

Ebola Forecasting - Error Analysis

Andy Shen

Data Input and Cleaning

We assume the most accurate dataset is the most recent dataset of the outbreak. We tally the cases such that there is a running total of infections at each date.

We then import the projections from the Hawkes and Recursive models. For these predictions, the date preceding the forecasts is the last date of that dataset with at least one case. The forecasted numbers then predict the additional number of infections 7, 14, and 21 days after that date, respectively.

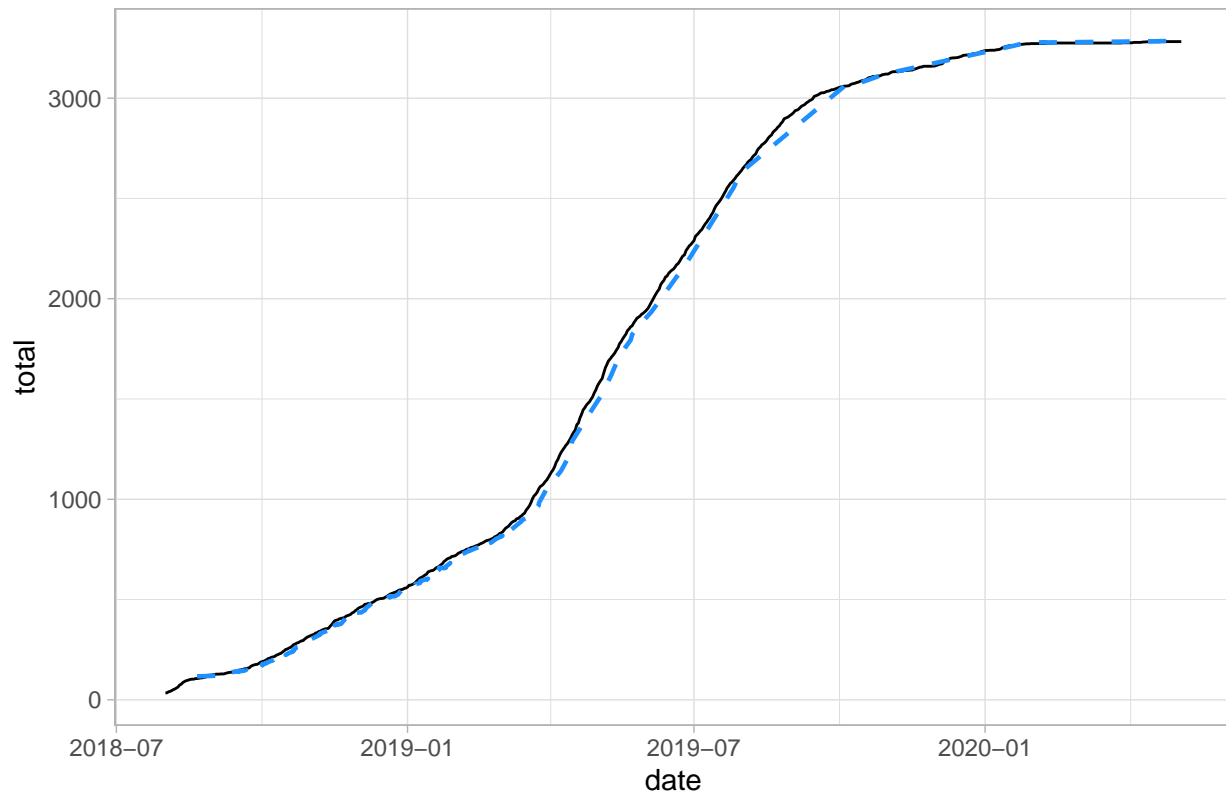
The code has been omitted for readability.

Hawkes Complete Outbreak Analysis

7-Day Forecast Analysis

Figure 1 below shows the Hawkes 7-Day Forecasts for all simulations with respect to the recorded number of infections. The RMSE values for all forecasts are included in Table 1.

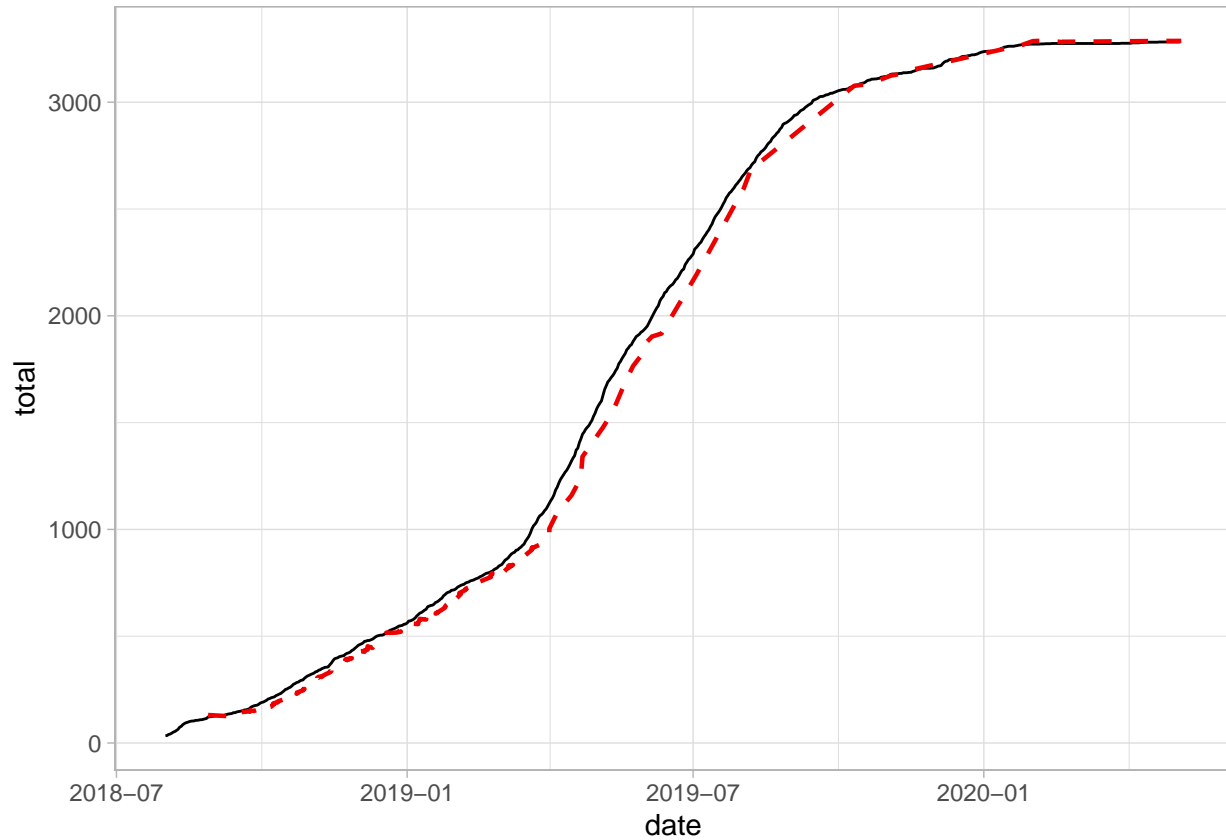
Figure 1: Hawkes 7-Day Forecasts for All Simulations



The 7-day Hawkes model generally shows accurate predictions throughout the duration of the pandemic, with slight under-prediction in 2019 during the middle of the pandemic. The Hawkes 7-day forecasts appear to predict the case counts at the beginning and end of the pandemic quite accurately.

14-Day Forecast Analysis

Figure 2 below shows the Hawkes 14-Day Forecasts for all simulations with respect to the recorded number of infections.

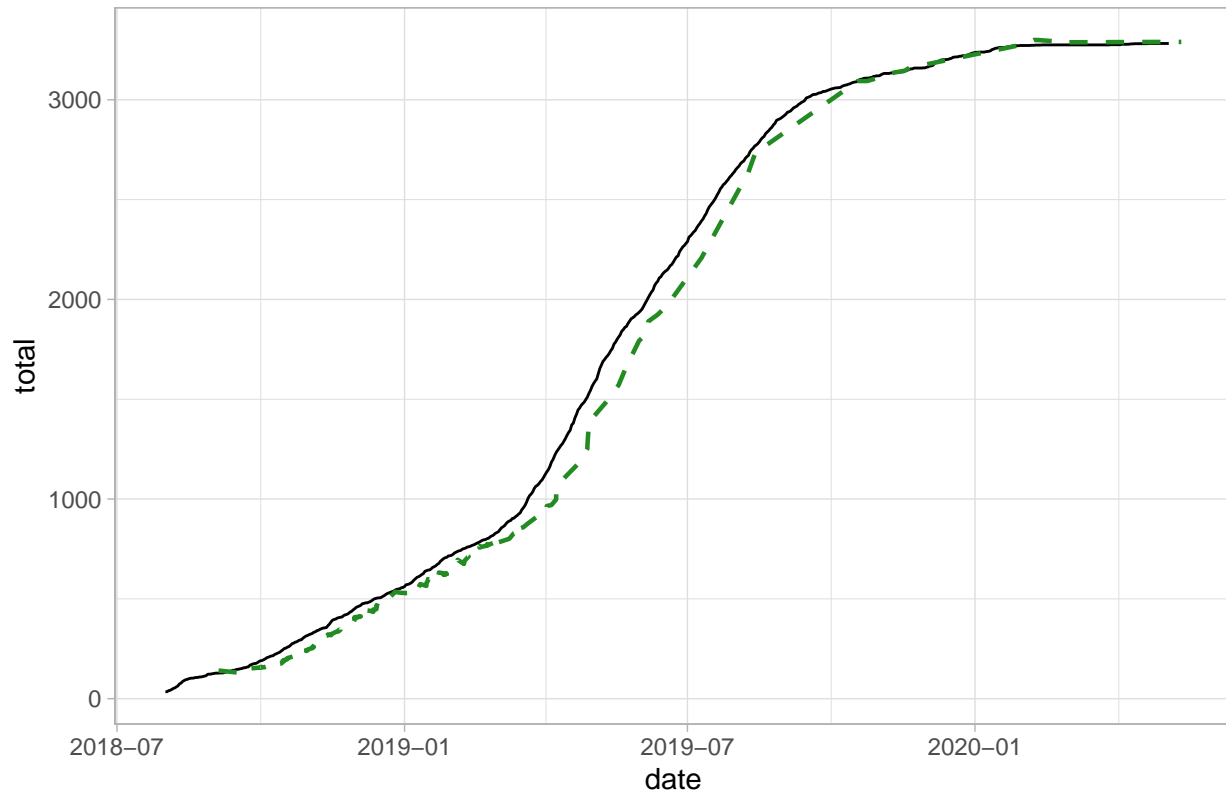


The 14-day Hawkes predictions tend to slightly under-predict the true case counts in the beginning and middle portions of the pandemic, with more accurate prediction in 2020 as the pandemic comes to an end. The largest prediction discrepancy is during 2019.

21-Day Forecast Analysis

Figure 3 below shows the Hawkes 21-Day Forecasts for all simulations with respect to the recorded number of infections.

Figure 3: Hawkes 21-Day Forecasts for All Simulations



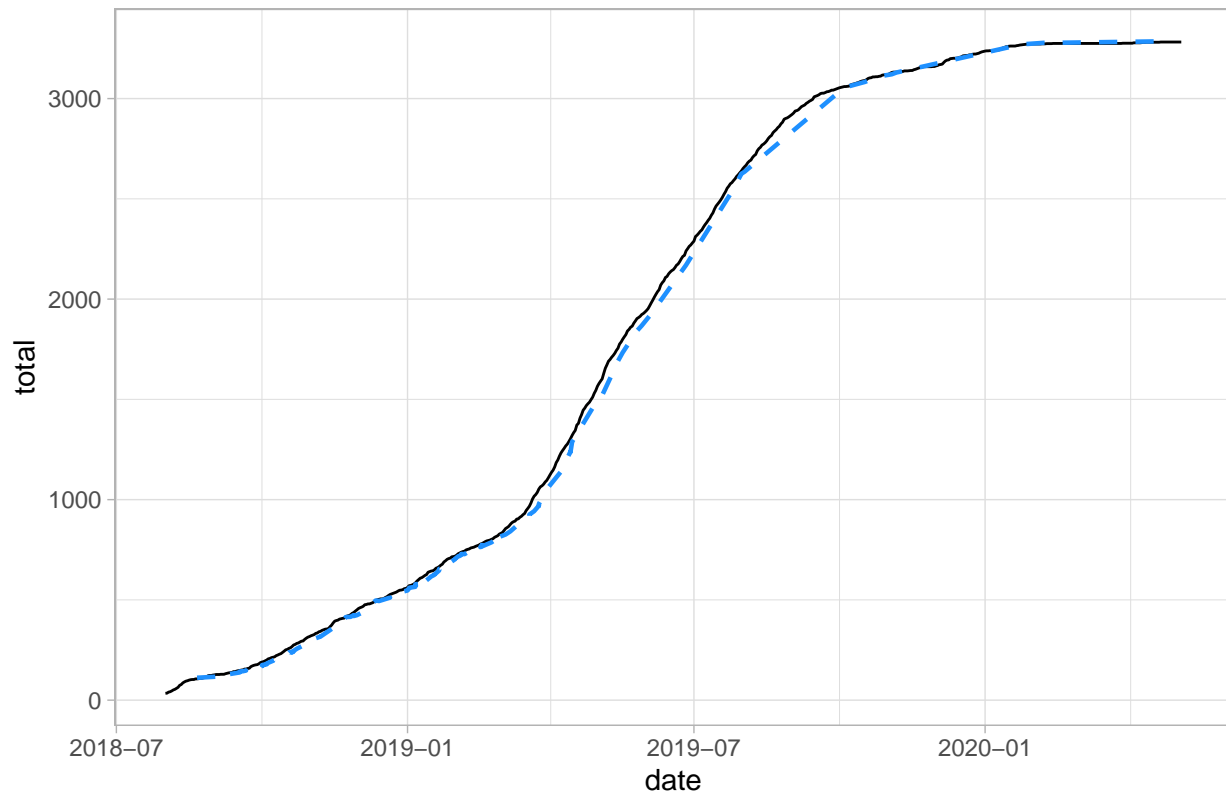
The 21-day Hawkes projections follow a similar pattern as 14-day Hawkes projections, with large under-predictions in the middle of the pandemic and more accurate projections in 2020 towards the end.

Recursive Complete Outbreak Analysis

7-Day Forecast Analysis

Figure 4 below shows the Recursive 7-Day Forecasts for all simulations with respect to the recorded number of infections.

Figure 4: Recursive 7-Day Forecasts for All Simulations

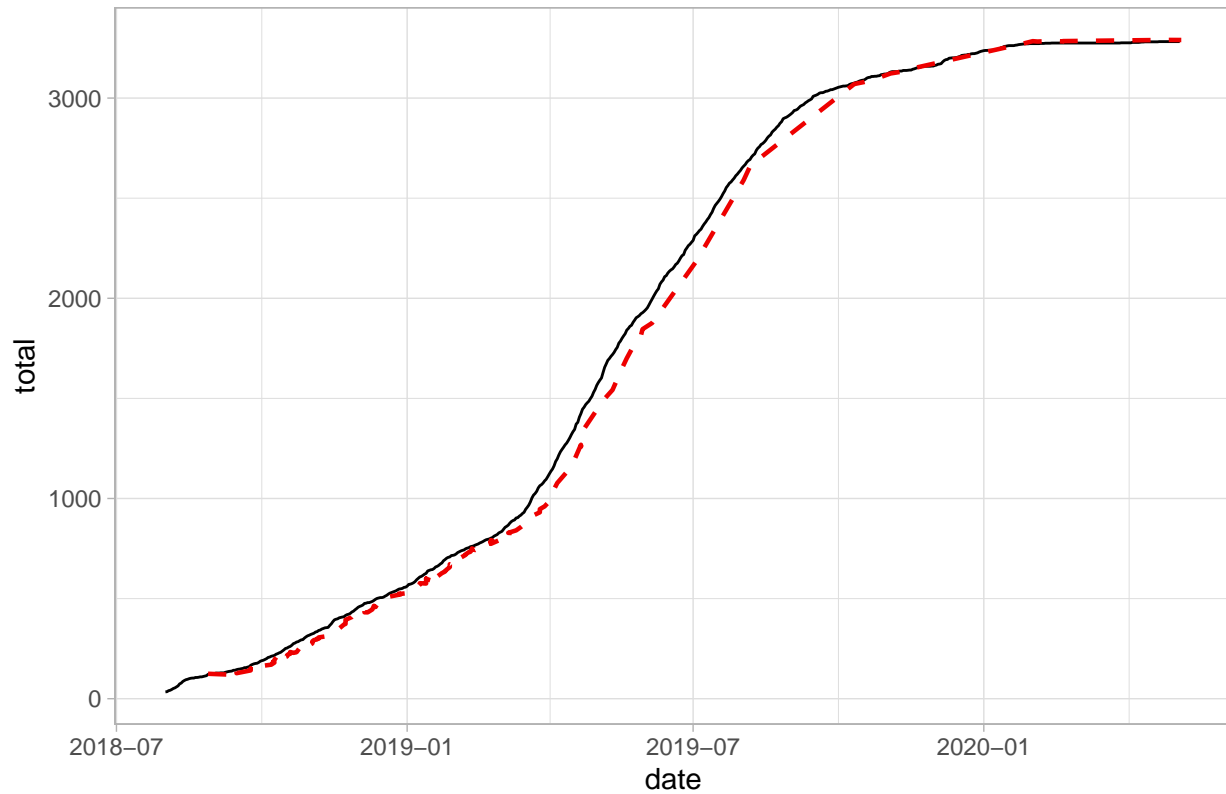


In general, the 7-day Recursive projections tend to under-predict the actual case counts. The largest errors generally occur during mid-2019, which is in the middle of the pandemic, whereas the model tends to have better prediction in the beginning and towards the end of the pandemic.

14-Day Forecast Analysis

Figure 5 below shows the Recursive 14-Day Forecasts for all simulations with respect to the recorded number of infections.

Figure 5: Recursive 14-Day Forecasts for All Simulations

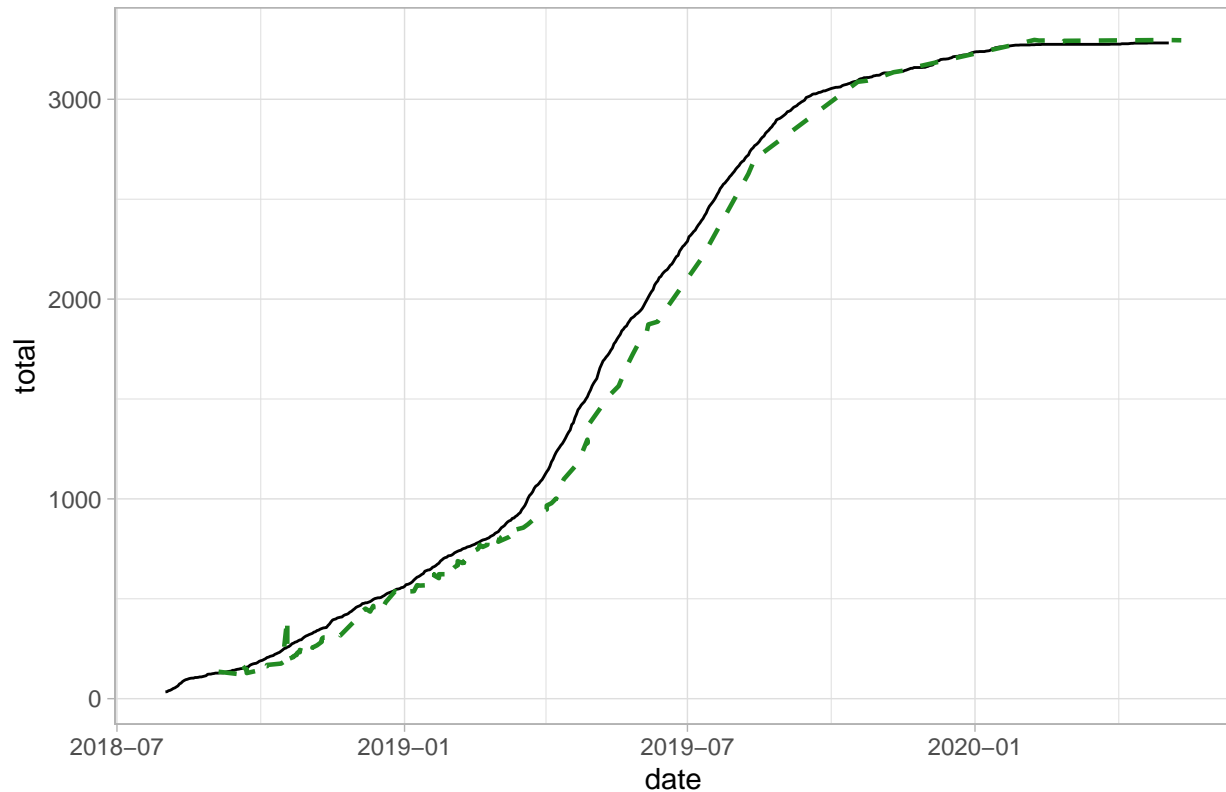


The 14-day Recursive forecasts tend to under-predict the actual case counts during most of the pandemic, and has better prediction at the end of the pandemic.

21-Day Forecast Analysis

Figure 6 below shows the Recursive 21-Day Forecasts for all simulations with respect to the recorded number of infections.

Figure 6: Recursive 7-Day Forecasts for All Simulations



Similar to the 14-day forecasts, the 21-day Recursive forecasts tend to under-predict the true case counts in the beginning and middle of the pandemic, but is more accurate towards the end of the pandemic with slight over-prediction.

RMSE for Full Hawkes and Recursive Forecasts

We compute the Root Mean Square Error (RMSE) of the 7, 14, and 21-day forecasts for both the Hawkes and Recursive models. The RMSE is computed as

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$

where N is the total number of observations.

The table below (Table 1) shows the RMSE values for the Hawkes and Recursive models, with respect to every simulated forecast during the outbreak.

Table 1: RMSE values for Hawkes and Recursive Models for all simulations.

	Hawkes	Recursive
7-day	27.90	29.67
14-day	58.68	60.70
21-day	88.44	92.22

We see that the Hawkes model forecasts have a consistently lower RMSE than those of the Recursive model for all three prediction days, when looking at all simulations.

Hawkes Partial Outbreak Analysis

After analyzing the entire model, we then adjust our analysis to account for the varying number of forecasts performed during a year. The pandemic began in mid-2018 and lasted until mid-2020. Since more forecasts were performed in the earlier parts of the pandemic compared to the later portion, we adjust our data to reflect an equal number of simulations per year, so that each year as a proportionate number of forecasts.