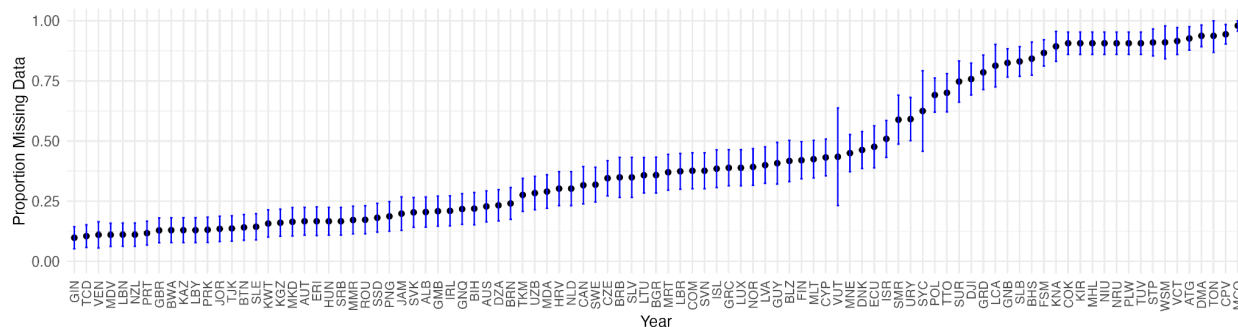
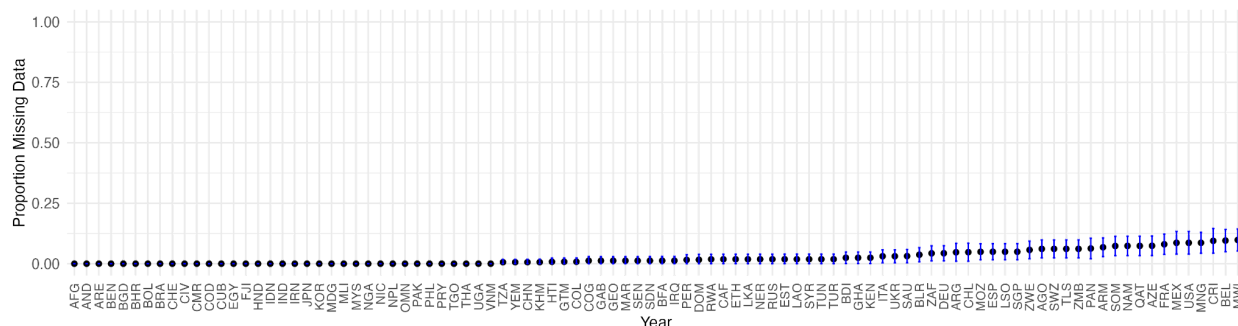


Figure 2b.

Proportion of missing data points by country with 95% confidence interval shown in blue (continuation from above)



The case incidence was used to create binary outbreak outcomes at varying thresholds ranging from 2 to 20 cases per million population within a month. Further outbreak-related variables were derived from binary outbreak outcomes, such as the number of months since the previous outbreak occurred, the proportion of outbreak months during a 12, 24, and 60 month rolling window, and the total number of outbreaks observed within the time series. Lastly, case incidence was used to create z-scores representing monthly case incidence standardized over periods of 12, 36, or 60 months. Standardizing case incidence in this way may be more informative for the model to detect unusually high or low measles incidence that deviates from historical norms, regardless of absolute numbers.

Climate Data

Climate data, specifically temperature⁴ and precipitation⁵, were included as predictors as they are associated with measles seasonality, particularly in countries with recurring outbreaks.

Precipitation

Precipitation datasets for 2012 – present were acquired from NOAA as gridded, temporal data in the .nc format. Precipitation data is available as a monthly mean value for each grid. These files were processed by the data ingestion pipeline to create a country-level mean monthly value by geocoding each grid to a global shapefile and extracting the country, then taking the average of all grid values in a country.

Temperature

Temperature data for 2012 – present were acquired via API from Our World in Data in the form of monthly average surface temperatures by country.

Social and Demographic Data

This category of data includes public health preparedness and response, socioeconomic, political, and demographic measures. These variables were assessed as predictors with the reasoning that they may influence the size of the susceptible population (e.g., measles vaccination coverage, birth rate), the likelihood of disease importation (e.g., air travel patterns, migration), the capacity of public health agencies to respond to outbreaks (e.g., public health spending, public health capacity scores), and access to or trust in preventative and response care (e.g., socioeconomic and political variables). Many of these variables were available as table downloads with an annual resolution and did not require processing beyond merging with the case data by year.

Case Data Alignment

To prepare the data for model fitting and selection, the case data was aligned so that a prediction would be made in a period of low measles activity preceding the coming measles season. The cutoff month for the prediction window was first determined through the application of seasonal

⁴ Our World in Data. "Monthly average surface temperatures by year." Retrieved from <https://ourworldindata.org/grapher/monthly-average-surface-temperatures-by-year.csv?v=1&csvType=full&useColumnShortNames=false>. Accessed 25 June 2025.

⁵ National Oceanic and Atmospheric Administration. "CPC Global Unified Temperature." National Oceanic and Atmospheric Administration, <https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html>. Accessed 25 June 2025.

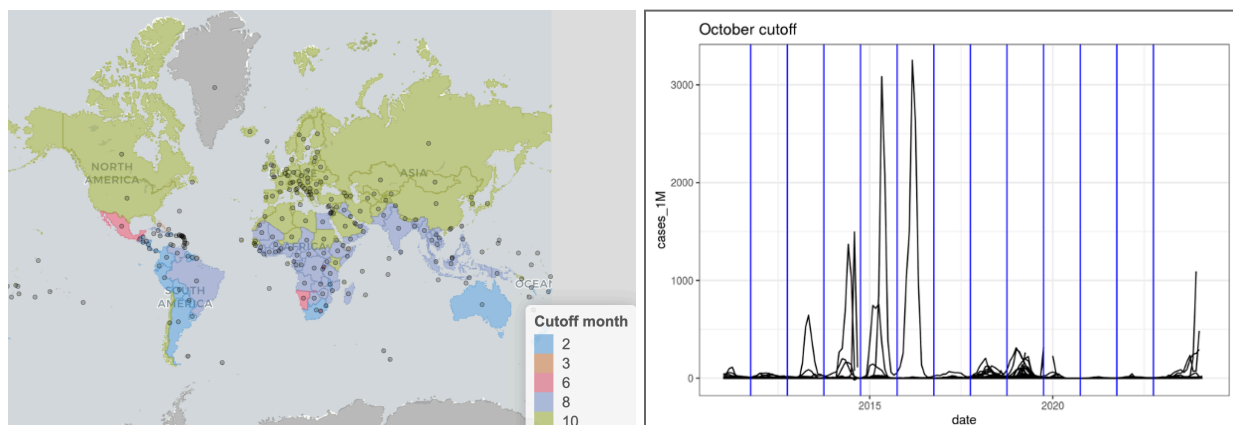
clustering of the case data (as described below). A prediction window of 9 months was then created for each country for model fitting and selection.

Seasonal Clustering

Given that measles outbreaks do not occur in sync across the globe, the cutoff date of the prediction window in the model evaluation was allowed to vary by seasonal patterns of measles outbreaks within the country. A time series k-means clustering analysis was performed using precipitation, temperature, and case data to group countries with similar patterns in measles outbreaks and determine the best month for initiating the cutoff date. Figure 3 on the left shows a map of the month chosen for the cutoff date by country. An example of how this cutoff month corresponds with patterns of measles cases is shown in Figure 3 on the right.

Figure 3.

A map showing the case alignment cutoff month by country (left). Time-series measles incidence per million population with seasonal cutoffs shown as vertical blue lines (right).



Prediction Window Cutoff

Cutoff months were determined by the clustering methods outlined above, but determining the year of the prediction window cutoff was performed by choosing the latest year that had at least 6 months of reported data after the cutoff month for the validation cutoff period (as of September 2025). This method ensured models were being evaluated using a sufficient amount of data across countries.

Preprocessor

Input data went through preprocessing prior to model fitting to 1) interpolate or drop missing data when necessary, 2) ensure proper case data alignment during global-local model fitting, and 3)



transform data types when necessary. The most common types of data preprocessing applied included dropping missing values or replacing them with 0, linearly interpolating annual predictors, and encoding categorical data to integers. A list of all applied methods from the alpha model and their description is available in Table 1. The type of processing was dependent on the temporal resolution and completeness of the data. The processing method(s) applied to each of the input variables are listed in Table A2 in the appendix as well as [input_variable_description.csv](#) in github.

Independent Variable Projection

Predictor variables are forward-projected using univariate time series forecasting models following the application of the preprocessor and predictor lags. In the case of training runs, predictors in the test set are first truncated using the cutoff month for the projection window before being forward-projected from the training data to ensure proper blinding before forecasting. For future outbreak predictions, the same process is applied albeit from the end of the available data and without truncation of the input data set. These predictor projections are then aligned in a futures table on which a fitted model may then draw its predictions. At present, we are using the StatsForecast implementation of AutoETS for predictor projections as it readily incorporates seasonality, largely automates the underlying fitting processes, and showed strong performance in community model comparisons.

Table 1.

A list and description of all preprocessor methods

Preprocessor method	Description
zero_fill	Fill 0 for all missing values
back_fill	First observation carried back
forward_fill	Last observation carried forward
back_truncate	Drop rows with leading NAs
forward_truncate	Drop NAs at end of data
first_valid_overwrite	First non-NA value fills entire column
average_valid_overwrite	Average of non-NA value fills entire column
linear	Fill NA with linear prediction
pass_unchanged	No processing applied
january_only	For years with repeated data for every month, drop all values but the first
yn_to_int	Cast str to no:0, yes:1
interpolate_via_{str}	interp1d {method}

remap_by_{str}	Using a dict/json structure, remap expected categorical values to numeric. Case insensitive, maps underscores to spaces.
zero_min	Replace < 0 values with 0
check_coverage_{str}	Reject columns below an acceptable coverage level
check_gaps_{str}	Reject columns with gaps beyond an acceptable length

Data Landscape Assessment

The process of building and deploying the AIM-F Measles Forecasting Model highlights several persistent challenges related to the data landscape. We identified three key pain points in data accessibility during our efforts of assembling the data needed for developing this model: 1) limited automated data access, 2) costs associated with critical data access, and 3) inconsistent and opaque data update schedules. These points are elaborated below.

Limited Automated Data Access

A significant limitation in measles forecasting is that most relevant datasets are not available for automated acquisition through APIs. Instead, many sources only offer manual downloads in non-standardized formats (e.g., CSV, Excel, PDF, or custom table downloads). This absence of programmatic access restricts the capacity to implement streamlined web scraping pipelines or perform continuous, end-to-end automated forecasting. As a result, teams are required to devote substantial time and effort to manually download, clean, and harmonize new data releases before updating models. These manual processes introduce delays and the potential for human error, further hampering rapid and scalable public health planning.

Costs Associated With Critical Data Access

Some datasets, especially those covering mobility, require licensing fees for access. For this model, air travel patterns are a key predictor variable that requires a license for access. Although the modified data can be shared open source in this model, the underlying data would not be publicly available when attempting to refresh data, expand models to new geographies, or rapidly iterate on predictor variables. These resource constraints sometimes mean essential data (e.g., timely migration statistics or granular air travel data) is entirely unavailable, limiting the robustness and generalizability of the forecasting framework.

Inconsistent and Opaque Data Update Schedules

Data update schedules vary widely across sources. For example, case and climate datasets are generally updated regularly (although it is clear that not all countries report case data with the

same timeliness), but social and demographic datasets may be released sporadically and not on a predictable timetable. Publication dates are frequently not communicated in advance, making it difficult to anticipate or automate dataset refreshes. This inconsistency leads to asynchronous update cycles in different model components and introduces temporal misalignments in the input data. In practice, the forecasting model may be working with outdated case or predictor information for significant periods, increasing model uncertainty and reducing confidence in predictions for public health planning.

Implications for Model Development and Use

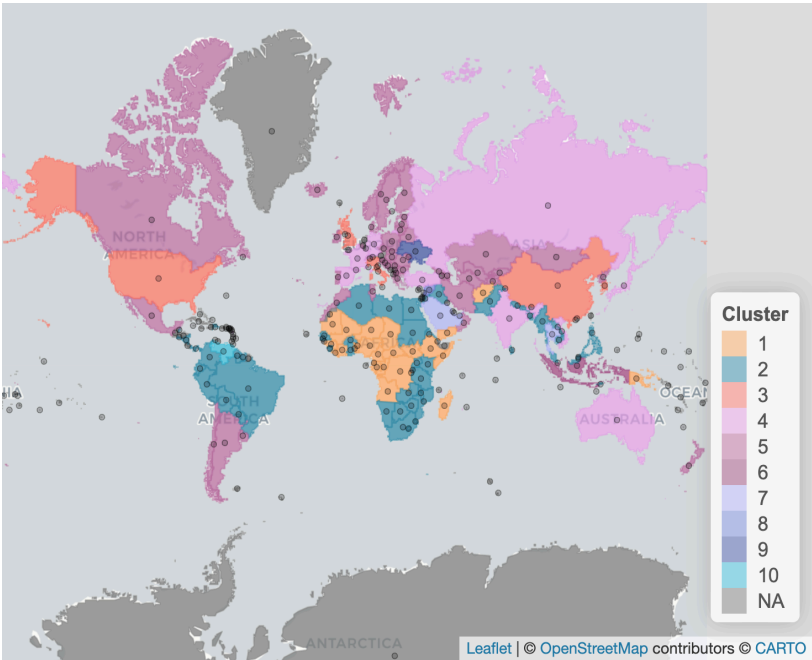
The AIM-F Measles Forecasting Model operates within a challenging and fragmented data landscape, characterized by limited automated access to key datasets, financial barriers to acquiring some predictors, and inconsistent, often opaque data update schedules across sources. To mitigate these difficulties, we have developed a unified data ingestion pipeline to assemble diverse inputs into a standardized file, facilitating repeatable model workflows. An automated preprocessor script ensures consistent handling by aligning case data, transforming data types, and systematically managing missing values through interpolation or exclusion. Proper case data alignment is enforced during model fitting to address temporal mismatches, and categorical or time-series predictors are transformed as needed to support robust analysis. While these interventions improve the reliability and comparability of model inputs, residual uncertainty may persist due to asynchronous updates or data gaps, and users should interpret outputs with an understanding of these ongoing challenges.

Country Clustering

The quality of case and predictor variable data varies by country, with some countries experiencing gaps in data or lags in data updates. The models fit to these countries tend to perform poorly due to low data quality. With the goal of using the data from other countries with similar patterns in case and predictor data, we assigned countries to clusters based on their measles case and predictor data.

We used a hierarchical clustering algorithm and dynamic time warping to calculate the distance matrices between the time series case and predictor data. The clustering algorithm yielded the dendrogram in Figure 4.

Figure 5. Global map of countries by cluster



Outcome Variable

To create an outcome variable that we could use to prioritize countries by their predicted future measles activity, we developed a prioritization tiering system (Table 2). First, we created a threshold number of cases per population which would be used to identify outbreak months. Threshold values of 2, 5, 10, and 20 cases per million population were evaluated and 5 cases per million was selected as it gave a reasonable proportion of positive months (approximately 0.2) across affected countries while also flagging most countries with significant outbreak history. A prediction window of 9 months was chosen to balance the need for future planning and the increase in uncertainty the further in the future the model predicts.

The number of observed and predicted outbreak months in the 9-month prediction window was then used to place countries in prioritized tiers as shown in the table below. When spot-checked against a few time points, this grouping generally resulted in 10-20 countries per tier when using the observed case data.

Table 2.

Tiers constructed for country prioritization by predicted future measles activity

Prioritization Tier	Number of outbreak months in the prediction window
A - Higher	7 or more months
B - Moderate	4 – 6 months
C - Lower	1 – 3 months
None	0 months

Based on the alignment of a country's observed and predicted tier, we assigned a country's prediction as acceptable, over, or under (Table 3). To bias against under-prediction (given the more detrimental public health impact of missing a coming outbreak), we wanted to allow the model an acceptable amount of overprediction. Therefore, we defined an acceptable prediction as predicting the correct tier or one tier higher. Over-prediction was defined as a predicted tier more than one tier higher than the correct tier (yellow cells). Under-prediction was defined as predicting a lower tier (or no tier) than the correct tier (red cells).

Table 3.

This schema illustrates how a country's observed and predicted Prioritization Tiers are categorized as acceptable (green), over (yellow), or under (red) predictions.

Prioritization Tier		Observed			
		A	B	C	None
Predicted	A - Higher				
	B - Moderate				
	C - Lower				
	None				

Model fitting and selection

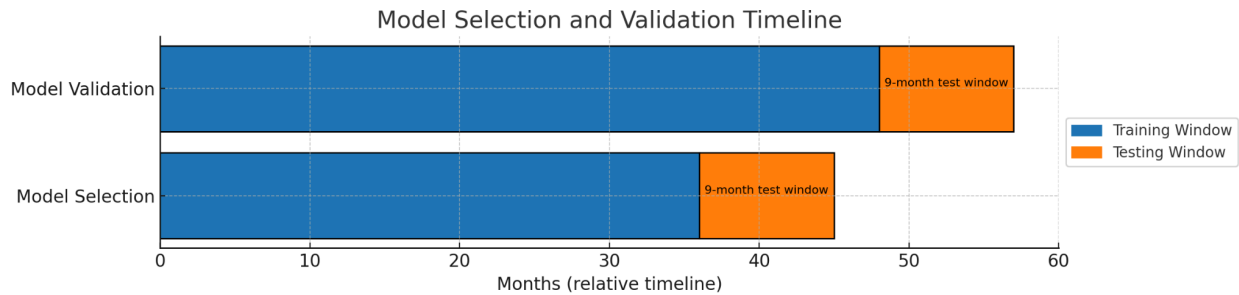
A grid search approach was used to find the best-fitting model for each country for the final model phase. All combinations of countries and country clusters, ML model methods (see Table A3), and up to three predictor variables were tested (see Table 3 for candidate predictor variables) in two sets of model sweeps: one for model selection and one for model validation. Predictor variable sets varied by country and were determined for each country through a three step process: process of 1) perform univariate statistical analysis of the relationship of the candidate predictor variable and outbreak status, keeping those with a significant relationship 2) of the significant variables, perform a test for multicollinearity and maintain a country-specific list of correlated variables, and 3) create a list of all possible single, two, or three predictor variable sets, constraining the sets to not include correlated variables. In addition, all models contained lagged temperature and precipitation variables. The code used to create the candidate predictor variable sets for each country can be found in [github final_model/create_final_model_metadata.R](#). Each model was run with 15 replicates and the replicate with the median number of predicted cases in the test window was used in downstream model assessments.

We first assessed models during a model selection window, where the train/test split occurred one year earlier than what was determined in the Prediction Window Cutoff section above. Then, we moved the train/test split one year forward and ran the same models again for a model validation testing period. A hypothetical timeline is shown in Figure X.

Figure X.



A hypothetical timeline of training and test window periods for model selection and model validation runs. Cutoff dates varied by country depending on seasonal factors and the availability of case data.



Country inclusion criteria

In order to increase confidence in model projections, we imposed a minimum number of months of available case data for country inclusion. Countries with fewer than 36 months of case data available⁶ in the WHO measles and rubella provisional monthly dataset were excluded from model fitting.

Furthermore, the goal of this model is to provide decision-making support for prioritization of countries needing supplementary immunization. With that goal in mind, we excluded countries that never had an outbreak year reported⁷, defined as reaching 20 cumulative cases per million population within a calendar year from our modeling scope. While these countries were not targeted in our model development and selection process, their data were used in country cluster model training.

Missing value handling

While some missing values for country-level predictor data could be reasonably projected or interpolated (seasonal climate data for example), as a general rule truncation was applied via the preprocessor in cases where predictors could not be defensibly projected. In such cases, all

⁶ ATG, BHS, COK, CPV, DMA, FSM, GNB, GRD, KIR, KNA, LCA, MCO, MHL, NIU, NRU, PLW, SLB, STP, SUR, SYC, TON, TUV, VCT, VUT, WSM. *Note countries may meet multiple exclusion criteria.*

⁷ AND, ARG, ATG, AUS, BHS, BLZ, BOL, BRB, BRN, CHL, COK, COL, CPV, CRI, CUB, CYP, DMA, DNK, DOM, ECU, FIN, GRD, GTM, GUY, HND, HTI, HUN, IRN, JAM, JOR, JPN, KIR, KNA, KOR, KWT, LCA, LVA, MAR, MEX, MHL, NIC, NIU, NOR, NRU, PAN, PER, PLW, PRK, PRT, PRY, RWA, SLV, STP, SUR, SWE, SWZ, TKM, TTO, TUV, URY, USA, VCT, ZAF. *Note countries may meet multiple exclusion criteria.*



outcomes and predictors for a given country were likewise truncated to ensure data completeness within training data rows. Following this, several quality checks were applied to ensure that predictors and countries alike had sufficient data for fitting. Predictors that did not have at least two unique non-NAN values within the training data were dropped. Likewise, countries were dropped if after preprocessing they did not have at least 6 months of case data in the prediction window, data for one or more viable predictor variables, or at least twice as many rows in the prepared training data as were targeted in the prediction window. In the case of a single country fit, non-viable predictor variables were logged and dropped. Likewise, in the case of fitting country clusters using global-local model fitting, countries were logged and dropped if they were missing one or more predictor variables to ensure the best fit to the available data, though this behavior may be toggled to drop predictors instead.

Model assessment

Of these runs, we chose the best predictor set by country using the mean absolute error (MAE) in the model selection test window for nine different subsets of model configurations for further analysis. The nine subsets were each possible combination of single model, ensemble model, single country fit, and cluster fit (Table 4).

Table 4.

Each possible subset (rows) of possible model configurations considered, the number of countries able to be modeled through the validation test window, and the proportion of countries with acceptable predictions. To illustrate which fits and models are included in the subset, cells are shaded and marked with a ✓ to indicate inclusion. The selected configuration is bolded.

Model configuration	Countries modeled through validation	Prop acceptable	Single country fit	Cluster fit	Single Model	Ensemble
Single country fit + single model	112	0.768	✓		✓	
Single country fit + ensemble	105	0.762	✓			✓
Single country fit (any model)	109	0.734	✓		✓	✓
Cluster fit + single model	102	0.696		✓	✓	
Cluster fit + ensemble model	22	0.777		✓		✓

Cluster fit (any model)	97	0.701		✓	✓	✓
Single model (any fit)	106	0.726	✓	✓	✓	
Ensemble model (any fit)	76	0.723	✓	✓		✓
Any fit, any model	104	0.731	✓	✓	✓	✓

Once the best predictor set for each country and configuration was selected based on the test MAE in the model selection window, we assessed each model's performance in the validation window according to our [Prediction Tier methodology](#).

Once tiers were assigned, we calculated the proportion of countries with acceptable predictions in the validation test window and the number of countries that were able to be modeled through the validation window (Table 4), and chose the model configuration with the best balance between proportion of acceptable countries and the total number of countries able to be modeled. The best configuration according to this criteria was the single country fit + single model, with 76.8% of 112 modeled countries having an acceptable prediction (see [Model performance and validation](#) for more details).

We proceeded with the best predictor sets for each country from the single country fit + single model as our top model for the final model phase. The top model for each country could contain up to three predictors (not including mean temperature and precipitation) in their predictor sets, with 30% of countries having one predictor, 38% of countries having two predictors, and 32% having three predictors in their top model. The top five predictor variables that appeared in the top models of the highest proportion of countries were: 1) the number of months since the last SIA campaign, 2) the proportion of previous months in outbreak (5 cases per M) during a rolling 12 month period, 3) the proportion of previous months in outbreak (5 cases per M) during a rolling 24 month period, 4) the number of migrations per thousand population, and 5) the number of months since the last outbreak (20 cumulative cases per M). Each predictor variable present in the top model set is listed in Table 5, along with the proportion of countries containing the variable in the predictor set.

Table 5.

Most common predictor variables included in the final model. Note: since the models are multivariate, the proportions will not sum to 1.

Variable	Proportion of selected country models containing the variable
Months since last SIA campaign	0.21

Proportion of previous rolling 12 months in outbreak (5 cases per M)	0.188
Proportion of previous rolling 24 months in outbreak (5 cases per M)	0.170
Migrations per thousand population	0.170
Months since last outbreak (20 cumulative cases per M)	0.161
Monthly air passengers arriving from countries with seasonal measles outbreaks	0.152
Cases per M 12 month z-score	0.152
Birth rate per thousand population	0.143
Proportion of previous rolling 60 months in outbreak (5 cases per M)	0.134
MCV2 coverage	0.134
Months since last outbreak (20 cases per M)	0.116
Monthly air passengers visiting countries with seasonal measles outbreaks	0.116
Cases per M 36 month z-score	0.071
MCV1	0.063
Cases per M 60 month z-score	0.036

Since the model configuration that yielded the highest percentage of acceptable predictions only included single country fits and single models, no ensemble models were present in the final model, however all five single ML model options were represented with Gradient boosting being the most common and XGB Regressor being the second (Table 6).

Table 6.

Proportion of selected country models utilizing the machine learning method

Machine Learning Model	Proportion of selected country models utilizing the machine learning method
Gradient boosting	0.232
XGB Regressor	0.223
Bagging Regressor	0.214
CatBoost	0.205
Random Forest	0.125

Model performance and validation

Rigorous validation was conducted to assess the predictive performance of the model following selection of the final country-specific models. For each country, predicted measles activity was evaluated against observed case data using a 9-month prediction window, ensuring that no validation data influenced model training. This window was shifted one year forward from the model selection prediction window, which was used to select the best performing model by country.

Prediction Tier validation

The data for the prediction window were again used to assign each country's observed and predicted prioritization tiers (Table 7).

Table 7.

Observed and predicted prioritization tiers for model validation

Prioritization Tier		Observed			
		A	B	C	None
Predicted	A - Higher	6	5	2	7
	B - Moderate	0	1	3	4
	C - Lower	0	4	3	9
	None	0	1	8	59

Overall, we found that the model performed well with acceptable predictions made for 76.8% of predicted countries, 11.6% were over-predicted, and 11.6% under-predicted within their assigned prioritization tiers (Table 8). A Fisher's exact test performed on the results reveals a highly significant correlation between the observed and predicted tiers ($p = 2.18 \times 10^{-7}$).

Table 8.

Summary of prioritization tier prediction for validation (based on Table 7)

Prediction	Number of countries	Proportion of total
Over-predicted	13	0.116
Acceptable	86	0.767
Under-predicted	13	0.116

The model demonstrated strong performance in predicting outbreaks in countries with no observed outbreaks, correctly classifying 59 out of 68 (86.8%) of these countries. It also accurately categorized all 6 countries (100%) that fell into prediction tier A based on observed case data. However, the model's accuracy decreased for mid-tier predictions. For countries in prediction tier B, only 6 out of 11 (54.5%) were correctly classified, and for tier C, the model correctly classified only 6 out of 16 (37.5%) countries.

We also looked at the output by month throughout the training period and the 9-month prediction window. A positive month was defined as having more than or equal to 5 reported measles cases per million population. For each month, a result was coded using the binary alignment of the observed and predicted outbreak data, as shown in Table 9.

Table 9.

Description of result categories for monthly predictions

Result	Observed	Predicted
True Positive (dark green)	≥ 5 cases per million	≥ 5 cases per million
True Negative (light green)	< 5 cases per million	< 5 cases per million
False Positive (dark purple)	< 5 cases per million	≥ 5 cases per million
False Negative (light purple)	≥ 5 cases per million	< 5 cases per million

Results by month for the training period and prediction window are shown by the observed prioritization tier in Figures 6a - d. For visualization purposes, the data is aligned relative to the

cutoff date which is labeled as time 0 and shown by a vertical white line and only the 100 months leading up to the cutoff date are shown.

Figure 6a.

Heatmap for monthly results for countries in **Prioritization Tier A** based on observed data for the validation prediction window.

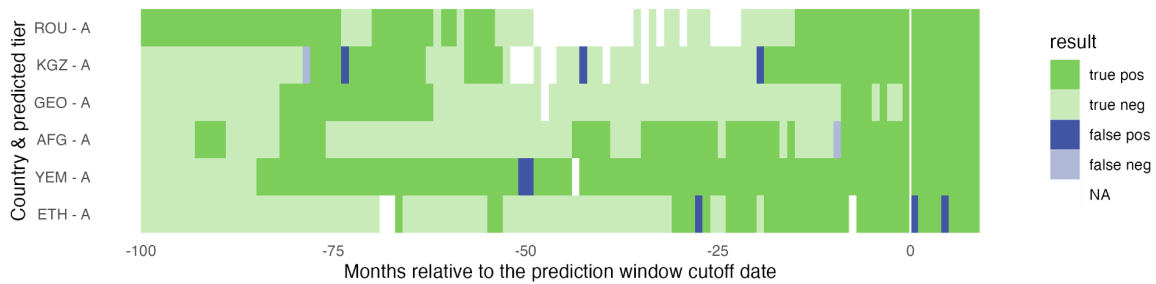


Figure 6b.

Heatmap for monthly results for countries in **Prioritization Tier B** based on observed data for the validation prediction window.

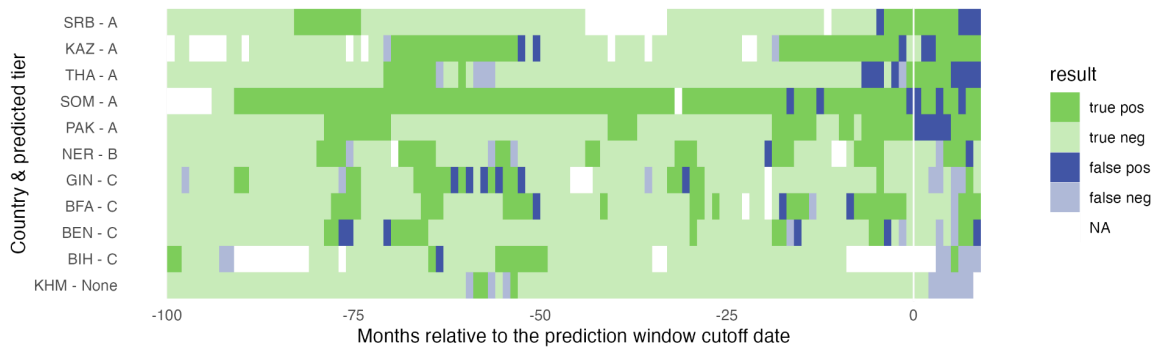


Figure 6c.

Heatmap for monthly results for countries in **Prioritization Tier C** based on observed data for the validation prediction window.

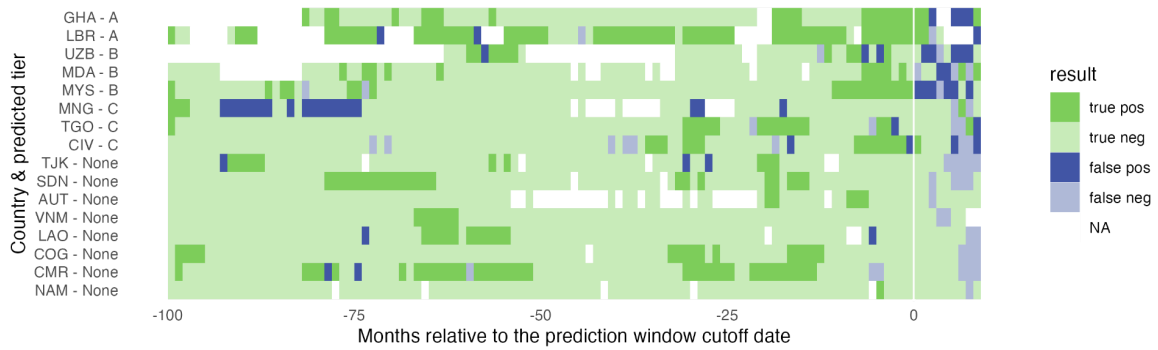
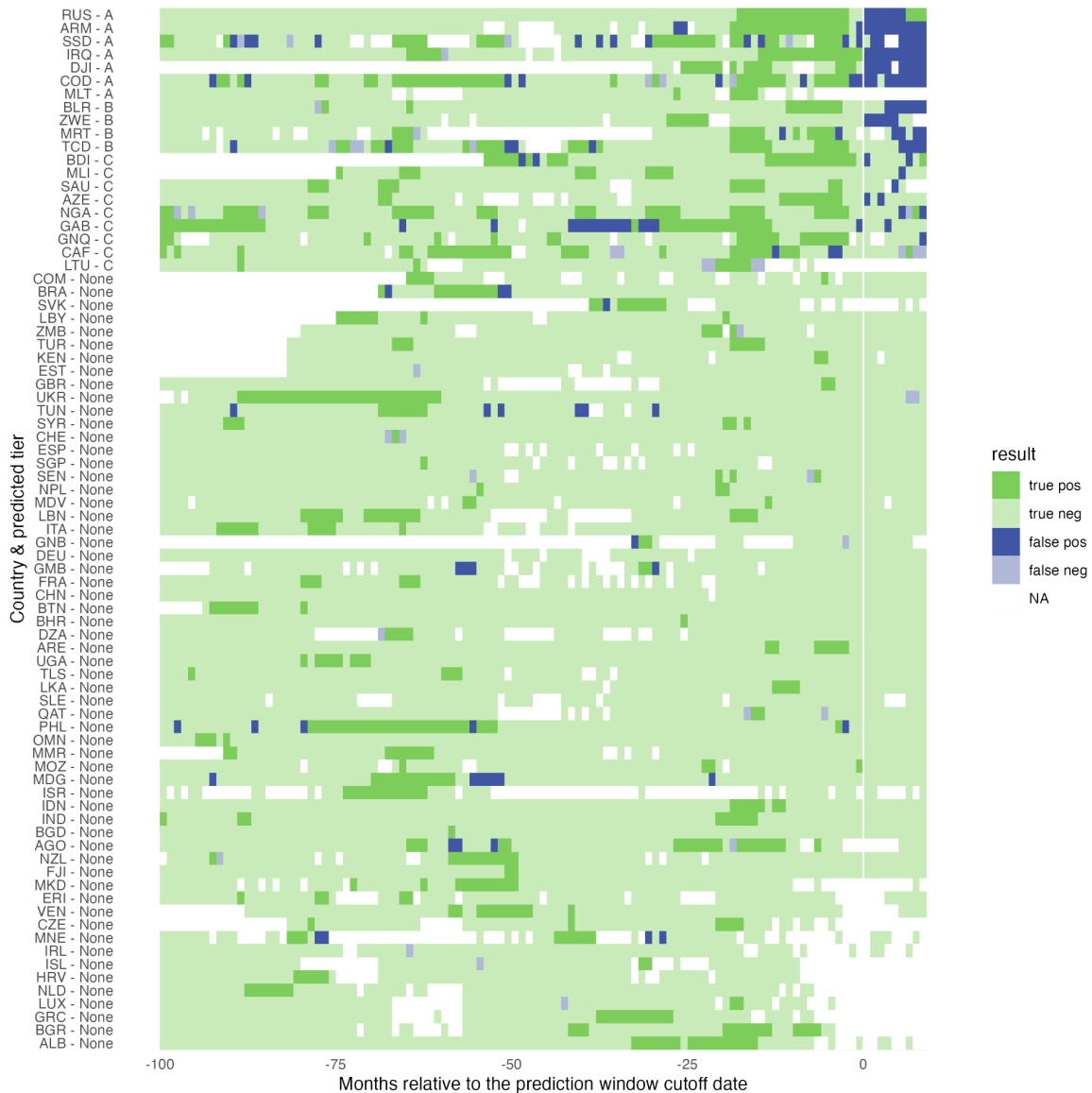


Figure 6d.

Heatmap for monthly results for countries not placed in a tier based on observed data for the validation prediction window.



In total, we modeled 15,152 months across all 112 modeled countries with sufficient input data (14,154 for the training period and 998 in the prediction window). Overall, the model correctly predicted (true positive or true negative) 95% of the months, with a slight increase to 97% for the training months and a decrease to 80% for the prediction window, as expected (Tables 13 a and

b). This gave a sensitivity (proportion of observed positives that are accurately predicted) of 64% and specificity (proportion of observed negatives that are accurately predicted) of 83% for the prediction window.

Table 13a.

*Accuracy of monthly predictions for the **training period***

Result		Observed	
		Positive	Negative
Predicted	Positive	1,970 (14%)	370 (2%)
	Negative	102 (1%)	11,712 (83%)

Table 13b.

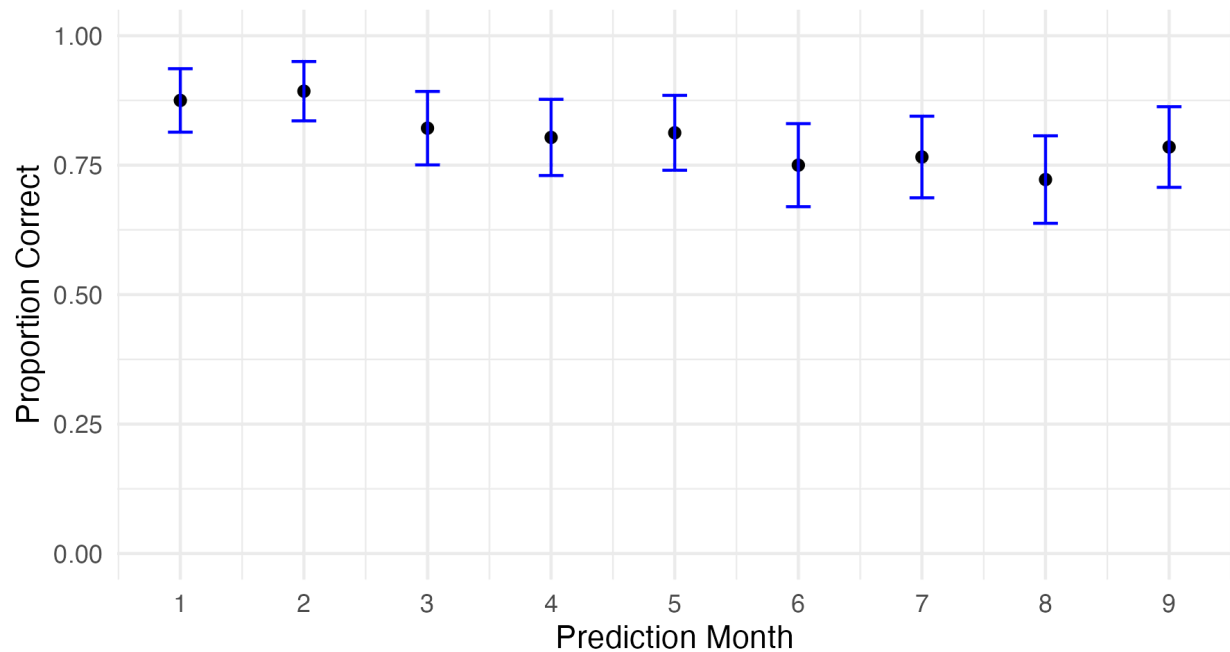
*Accuracy of monthly predictions for the **prediction window***

Result		Observed	
		Positive	Negative
Predicted	Positive	105 (10%)	138 (14%)
	Negative	58 (6%)	697 (70%)

To assess how model accuracy changes moving further into the future, we calculated the proportion of correct predictions by month throughout the 9-month prediction window (Figure 7). The proportion correct ranged from 0.89 for month 2 (September 2024) to 0.72 in month 8 (March 2025), with a general pattern of decreasing the further into the future the prediction was.

Figure 7.

Proportion of correct predictions by month for the prediction window with 95% confidence intervals shown with blue bars.

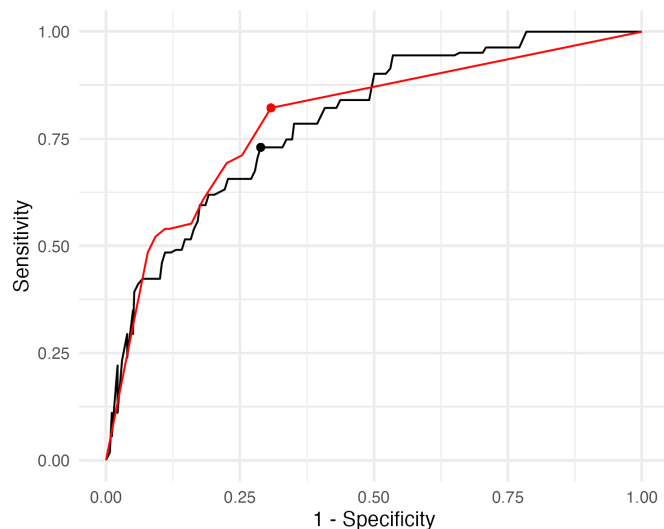


Comparison to baseline

We created a simple baseline model based on the proportion of months observing an outbreak prior to the validation window to compare the results of our ML model. We compared sensitivity (Se), specificity (Sp), and accuracy as indicated by the Youden index optimal predicted probability threshold on the ROC curve (Fig. X).

Figure X.

The ROC curve of the ML model (red) and the baseline model (black) with Youden's optimal threshold indicated by the point.



Our ML model provided an improvement over the baseline model in sensitivity and accuracy, but slightly under-performed in specificity (Table X).

Table X.

Sensitivity (Se), specificity (Sp), and Accuracy of the baseline and machine-learning (ML) forecast model.

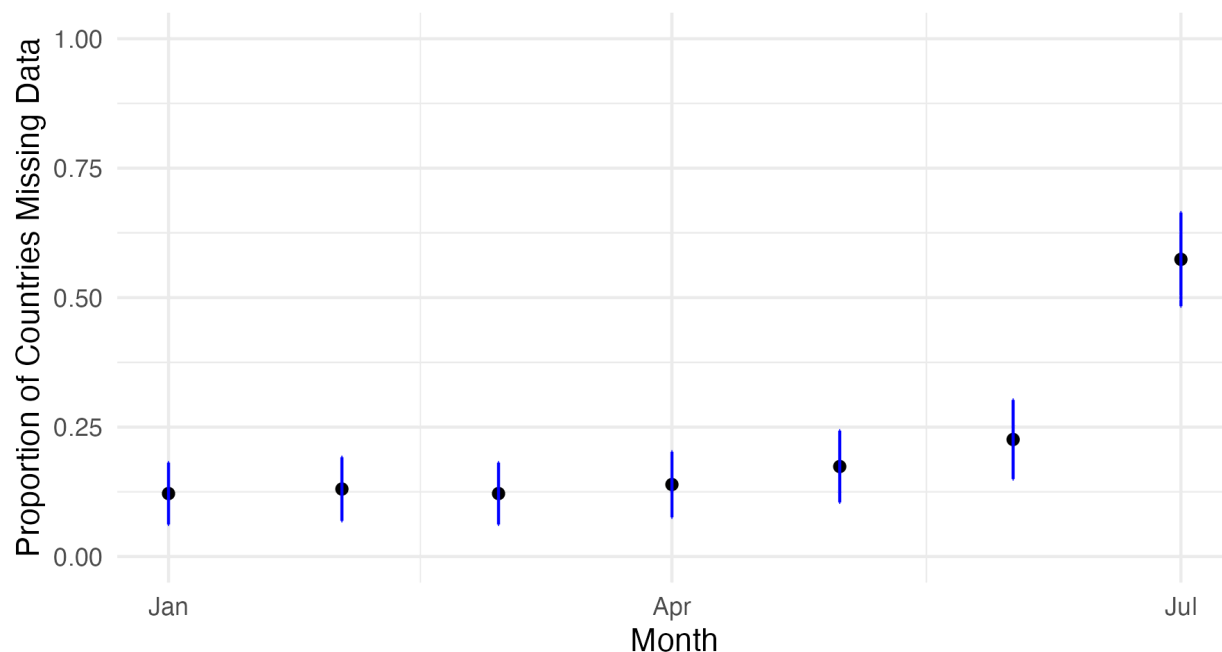
Model	Se	Sp	Accuracy
ML	0.822	0.692	0.740
Baseline	0.730	0.711	0.714

Future Projections

We used the fitted models to create future predictions with a prediction window starting in August 2025 through April 2026. It should be noted that model selection and validation were performed so that each designated measles season fell within the prediction window as determined by our cutoff dates. With these predictions, however, we are using the same August through April window for all countries, which may not encompass the entire measles season for each country. Additionally, due to delays in reporting, several countries had missing data in the months leading up to the August prediction window (Figure 8) with approximately 23% missing in June and 57% in July which will likely impact the robustness of their predictions.

Figure 8.

Proportion of countries with missing reported data on measles cases by month for 2025.



Of the 115 countries that we modeled, 14 countries were predicted in Tier A, 8 in Tier B, and 12 in Tier C (Table 14). Several countries (highlighted in blue) in each tier were also listed as targets for SIAs in 2025 on the Measles and Rubella Partnership website.⁸

Table 14.

Predicted country tiers based on future projections of reported measles cases for November 2024 - June 2025

Prioritization Tier	Countries
A - Higher (7 or more outbreak months)	Afghanistan, Georgia, Equatorial Guinea, Israel, Kazakhstan, Kyrgyzstan, Laos, Liberia, Montenegro, Mongolia, Romania, Russia, Sudan, Somalia, Serbia, Ukraine, Yemen
B - Moderate (4 – 6 outbreak months)	Bosnia and Herzegovina, Democratic Republic of the Congo, Solomon Islands, South Sudan

⁸ Measles and Rubella Partnership. 2024 Measles SIA Calendar. <https://measlesrubellapartnership.org/resources/sia-schedule/2024-measles-sia-calendar/>. Accessed 11 December 2024

C - Lower (1 – 3 outbreak months)	Burundi, Benin, Ethiopia, Gabon, Ghana, Indonesia, Lithuania, Moldova, Mali, Nigeria, Netherlands, Chad, Tajikistan, Uzbekistan
None (No outbreak months)	Angola, Albania, United Arab Emirates, Armenia, Austria, Azerbaijan, Belgium, Burkina Faso, Bangladesh, Bulgaria, Bahrain, Belarus, Brazil, Bhutan, Central African Republic, Switzerland, China, Côte d'Ivoire, Cameroon, Republic of the Congo, Comoros, Czechia, Germany, Djibouti, Algeria, Eritrea, Spain, Estonia, Fiji, France, United Kingdom, Guinea, Gambia, Guinea-Bissau, Greece, Croatia, India, Ireland, Iraq, Iceland, Italy, Kenya, Cambodia, Lebanon, Libya, Sri Lanka, Luxembourg, Madagascar, Maldives, North Macedonia, Malta, Myanmar, Mozambique, Mauritania, Malaysia, Namibia, Niger, Nepal, New Zealand, Oman, Pakistan, Philippines, Papua New Guinea, Qatar, Saudi Arabia, Senegal, Singapore, Sierra Leone, Slovakia, Syria, Togo, Thailand, Timor-Leste, Tunisia, Turkey, Uganda, Venezuela, Vietnam, Zambia, Zimbabwe
Not modeled due to insufficient data or meeting exclusion criteria	Andorra, Argentina, Antigua and Barbuda, Australia, Bahamas, Belize, Bolivia, Barbados, Bahrain, Botswana, Canada, Chile, Cook Islands, Colombia, Cape Verde, Costa Rica, Cuba, Cyprus, Dominica, Denmark, Dominican Republic, Ecuador, Egypt, Finland, Federated States of Micronesia, Grenada, Guatemala, Guyana, Honduras, Haiti, Hungary, Iran, Jamaica, Jordan, Japan, Kiribati, Saint Kitts and Nevis, South Korea, Kuwait, Saint Lucia, Lesotho, Latvia, Morocco, Monaco, Mexico, Marshall Islands, Malawi, Nicaragua, Niue, Norway, Nauru, Panama, Peru, Palau, Poland, North Korea, Portugal, Paraguay, Rwanda, El Salvador, San Marino, Sao Tome and Principe, Suriname, Slovenia, Sweden, Eswatini, Seychelles, Turkmenistan, Tonga, Trinidad and Tobago, Tuvalu, Tanzania, Uruguay, United States, Saint Vincent and the Grenadines, Vanuatu, Samoa, South Africa

Discussion and Commentary

The Effect of Seasonality and Data Quality

Throughout this study, we have explored rich, multidimensional sweeps of models, hyperparameters, and secondary methods to improve predictive performance. While we have developed a strong method for searching this space and identifying candidate model-hyperparameter pairings, characterizing the underlying dynamics that drive or compromise country-specific performance metrics remains challenging.

Data sets may be degraded in unique ways, each with significant implications for both local interpretation and the generalizability of models. For example, a long stretch of missing data during periods of low case burden may reflect an intentionally down-regulated surveillance infrastructure. By contrast, single missing months during an epidemic may signal reporting failures or strained response systems.

To explore these dynamics, we implemented the following four data quality metrics (Table 15).

Table 15.

Data quality metrics

Metric	Definition
Coverage	The percentage of expected values present for a given data range.
Mean segment length	The average length of uninterrupted records.
Proportion adjacent to gaps	The proportion of cumulative measured signal within one month of a missing record.
Permutation entropy	A complexity measure of the predictability of a system over time.

We likewise observed that while some model configurations performed well in countries with highly seasonal, endemic measles, these same configurations did not necessarily generalize to countries with singular or otherwise sporadic outbreaks. Defining seasonality itself proved challenging, as features such as annual peak timing, peak amplitude, or curve shape often vary within an otherwise endemic country across years.

To address this, we defined a range of **seasonality metrics** (Table 16) to describe the relative extent of annual trends in country-level measles incidence.

Table 16.
Seasonality metrics

Metric	Definition
Annual autocorrelation	Strength of annual autocorrelation within a curve.
Peak interval consistency	Variation in durations of intervals between peaks.
Peak power ratio	Ratio of peak power to total power in a Fast Fourier Transform (FFT) spectrum.
Seasonal consistency	Consistency of seasonal component variation across years.
Combined seasonality score	Mean of the normalized component seasonality scores.

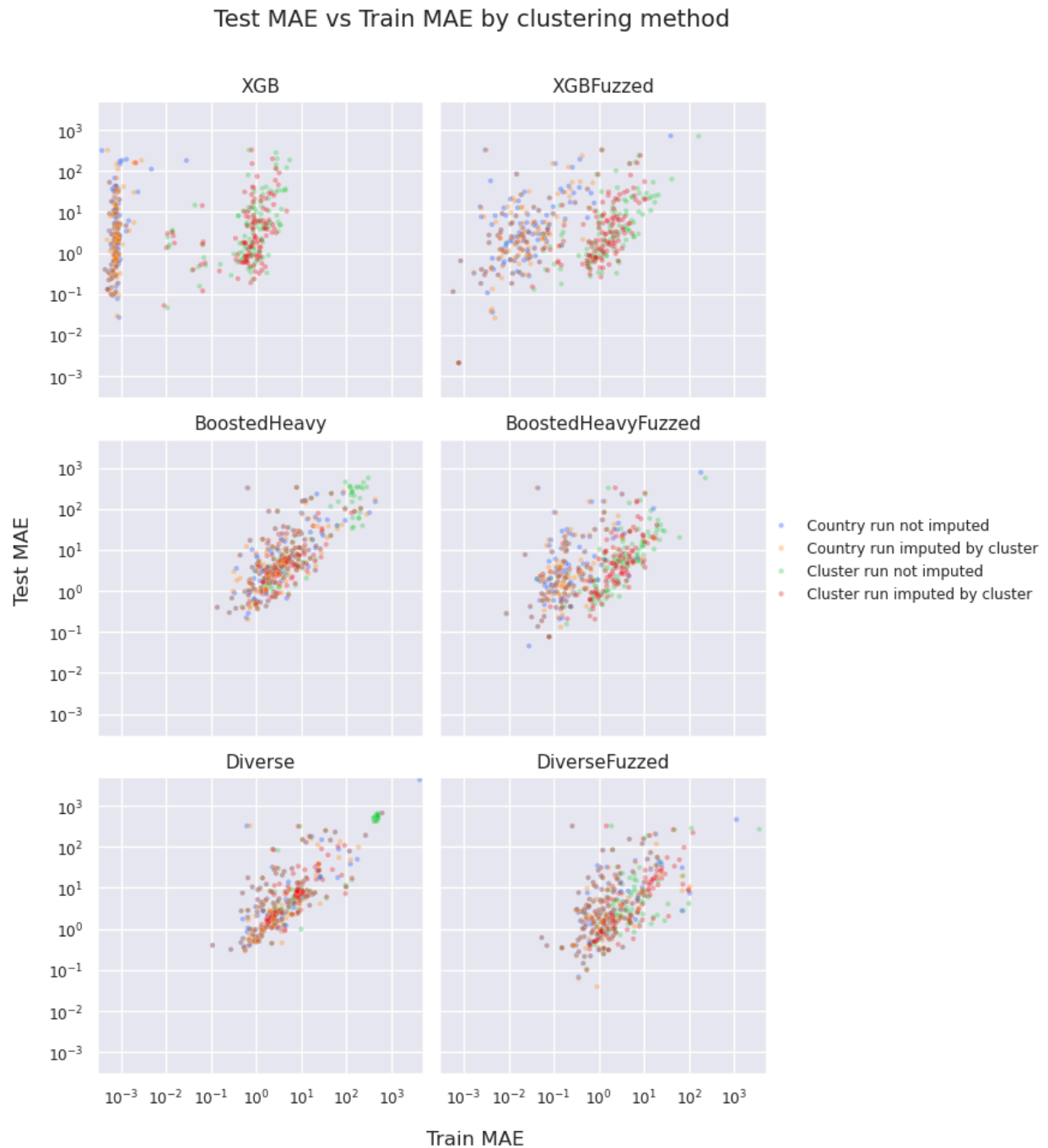
With these input data metrics defined, we conducted a sweep of model runs to identify general trends in performance as a function of model selection, fuzzing strategy, and data imputation strategies. Selected models included XGBoost, the Diverse ensemble, and the BoostedHeavy ensemble. Each model was run with and without fuzzing. Predictors included MCV1 coverage, mean precipitation, mean temperature, births per 1,000, and passengers from ISO3 reporting cases.

Data quality metrics (Table 15) were calculated using raw incidence data for each country. Likewise, the combined seasonality score was derived from country-level incidence data. Two additional source data sets were then created by imputing missing independent variables via KNN imputation: (i) imputation of singular missing months within countries, and (ii) imputation of all missing months within country clusters.

Each country was then evaluated individually, within its cluster (via global-local), and within a global cluster of all countries, under all combinations of predictor imputation strategy, model type, and fuzzing configuration. Parameters were matched to those used in validation runs. Incidence predictions in the withheld test set were evaluated by their test error (MAE) and positive predictive power, the latter based on a binary outbreak threshold of two cases per million.

Figure 9.

Relationship between test fit performance vs training fit performance by clustering method.



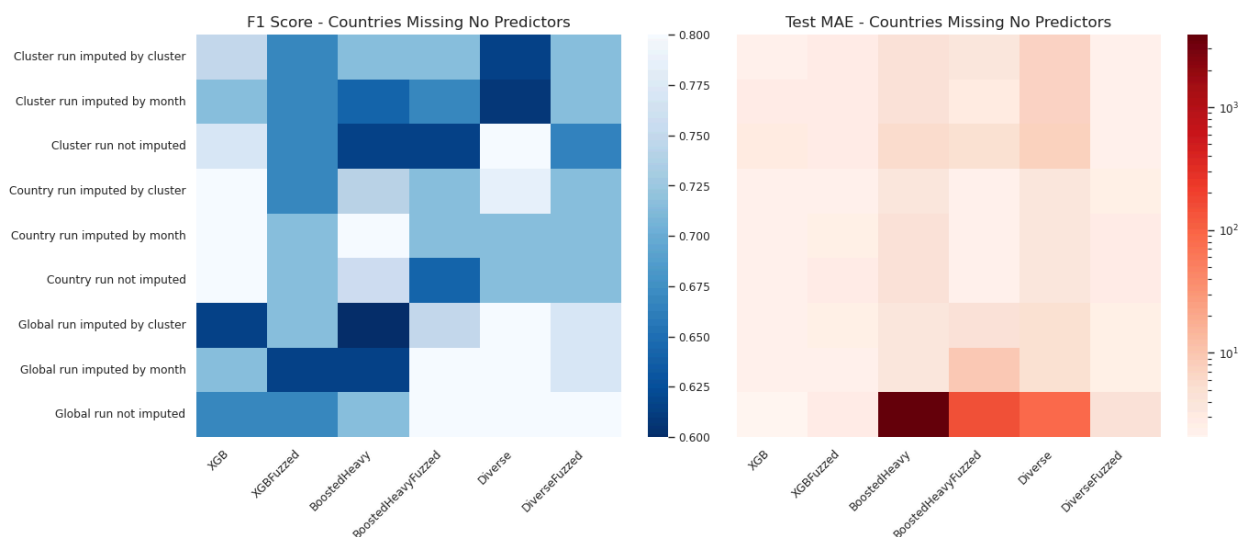
Although we hoped this would reveal informative partitions of model performance by seasonality or data quality, no consistent utility niches of model configurations emerged. Instead, only broad

performance differences were observed (ADD APPENDICES). Country-level runs generally outperformed both cluster- and global-level runs while also appearing to benefit from missing-data imputation.

The most consistent trend observed was a power-law relationship between test-set MAE and training-set MAE (Figure 9). This raised the question of how much of the surplus error in the test set, relative to the training set, could be attributed to differences in data quality and seasonality. To examine this further, we selected countries with observed outbreaks and complete predictor observations ($n = 119$) and evaluated them by median F1 score and test MAE (Figure 10). From this subset, we identified XGB country-level runs without fuzzing, using the cluster-imputed data set, as the overall best-performing configuration.

Figure 10.

Performance metric comparison by model and imputation strategy.



Lastly, we conducted an exploratory analysis to evaluate whether seasonality and data-quality metrics could serve as reasonable predictors of model error. The motivation here was not to optimize performance of any particular forecasting model, but rather to assess whether metadata describing the underlying time series could be used to anticipate how well a model is likely to perform. To test this, we constructed regression models that used only country-level measures of seasonality and data quality within the incidence data as inputs, with the observed test error (MAE) as the prediction target. Because test error values were highly skewed and contained substantial outliers, the target distribution was log-transformed prior to modeling, and performance was evaluated using a range of complementary error metrics (Table 17).

Table 17.

Error prediction performance by model

model	R2	RMSE	MAE	MedianAE
BoostedHeavy	-4.45E-02	3.14E+03	6.54E+02	1.02E+01
Diverse	-2.18E-03	7.25E+02	6.70E+01	4.31E+00
XGB	7.92E-01	1.97E+01	5.03E+00	4.92E-01

The results showed a mixed picture. For the BoostedHeavy and Diverse projections, the regression models explained little to no variance in the observed error (R^2 values near zero). At the same time, their median absolute errors were quite low, indicating that the models could often predict error reasonably well under typical conditions, but struggled in the presence of rare albeit extreme cases. This points to the challenge of learning a stable mapping between data-quality metrics and test error when the latter is dominated by heavy-tailed behavior.

Figure 11a.

Data metric correlation to test set error

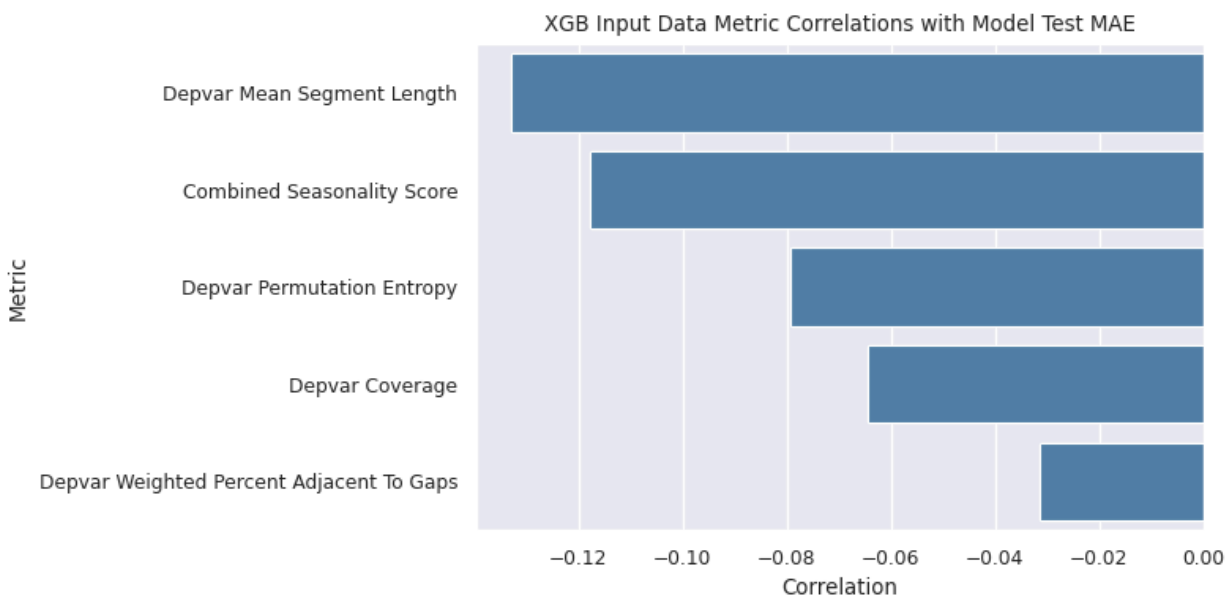
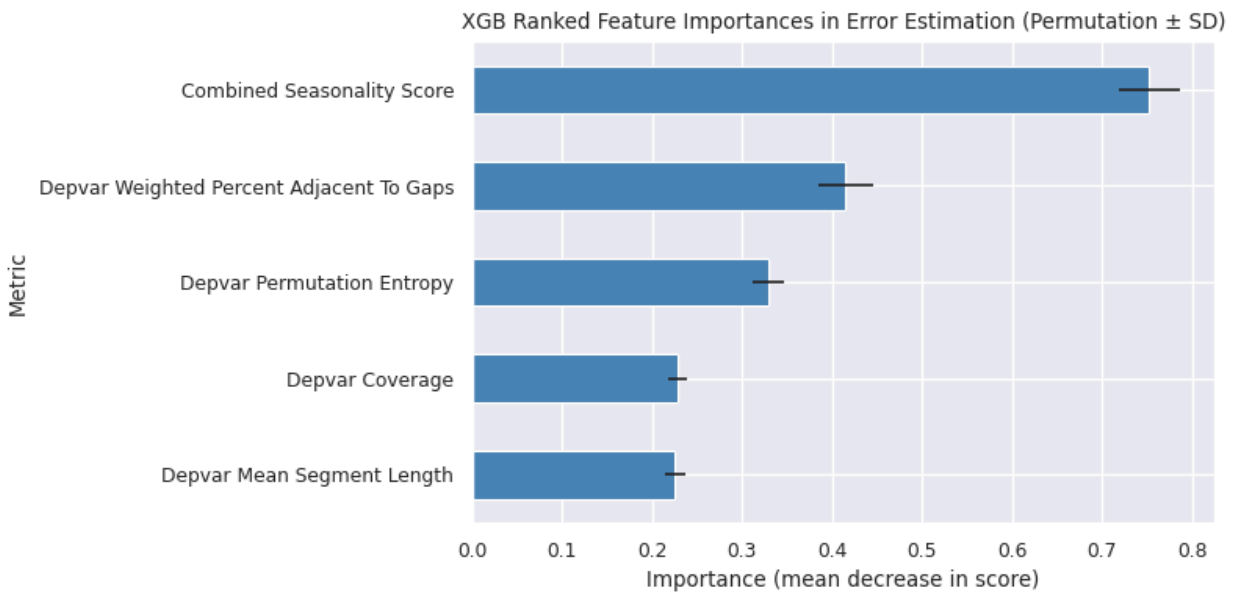


Figure 11b.

Data metric feature importance in error estimation



By contrast, the XGB projections showed a much clearer relationship: seasonality and quality metrics were able to explain roughly 79% of the variance in test error, with low median error and strong robustness to outliers. This suggests that, at least for certain model classes, metadata about the time series carries meaningful information about predictive difficulty.

Taken together, these findings indicate that the feasibility of estimating model error from seasonality and data-quality metrics depends strongly on the model in question and the stability of its error distribution. However, while the approach does not appear universally reliable, the XGB results demonstrate that such metadata can, under the right conditions, be informative about expected error bounds in projected epicurves. This opens a potential path for incorporating diagnostic features of the data itself into future model selection and evaluation frameworks.

Conclusion

Leveraging a country-specific modeling approach and a staged grid search for model fitting and predictor selection, the AIM-F Measles Forecasting Model achieves a high level of accuracy in identifying outbreak months. This strategy enables the model to tailor predictive methods and inputs to the unique epidemiological and data characteristics of each country, leading to robust performance in forecasting measles risk. The model's ability to prioritize countries for Supplementary Immunization Activities (SIA) based on predicted outbreak months offers a valuable tool for public health decision-making. Future projections indicate that the model can effectively guide vaccine deployment efforts, ultimately contributing to the reduction of measles-related mortality and morbidity globally.

Legal Notice & Disclaimers

No model of events is, or could be, an exact representation of reality, and models rely on various data, methodologies, and assumptions, some of which are subjective and/or subject to uncertainty, and may differ materially from models produced by other organizations. Consequently, there can be no assurance that the Ginkgo Biosecurity model(s) described herein represents an accurate estimation of the future.

Appendix

Figure A1.

Measles case data presence by country and date. Countries are grouped by World Bank income classification and sorted within groups by proportion of dates present. The color of the country ISO3 code corresponds to the prediction accuracy during validation assessment.

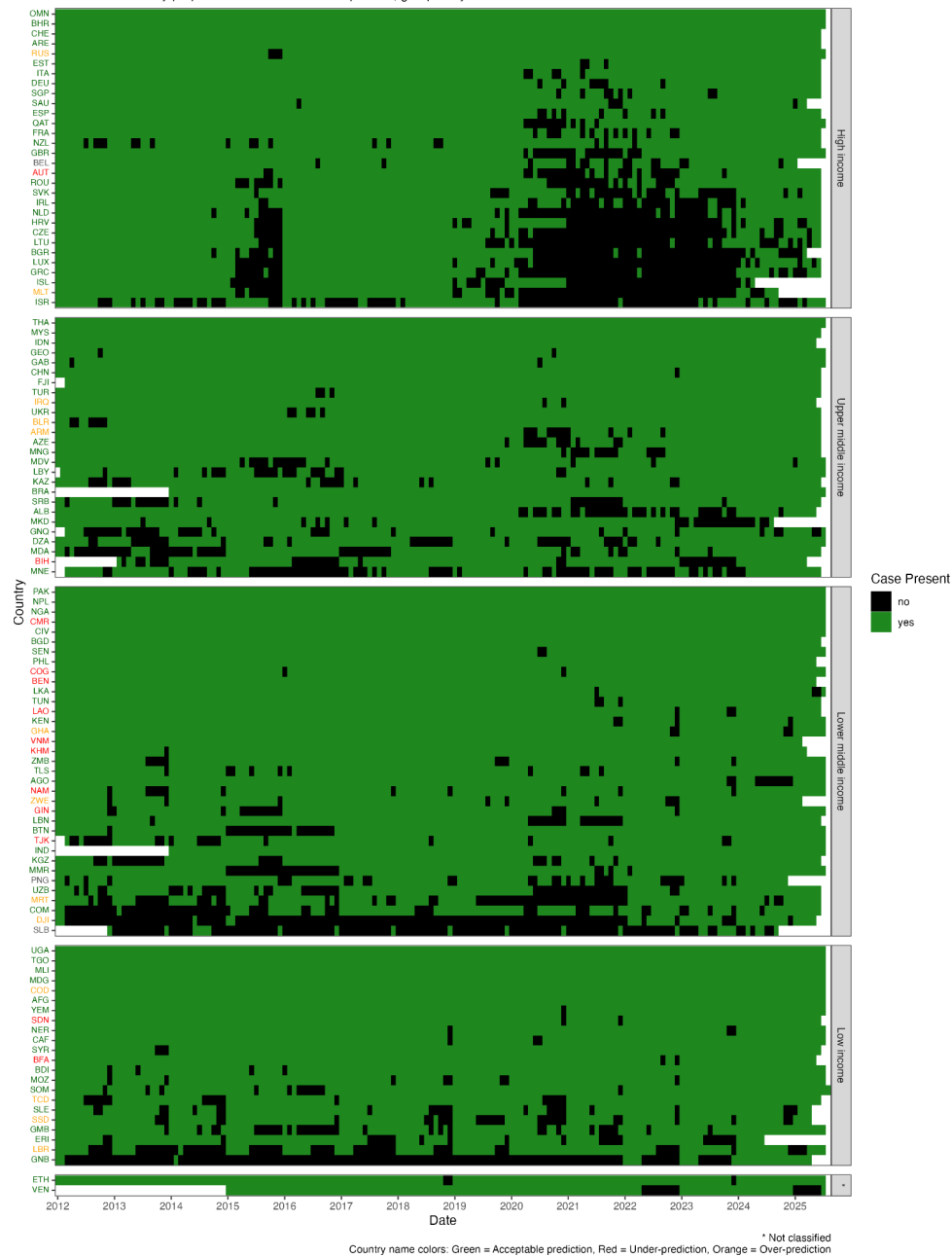


Table A1.

A list of countries assigned to each cluster by time-series hierarchical clustering analysis. These clusters were used in global-local model fitting.

Cluster	Countries
1	Afghanistan, Angola, Benin, Burundi, Cabo Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Equatorial Guinea, Ethiopia, Gabon, Gambia, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Nigeria, Papua New Guinea, Rwanda, Sao Tome and Principe, Seychelles, Somalia, South Sudan, Togo, Yemen
2	Algeria, Antigua and Barbuda, Bahamas, Bangladesh, Barbados, Belize, Bhutan, Bolivia, Botswana, Brazil, Brunei, Burkina Faso, Cambodia, Colombia, Cook Islands, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Eritrea, Eswatini, Fiji, Ghana, Grenada, Guatemala, Guyana, Haiti, Honduras, Iraq, Jamaica, Jordan, Kenya, Kiribati, Laos, Lesotho, Libya, Malawi, Malaysia, Maldives, Marshall Islands, Micronesia, Mozambique, Myanmar, Namibia, Nauru, Nepal, Nicaragua, Niue, Pakistan, Palau, Panama, Paraguay, Peru, Philippines, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, Senegal, Sierra Leone, Singapore, Solomon Islands, South Africa, Sri Lanka, Sudan, Suriname, Timor-Leste, Tonga, Trinidad and Tobago, Tuvalu, Uganda, Tanzania, Vanuatu, Zambia, Zimbabwe
3	China, Italy, Republic of Korea, United Kingdom, United States of America
4	Australia, France, Germany, India, Japan, Russian Federation, Spain, Türkiye, Vietnam
5	Albania, Andorra, Argentina, Armenia, Austria, Azerbaijan, Bahrain, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Canada, Chile, Croatia, Cyprus, Czechia, Democratic People's Republic of Korea, Denmark, Djibouti, Estonia, Finland, Georgia, Greece, Hungary, Iceland, Iran, Ireland, Israel, Kazakhstan, Kuwait, Kyrgyzstan, Latvia, Lebanon, Lithuania, Luxembourg, Malta, Mexico, Monaco, Mongolia, Montenegro, Morocco, Netherlands, New Zealand, North Macedonia, Norway, Oman, Poland, Portugal, Qatar, Republic of Moldova, Romania, San Marino, Serbia, Slovakia, Slovenia, Sweden, Switzerland, Tajikistan, Tunisia, Turkmenistan, Uruguay, Uzbekistan
6	Indonesia
7	Saudi Arabia, Thailand, United Arab Emirates
8	Syria



9	Ukraine
10	Venezuela

Table A2.

Preprocessor methods applied by input variable

Method(s)	Applied variables
pass_unchanged	unicef_region, cluster
january_only interpolate_via_linear	population_jan
july_only interpolate_via_linear	population_july
january_only divide_by_12 interpolate_via_linear	birth_per_1k, births, migrations, migrations_per_1k
january_only divide_by_12 interpolate_via_linear forward_fill back_fill	MCV1, MCV2
back_truncate interpolate_via_cubic zero_min	cuml_cases, mean_precip_mm_per_day, total_precip_mm_per_day
back_truncate yn_to_int	outbreak_20_per_M, outbreak_2_per_M, outbreak_5_per_M, outbreak_20_cuml_per_M
back_truncate	mnths_since_outbreak_20_per_M, num_outbreak_20_per_M, mnths_since_outbreak_2_per_M, num_outbreak_2_per_M, mnths_since_outbreak_20_cuml_per_M, num_outbreak_20_cuml_per_M, incoming_air_passengers, outgoing_air_passengers
back_truncate interpolate_via_cubic	mean_max_temp
back_truncate forward_truncate interpolate_via_cubic zero_min	cases, cases_1M, cuml_cases_1M
zero_fill	passengers_from_iso3_reporting_cases, passengers_to_iso3_reporting_cases

Table A3.

A list of candidate classifier models entered into the simulated annealing model selection algorithm.

Classifier model	Documentation
------------------	---------------



XGBRegressor	https://xgboost.readthedocs.io/en/stable/python/
Gradient Boosting	https://scikit-learn.org/stable/auto_examples/ensemble/plot_gradient_boosting_regression.html
CatBoost	https://catboost.ai/docs/en/concepts/python-reference_catboostregressor.html
Random Forest	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
Bagging Regressor	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html

Table A4.

Characteristics of the selected best model for each of the 115 countries with the observed tier and predicted tier for the validation phase

Country	Predictor set	ML model	Obs Tier	Sim Tier	Prediction
Afghanistan	mnths_since_outbreak_20_cuml_per_M, prop_prev_rolling_60_mnths_outbreak_5M, cases_1M_36z	CatBoost	A	A	acceptable
Angola	mnths_since_outbreak_20_per_M, cases_1M_36z	Bagging regressor	None	None	acceptable
Albania	mnths_since_SIA	XGBRegressor	None	None	acceptable
United Arab Emirates	mnths_since_SIA, incoming_air_passengers, prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	None	None	acceptable
Armenia	migrations_per_1k, prop_prev_rolling_60_mnths_outbreak_5M	Bagging regressor	None	A	over
Austria	prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	C	None	under
Azerbaijan	birth_per_1k, incoming_air_passengers, prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	None	C	acceptable
Burundi	mnths_since_outbreak_20_cuml_per_M, prop_prev_rolling_24_mnths_outbreak_5M	gradient boosting	None	C	acceptable
Belgium	cases_1M_12z	Bagging regressor			
Benin	migrations_per_1k, mnths_since_outbreak_20_cuml_per_M	Bagging regressor	B	C	under
Burkina Faso	mnths_since_outbreak_20_cuml_per_M, mnths_since_SIA, incoming_air_passengers	gradient boosting	B	C	under



Bangladesh	cases_1M_12z	gradient boosting	None	None	acceptable
Bulgaria	MCV2, mnths_since_SIA, prop_prev_rolling_60_mnths_outbreak_5M	CatBoost	None	None	acceptable
Bahrain	birth_per_1k	XGBRegressor	None	None	acceptable
Bosnia & Herzegovina	MCV2, incoming_air_passengers, prop_prev_rolling_60_mnths_outbreak_5M	gradient boosting	B	C	under
Belarus	outgoing_air_passengers, prop_prev_rolling_60_mnths_outbreak_5M	CatBoost	None	B	over
Brazil	mnths_since_outbreak_20_per_M, prop_prev_rolling_24_mnths_outbreak_5M	Bagging regressor	None	None	acceptable
Bhutan	birth_per_1k, prop_prev_rolling_12_mnths_outbreak_5M	XGBRegressor	None	None	acceptable
Central African Republic	mnths_since_outbreak_20_per_M	Random Forest	None	C	acceptable
Switzerland	prop_prev_rolling_60_mnths_outbreak_5M	Random Forest	None	None	acceptable
China	birth_per_1k, prop_prev_rolling_12_mnths_outbreak_5M	CatBoost	None	None	acceptable
Côte d'Ivoire	prop_prev_rolling_24_mnths_outbreak_5M	gradient boosting	C	C	acceptable
Cameroon	mnths_since_outbreak_20_cuml_per_M, incoming_air_passengers, cases_1M_12z	CatBoost	C	None	under
Congo - Kinshasa	prop_prev_rolling_12_mnths_outbreak_5M	Random Forest	None	A	over
Congo - Brazzaville	mnths_since_SIA, prop_prev_rolling_12_mnths_outbreak_5M, cases_1M_12z	XGBRegressor	C	None	under
Comoros	mnths_since_outbreak_20_per_M	gradient boosting	None	None	acceptable
Czechia	MCV2, cases_1M_12z	Random Forest	None	None	acceptable
Germany	migrations_per_1k, incoming_air_passengers	Bagging regressor	None	None	acceptable
Djibouti	mnths_since_outbreak_20_per_M, mnths_since_SIA, prop_prev_rolling_60_mnths_outbreak_5M	gradient boosting	None	A	over
Algeria	MCV2, prop_prev_rolling_12_mnths_outbreak_5M	Bagging regressor	None	None	acceptable
Eritrea	mnths_since_SIA, outgoing_air_passengers	CatBoost	None	None	acceptable



Spain	outgoing_air_passengers	CatBoost	None	None	acceptable
Estonia	cases_1M_60z	Random Forest	None	None	acceptable
Ethiopia	mnths_since_outbreak_20_cuml_per_M, prop_prev_rolling_24_mnths_outbreak_5M, cases_1M_12z	gradient boosting	A	A	acceptable
Fiji	outgoing_air_passengers	XGBRegressor	None	None	acceptable
France	prop_prev_rolling_12_mnths_outbreak_5M, cases_1M_36z	gradient boosting	None	None	acceptable
Gabon	prop_prev_rolling_60_mnths_outbreak_5M	Bagging regressor	None	C	acceptable
United Kingdom	cases_1M_12z	gradient boosting	None	None	acceptable
Georgia	MCV2, mnths_since_outbreak_20_per_M, prop_prev_rolling_24_mnths_outbreak_5M	CatBoost	A	A	acceptable
Ghana	MCV1, mnths_since_outbreak_20_cuml_per_M, incoming_air_passengers	XGBRegressor	C	A	over
Guinea	mnths_since_SIA, prop_prev_rolling_12_mnths_outbreak_5M	Random Forest	B	C	under
Gambia	mnths_since_SIA, incoming_air_passengers	Bagging regressor	None	None	acceptable
Guinea-Bissau	prop_prev_rolling_60_mnths_outbreak_5M	Random Forest	None	None	acceptable
Equatorial Guinea	prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	None	C	acceptable
Greece	migrations_per_1k	XGBRegressor	None	None	acceptable
Croatia	migrations_per_1k	XGBRegressor	None	None	acceptable
Indonesia	outgoing_air_passengers, prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	None	None	acceptable
India	MCV2, prop_prev_rolling_60_mnths_outbreak_5M, cases_1M_12z	XGBRegressor	None	None	acceptable
Ireland	mnths_since_SIA, prop_prev_rolling_60_mnths_outbreak_5M	gradient boosting	None	None	acceptable
Iraq	migrations_per_1k, mnths_since_SIA, prop_prev_rolling_12_mnths_outbreak_5M	gradient boosting	None	A	over
Iceland	MCV1	Random Forest	None	None	acceptable
Israel	migrations_per_1k, MCV1, prop_prev_rolling_12_mnths_outbreak_5M	gradient boosting	None	None	acceptable



Italy	MCV2, mnths_since_outbreak_20_cuml_per_M, prop_prev_rolling_12_mnths_outbreak_5M	XGBRegressor	None	None	acceptable
Kazakhstan	birth_per_1k, outgoing_air_passengers, prop_prev_rolling_12_mnths_outbreak_5M	gradient boosting	B	A	acceptable
Kenya	incoming_air_passengers, cases_1M_60z	gradient boosting	None	None	acceptable
Kyrgyzstan	mnths_since_outbreak_20_per_M	Random Forest	A	A	acceptable
Cambodia	MCV2	Random Forest	B	None	under
Laos	prop_prev_rolling_12_mnths_outbreak_5M	Bagging regressor	C	None	under
Lebanon	mnths_since_outbreak_20_per_M, cases_1M_12z	gradient boosting	None	None	acceptable
Liberia	mnths_since_outbreak_20_per_M, mnths_since_SIA, prop_prev_rolling_24_mnths_outbreak_5M	Bagging regressor	C	A	over
Libya	mnths_since_outbreak_20_per_M, prop_prev_rolling_12_mnths_outbreak_5M, cases_1M_36z	CatBoost	None	None	acceptable
Sri Lanka	mnths_since_outbreak_20_cuml_per_M, prop_prev_rolling_12_mnths_outbreak_5M	CatBoost	None	None	acceptable
Lithuania	incoming_air_passengers	CatBoost	None	C	acceptable
Luxembourg	birth_per_1k, MCV2	Bagging regressor	None	None	acceptable
Moldova	prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	C	B	acceptable
Madagascar	mnths_since_SIA, prop_prev_rolling_12_mnths_outbreak_5M	Bagging regressor	None	None	acceptable
Maldives	MCV2	CatBoost	None	None	acceptable
North Macedonia	migrations_per_1k, MCV2, mnths_since_outbreak_20_cuml_per_M	CatBoost	None	None	acceptable
Mali	mnths_since_outbreak_20_cuml_per_M, outgoing_air_passengers	XGBRegressor	None	C	acceptable
Malta	migrations_per_1k, prop_prev_rolling_60_mnths_outbreak_5M	CatBoost	None	A	over
Myanmar (Burma)	birth_per_1k, outgoing_air_passengers	gradient boosting	None	None	acceptable
Montenegro	migrations_per_1k, prop_prev_rolling_12_mnths_outbreak_5M	Bagging regressor	None	None	acceptable
Mongolia	prop_prev_rolling_24_mnths_outbreak_5M	Bagging regressor	C	C	acceptable



Mozambique	outgoing_air_passengers, cases_1M_36z	XGBRegressor	None	None	acceptable
Mauritania	birth_per_1k, incoming_air_passengers	Bagging regressor	None	B	over
Malaysia	mnths_since_SIA, cases_1M_12z	CatBoost	C	B	acceptable
Namibia	birth_per_1k, prop_prev_rolling_12_mnths_outbreak_5M, cases_1M_12z	CatBoost	C	None	under
Niger	incoming_air_passengers	gradient boosting	B	B	acceptable
Nigeria	mnths_since_outbreak_20_per_M, outgoing_air_passengers, incoming_air_passengers	CatBoost	None	C	acceptable
Netherlands	migrations_per_1k, prop_prev_rolling_24_mnths_outbreak_5M	Random Forest	None	None	acceptable
Nepal	incoming_air_passengers, cases_1M_12z	XGBRegressor	None	None	acceptable
New Zealand	migrations_per_1k, mnths_since_outbreak_20_per_M, mnths_since_SIA	Bagging regressor	None	None	acceptable
Oman	birth_per_1k	CatBoost	None	None	acceptable
Pakistan	birth_per_1k, mnths_since_outbreak_20_cuml_per_M, incoming_air_passengers	gradient boosting	B	A	acceptable
Philippines	mnths_since_SIA, prop_prev_rolling_24_mnths_outbreak_5M, cases_1M_12z	Bagging regressor	None	None	acceptable
Papua New Guinea	mnths_since_SIA, prop_prev_rolling_12_mnths_outbreak_5M	gradient boosting			
Qatar	birth_per_1k, MCV2	Bagging regressor	None	None	acceptable
Romania	migrations_per_1k, MCV1, mnths_since_outbreak_20_cuml_per_M	XGBRegressor	A	A	acceptable
Russia	mnths_since_outbreak_20_cuml_per_M, prop_prev_rolling_12_mnths_outbreak_5M	gradient boosting	None	A	over
Saudi Arabia	mnths_since_SIA, incoming_air_passengers	XGBRegressor	None	C	acceptable
Sudan	migrations_per_1k, MCV2, cases_1M_12z	gradient boosting	C	None	under
Senegal	migrations_per_1k, prop_prev_rolling_60_mnths_outbreak_5M	gradient boosting	None	None	acceptable
Singapore	MCV2	XGBRegressor	None	None	acceptable



Solomon Islands	prop_prev_rolling_24_mnths_outbreak_5M	Bagging regressor			
Sierra Leone	cases_1M_12z	Random Forest	None	None	acceptable
Somalia	migrations_per_1k, mnths_since_SIA, prop_prev_rolling_60_mnths_outbreak_5M	gradient boosting	B	A	acceptable
Serbia	birth_per_1k, MCV1	CatBoost	B	A	acceptable
South Sudan	mnths_since_outbreak_20_cuml_per_M, outgoing_air_passengers	Bagging regressor	None	A	over
Slovakia	mnths_since_outbreak_20_per_M, prop_prev_rolling_12_mnths_outbreak_5M	Bagging regressor	None	None	acceptable
Syria	mnths_since_SIA, cases_1M_36z	CatBoost	None	None	acceptable
Chad	MCV1, prop_prev_rolling_24_mnths_outbreak_5M	Random Forest	None	B	over
Togo	mnths_since_outbreak_20_cuml_per_M	CatBoost	C	C	acceptable
Thailand	MCV2, cases_1M_12z	Bagging regressor	B	A	acceptable
Tajikistan	outgoing_air_passengers, incoming_air_passengers, cases_1M_36z	Bagging regressor	C	None	under
Timor-Leste	cases_1M_12z	gradient boosting	None	None	acceptable
Tunisia	prop_prev_rolling_24_mnths_outbreak_5M	Bagging regressor	None	None	acceptable
Turkey	MCV1, prop_prev_rolling_12_mnths_outbreak_5M, cases_1M_60z	CatBoost	None	None	acceptable
Uganda	migrations_per_1k, mnths_since_SIA, cases_1M_12z	CatBoost	None	None	acceptable
Ukraine	birth_per_1k, mnths_since_SIA, prop_prev_rolling_60_mnths_outbreak_5M	XGBRegressor	None	None	acceptable
Uzbekistan	mnths_since_outbreak_20_cuml_per_M, mnths_since_SIA, prop_prev_rolling_24_mnths_outbreak_5M	Bagging regressor	C	B	acceptable
Venezuela	birth_per_1k, migrations_per_1k, mnths_since_SIA	XGBRegressor	None	None	acceptable
Vietnam	birth_per_1k, mnths_since_outbreak_20_cuml_per_M, cases_1M_36z	gradient boosting	C	None	under
Yemen	migrations_per_1k, mnths_since_SIA, prop_prev_rolling_12_mnths_outbreak_5M	Random Forest	A	A	acceptable
Zambia	outgoing_air_passengers, cases_1M_60z	gradient boosting	None	None	acceptable
Zimbabwe	prop_prev_rolling_24_mnths_outbreak_5M	XGBRegressor	None	B	over

Figure A2.