

Model Comparison Pipeline

Output 7

Dec 9, 2024

The Model Comparison Pipeline is an R markdown script that can be used to create an R Shiny dashboard enabling the comparison of candidate measles forecast models. Model methods and input parameters can be compared based on MSE, sensitivity, specificity, and by visual inspection of plots and statistical summary tables.

Github location

The model comparison R Shiny can be found at the following Github repository:

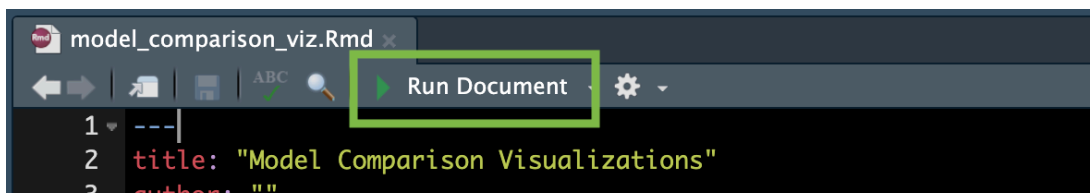
https://github.com/concentricbyginkgo/measles_forecasting. The Shiny script can be found in `model_comparison_pipeline/model_comparison_viz.Rmd` with accompanying sample input files:

- `model_comparison_pipeline/country_output/NGA.csv`
- `model_comparison_pipeline/country_output/UKR.csv`
- `model_comparison_pipeline/summary_output/summaryTable.csv`

Lines 38-40 in `model_comparison_viz.Rmd` specify the paths to these files, and may need to be updated to the user's directory structure.

Launching the Shiny app

The Shiny dashboard can be launched by either clicking the 'Run' icon (highlighted in green in the image below) in the Rstudio interface, or executing the following command in the R console `rmarkdown::render("<path to document>/model_comparison_viz.Rmd")`





Data input files

The Model Comparison Shiny requires two types of files: 1) a csv file of compiled model summary statistic output and 2) a csv of time series output compiled at the country-level (one file per country).

1) **summaryTable.csv**: The compiled model summary table should have the following fields:

method: The modeling method (string)

ISO3: The ISO3 country code of the modeled country (string).

country: The name of the modeled country (string).

predictor: The model predictor variable (string).

predictor_lag: The number of months the predictor is lagged (integer).

Test_MSE: The test MSE of the model (numeric).

Train_MSE: The training MSE of the model (numeric).

model_vars: a concatenated string of all the predictor variables used to fit the model (string).

indepVars: a concatenated string of all the predictor variables and their respective lags in format <predictor>:<lag> used to fit the model (string).

index: A unique identifier for the model run; this is used to pair model summary statistics with model time series output (integer).

sensitivity_2M: Model sensitivity when an outbreak threshold of 2 cases per million is applied. Calculated using the training dataset (numeric).

specificity_2M: Model specificity when an outbreak threshold of 2 cases per million is applied. Calculated using the training dataset (numeric).

sensitivity_20M: Model sensitivity when an outbreak threshold of 20 cases per million is applied. Calculated using the training dataset (numeric).

specificity_20M: Model specificity when an outbreak threshold of 20 cases per million is applied. Calculated using the training dataset (numeric).

sensitivity_20cM: Model sensitivity when an outbreak threshold of 20 cumulative cases per million (annual) is applied. Calculated using the training dataset (numeric).

specificity_20cM: Model specificity when an outbreak threshold of 20 cumulative cases per million (annual) is applied. Calculated using the training dataset (numeric).

test_sensitivity_2M: Model sensitivity when an outbreak threshold of 2 cases per million is applied. Calculated using the testing dataset (numeric).

test_specificity_2M: Model specificity when an outbreak threshold of 2 cases per million is applied. Calculated using the testing dataset (numeric).

test_sensitivity_20M: Model sensitivity when an outbreak threshold of 20 cases per million is applied. Calculated using the testing dataset (numeric).



test_specificity_20M: Model specificity when an outbreak threshold of 20 cases per million is applied. Calculated using the testing dataset (numeric).

test_sensitivity_20cM: Model sensitivity when an outbreak threshold of 20 cumulative cases per million (annual) is applied. Calculated using the testing dataset (numeric).

test_specificity_20cM: Model specificity when an outbreak threshold of 20 cumulative cases per million (annual) is applied. Calculated using the testing dataset (numeric).

2) <country IS03>.csv: **The compiled country-level time series output should have the following fields:**

index: A unique identifier for the model run; this is used to pair model summary statistics with model time series output (integer).

ds: date in format YYYY-MM-DD.

y: The observed cases per million population at time ds (numeric).

yhat1: The predicted cases per million population at time ds (numeric).

outbreak_observed_2M: Binary (Yes/No) outcome indicating if an outbreak was observed according to 2 cases per million population threshold (string).

outbreak_predicted_2M: Binary (Yes/No) outcome indicating if an outbreak was predicted according to 2 cases per million population threshold (string).

outbreak_observed_20M: Binary (Yes/No) outcome indicating if an outbreak was observed according to 20 cases per million population threshold (string).

outbreak_predicted_20M: Binary (Yes/No) outcome indicating if an outbreak was predicted according to 20 cases per million population threshold (string).

outbreak_observed_cuml_20M: Binary (Yes/No) outcome indicating if an outbreak was observed according to 20 annual cumulative cases per million population threshold (string).

outbreak_predicted_cuml_20M: Binary (Yes/No) outcome indicating if an outbreak was predicted according to 20 annual cumulative cases per million population threshold (string).

The time series data can be large if many models are compared simultaneously, so it is pre-processed to be compiled at the country-level and read dynamically by the Model Validation Shiny to avoid reading a large file into memory.



Landing page

Measles Output Visualizations

Model specifications

Test Metrics Segment Binary Outcome Analysis Binary Outcome Plot Metrics Correlation ← Assessment metrics

Selections

Select Country (All Countries available for Binary Outcome Analysis tab only)

AFG - Afghanistan

Test Metrics Tab - Heatmap: Select model vars

MCV1_Cov_1yo, mean_precip_mm_per_day, mean_max_temp

Test Metrics & Segment Tab - Number of runs in top/bottom groups

3

Top Model vars - sorted by Test MSE

Show 5 entries Search:

| | model_vars | Test_MSE |
|---|--|----------|
| 1 | population_jan, mean_precip_mm_per_day, mean_max_temp | 18.4248 |
| 2 | birth_per_1k, mean_precip_mm_per_day, mean_max_temp | 18.9734 |
| 3 | migrations_per_1k, mean_precip_mm_per_day, mean_max_temp | 21.545 |
| 4 | MCV2, mean_precip_mm_per_day, mean_max_temp | 22.4884 |
| 5 | MCV2_Cov_RecAge, mean_precip_mm_per_day, mean_max_temp | 22.4884 |

Showing 1 to 5 of 11 entries

PREVIOUS 1 2 3 NEXT

Heatmap of method and model vars

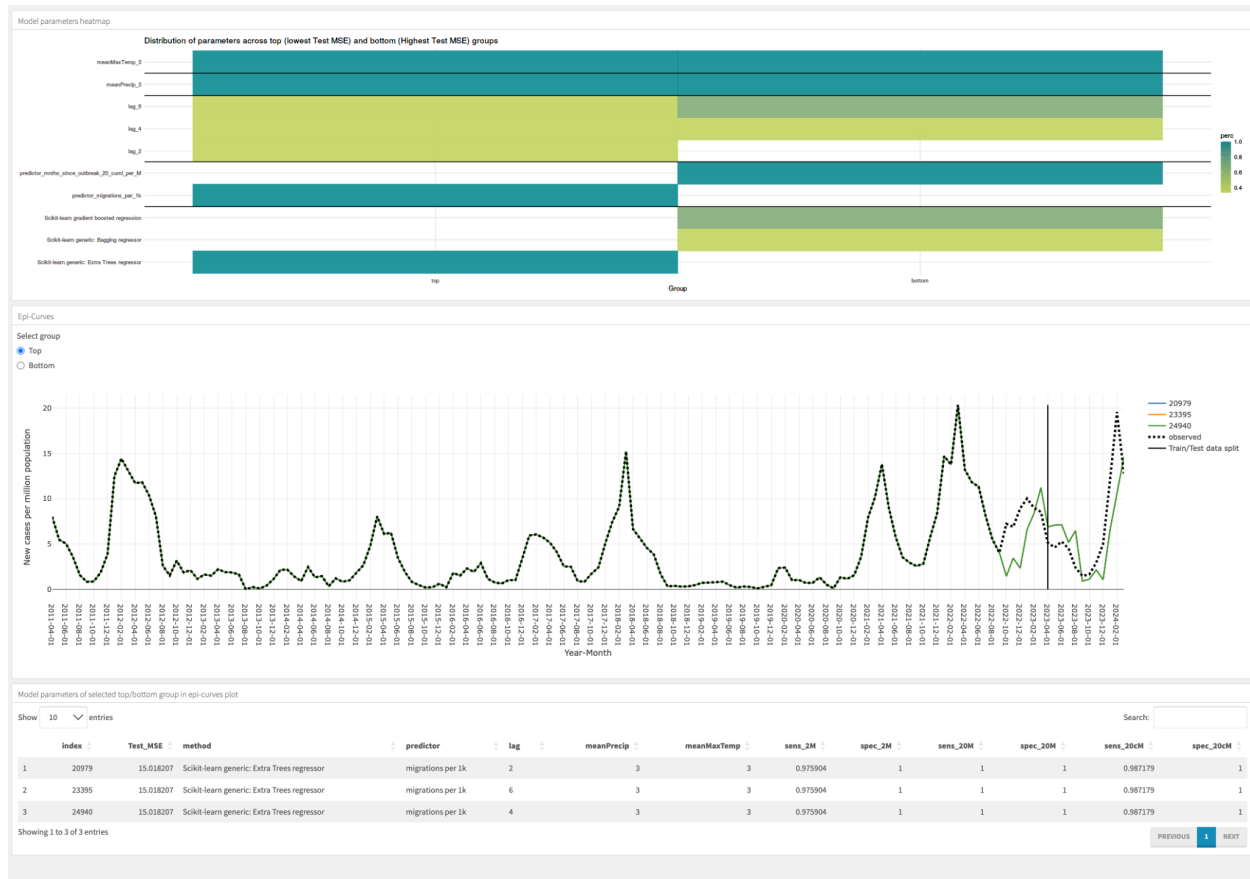
The red tabs at the top of the page will load different **assessment metrics** that can be used to assess model and method performance. The column on the left of the page is used to choose **model specifications** and further refine some of the assessment metrics.

Test Metrics

This tab contains (from top to bottom):

1. A table of the best performing models (according to the Test MSE metric) for the country selected in the left panel.
2. A heatmap showing the Test MSE values of combinations of methods and predictor lags for the specified country and model (both chosen on the left panel).
3. A bar chart comparing the Test MSE values of the best N method/model combinations (N is specified in the left panel).

Segment



This tab contains (from top to bottom):

1. A heatmap showing the percent of model parameters (methods, predictor variables, lags) across the top and bottom N models based on MSE (specified in the left panel). The darker green shades are more frequently present in the model group.
2. A time series showing the observed measles incidence and the predicted values according to the top N models. The vertical line shows the train/test split cutoff.
3. A table showing the metadata for the top N runs, along with sensitivity and specificity of the model for predicting outbreaks according to 3 different outbreak thresholds: 2 cases per million population, 20 cases per million population, and 20 cumulative cases yearly per million population.

Binary Outcome Analysis

Measles Output Visualizations
Test Metrics
Segment
Binary Outcome Analysis
Binary Outcome Plot
Metrics Correlation

Selections
 Select Country (All Countries available for Binary Outcome Analysis tab only)
 AFG - Afghanistan
 Test Metrics Tab - Heatmap: Select model vars
 mnth_since_outbreak_2_per_M, mean_precip_mm_per_day, mean_max_temp
 Test Metrics & Segment Tab - Number of runs in top/bottom groups
 5
 Binary outcome analysis - dataset
☒ Full
☐ Test
 Binary outcome analysis - outbreak threshold
☒ 20 cases per M
☐ 2 cases per M
☐ 20 cuml cases per M
 Binary outcome analysis - Run Index

Best Methods - Sensitivity
 Show 5 entries
 Search:

| | method | N |
|---|---|----|
| 1 | Scikit-learn generic: Extra Trees regressor | 46 |
| 2 | Scikit-learn generic: AdaBoost regressor | 1 |

 Showing 1 to 2 of 2 entries
 PREVIOUS 1 NEXT

Best Model - Sensitivity
 Show 5 entries
 Search:

| | indepVars | N |
|---|---|---|
| 1 | {'birth_per_1k': 0, 'mean_max_temp': 3, 'mean_precip_mm_per_day': 3} | 2 |
| 2 | {'mean_max_temp': 3, 'mean_precip_mm_per_day': 3, 'mnts_since_outbreak_20_cuml_per_M': 4} | 2 |
| 3 | {'mean_max_temp': 3, 'mean_precip_mm_per_day': 3, 'mnts_since_outbreak_20_cuml_per_M': 6} | 2 |
| 4 | {'mean_max_temp': 3, 'mean_precip_mm_per_day': 3, 'outgoing_air_passengers': 2} | 1 |
| 5 | {'MCV1_Cov_1yo': 0, 'mean_max_temp': 3, 'mean_precip_mm_per_day': 3} | 1 |

 Showing 1 to 5 of 44 entries
 PREVIOUS 1 2 3 4 5 ... 9 NEXT

Best Method - Specificity
 Show 5 entries
 Search:

| | method | N |
|---|---|----|
| 1 | Scikit-learn generic: Bagging regressor | 46 |
| 2 | Scikit-learn generic: AdaBoost regressor | 46 |
| 3 | Scikit-learn generic: Gradient Boosted Regression | 46 |
| 4 | Scikit-learn generic: Random Forest regressor | 46 |
| 5 | Scikit-learn generic: Extra Trees regressor | 46 |

 Showing 1 to 5 of 5 entries
 PREVIOUS 1 NEXT

Best Model - Specificity
 Show 5 entries
 Search:

| | indepVars | N |
|---|---|----|
| 1 | {'mean_max_temp': 3, 'mean_precip_mm_per_day': 3, 'mnts_since_outbreak_20_cuml_per_M': 4} | 10 |
| 2 | {'mean_max_temp': 3, 'mean_precip_mm_per_day': 3, 'mnts_since_outbreak_20_cuml_per_M': 6} | 10 |
| 3 | {'birth_per_1k': 2, 'mean_max_temp': 3, 'mean_precip_mm_per_day': 3} | 5 |
| 4 | {'MCV2': 2, 'mean_max_temp': 3, 'mean_precip_mm_per_day': 3} | 5 |
| 5 | {'MCV2_Cov_RecAge': 0, 'mean_max_temp': 3, 'mean_precip_mm_per_day': 3} | 5 |

 Showing 1 to 5 of 44 entries
 PREVIOUS 1 2 3 4 5 ... 9 NEXT

The binary outcome analysis assesses the model based on its ability to predict a measles outbreak according to three different measles incidence thresholds (select outbreak threshold in the left panel). Sensitivity and specificity can be derived from these binary predictions for the entire time series or the test dataset (specified on the left panel).

The analysis shows the best methods and predictor variables according to sensitivity (left column of tables) and to specificity (right column of tables). For each table, the number of times that method is represented in the highest sensitivity or specificity model is shown as N.

Within a country, these tables are most meaningful for comparing methods, not models (because each model/method combination is only represented one time, so only methods are widely repeated within the same country). However, an 'All Countries' option can be selected, which will show the best performing methods and models across countries.



Binary Outcome Plot



The binary outcome plot shows, for a single model run, the observed and predicted outbreak status according to the threshold specified on the left panel. This view allows the user to see where false negatives or false positives may occur during the time series. Additionally, a table below the plot summarizes the number of time steps in each cell of the contingency table.

A run index is required to view the plot, which is most easily accessed in the 'Segment' tab. An exhaustive list of all run index values is not currently available, so it is recommended to use an index that is revealed through the top N models returned in the Segment tab.