

# Ebola Forecasting - Residual Analysis

Andy Shen

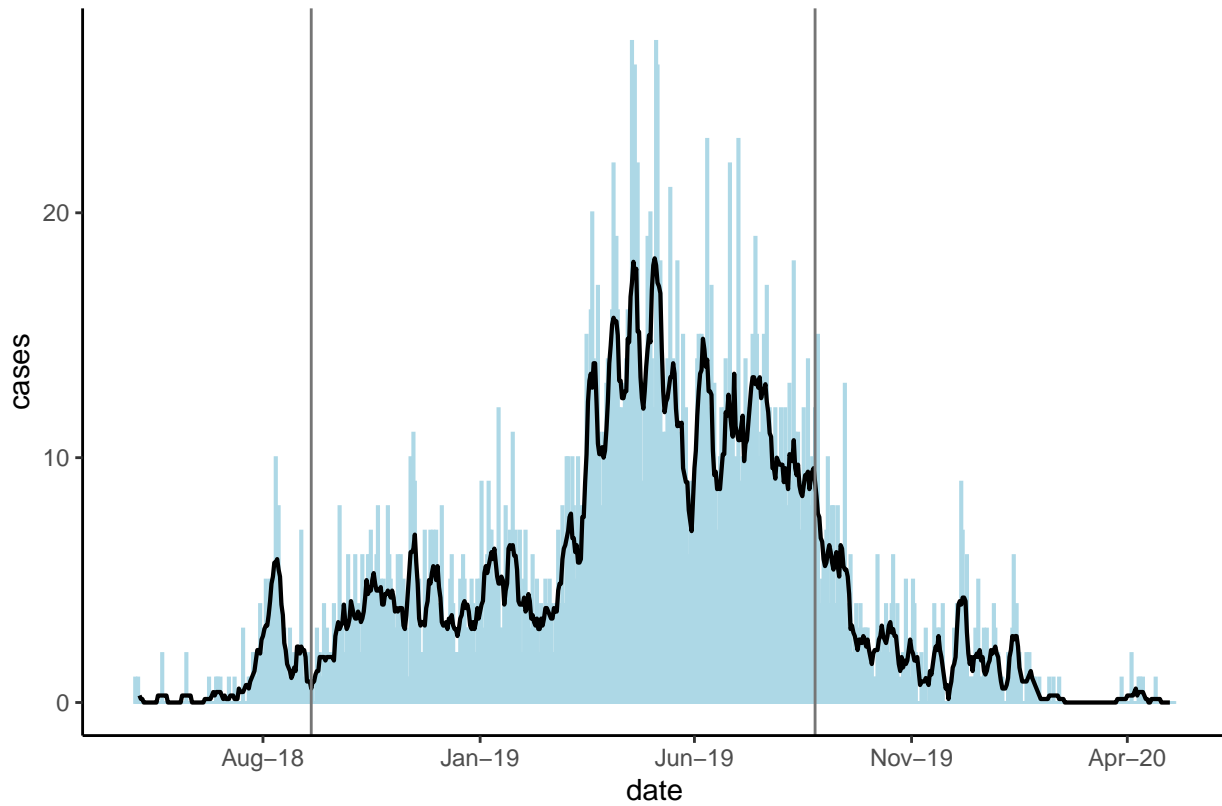
9/11/2020

## 1 Setup

We define the beginning, middle, and end of the outbreak such that the beginning is from May 3, 2018 to September 4, 2018, the middle from September 5, 2018 to August 25, 2019, and the end from August 26, 2019 to May 4, 2020.

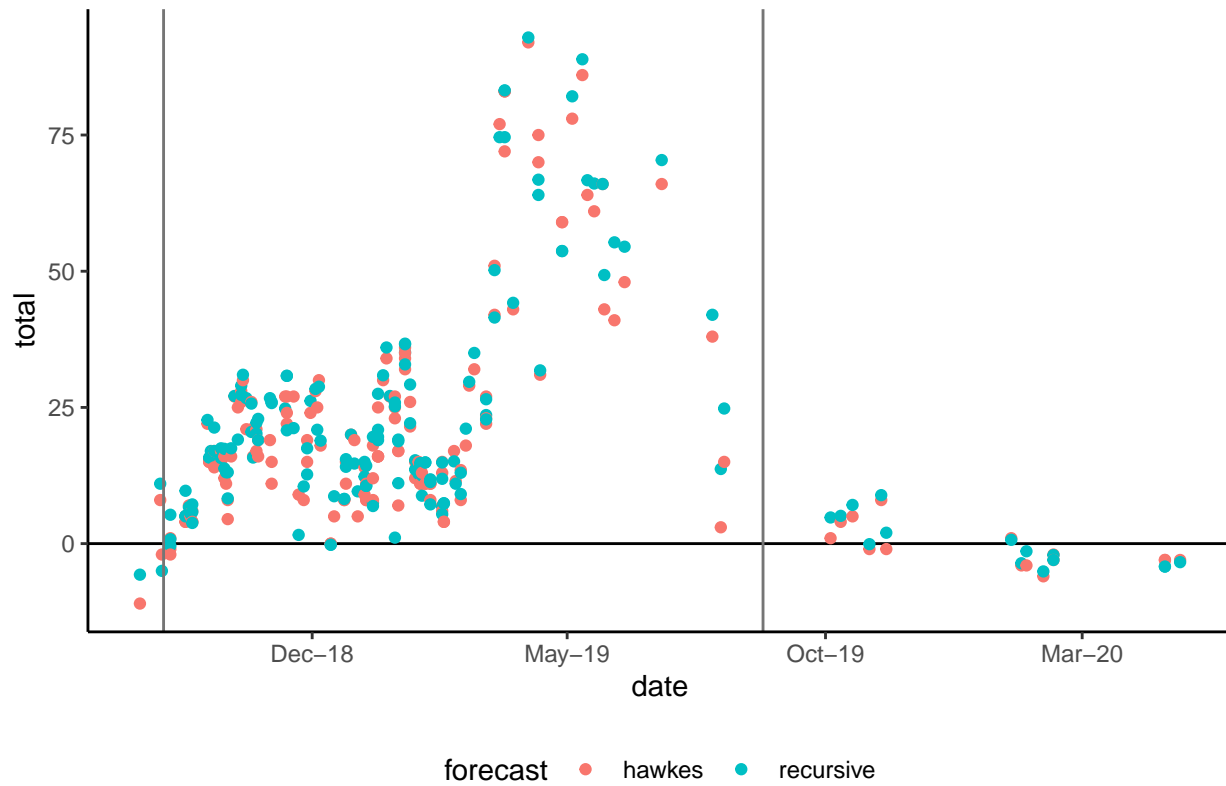
## 2 New Ebola Cases by Day

The figure below shows the number of new Ebola cases in West Africa by day during the outbreak. The grey lines represent boundaries between the beginning, middle and end of the outbreak (see section 1).



Ebola Cases by Day in West Africa with 7-day Rolling Average

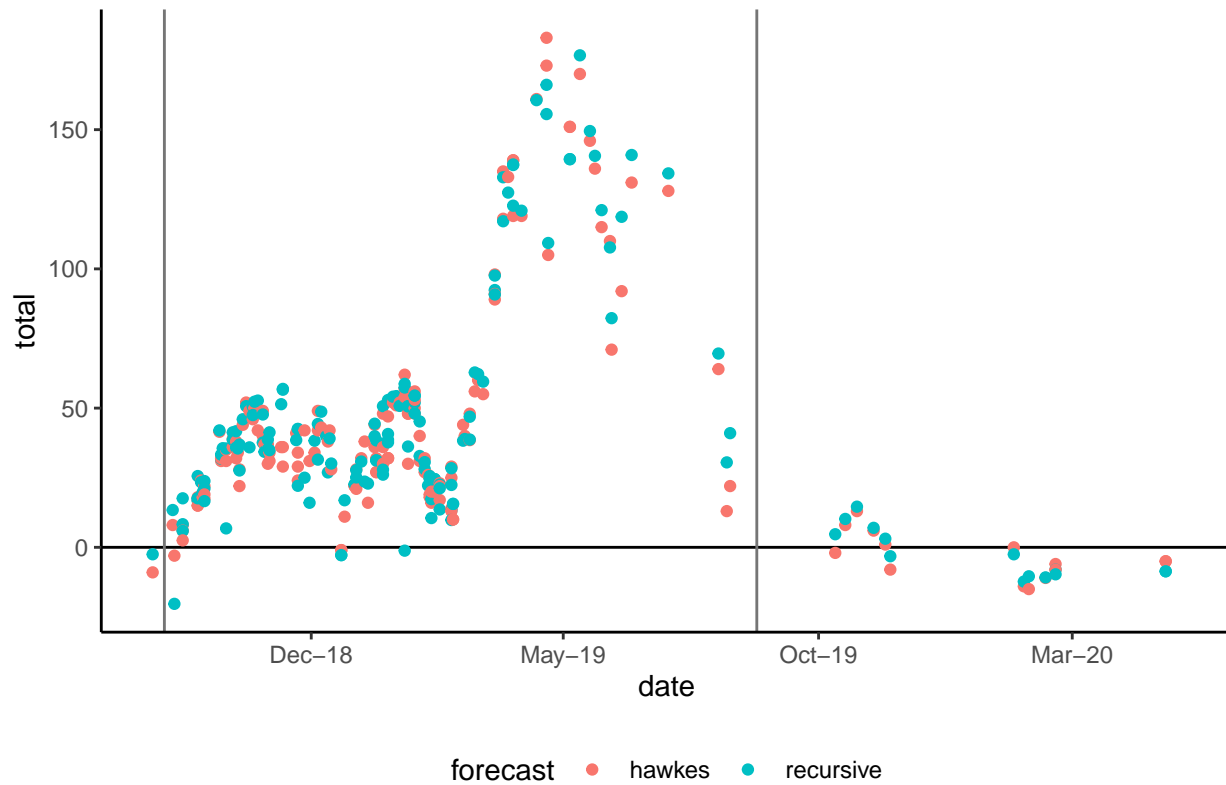
### 3 Hawkes vs Recursive 7-Day Residual Plot



Residual plot of Hawkes and Recursive 7-day models

## 4 Hawkes vs Recursive 14-Day Residual Plot

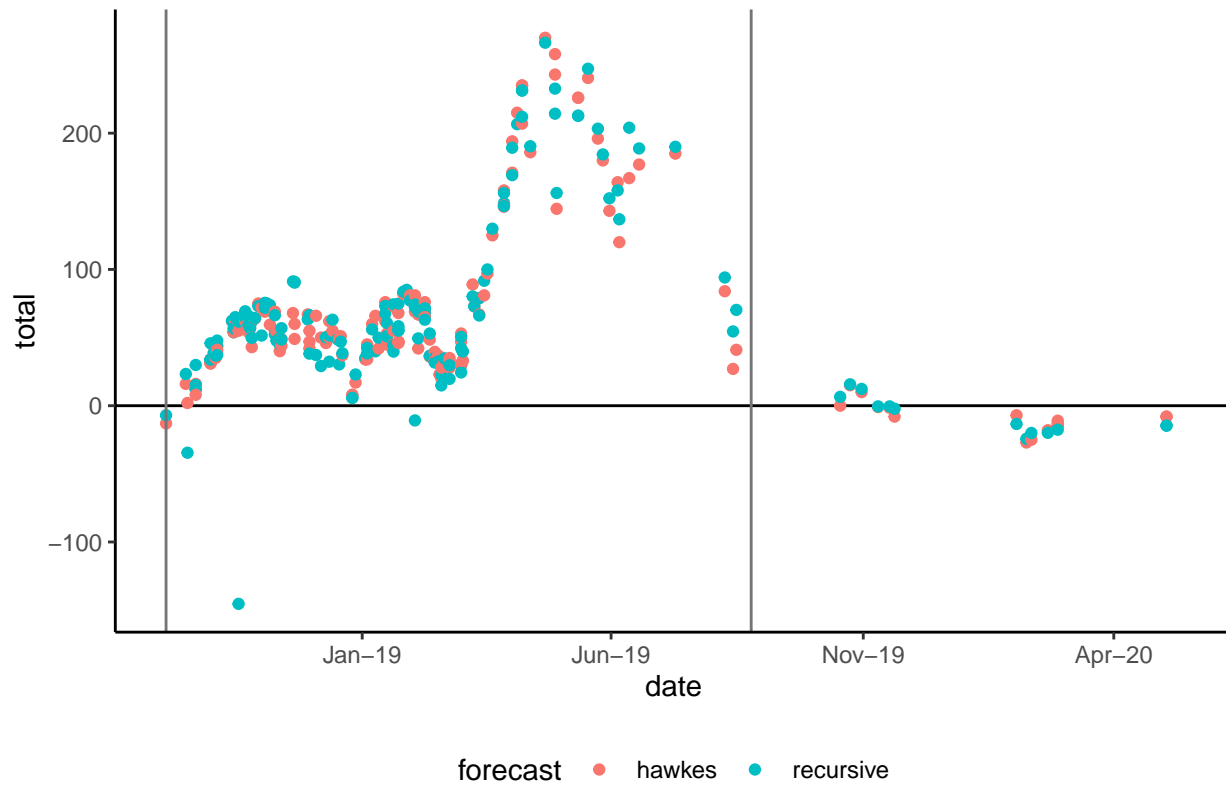
## Warning: Removed 2 rows containing missing values (geom\_point).



Residual plot of Hawkes and Recursive 21-day models

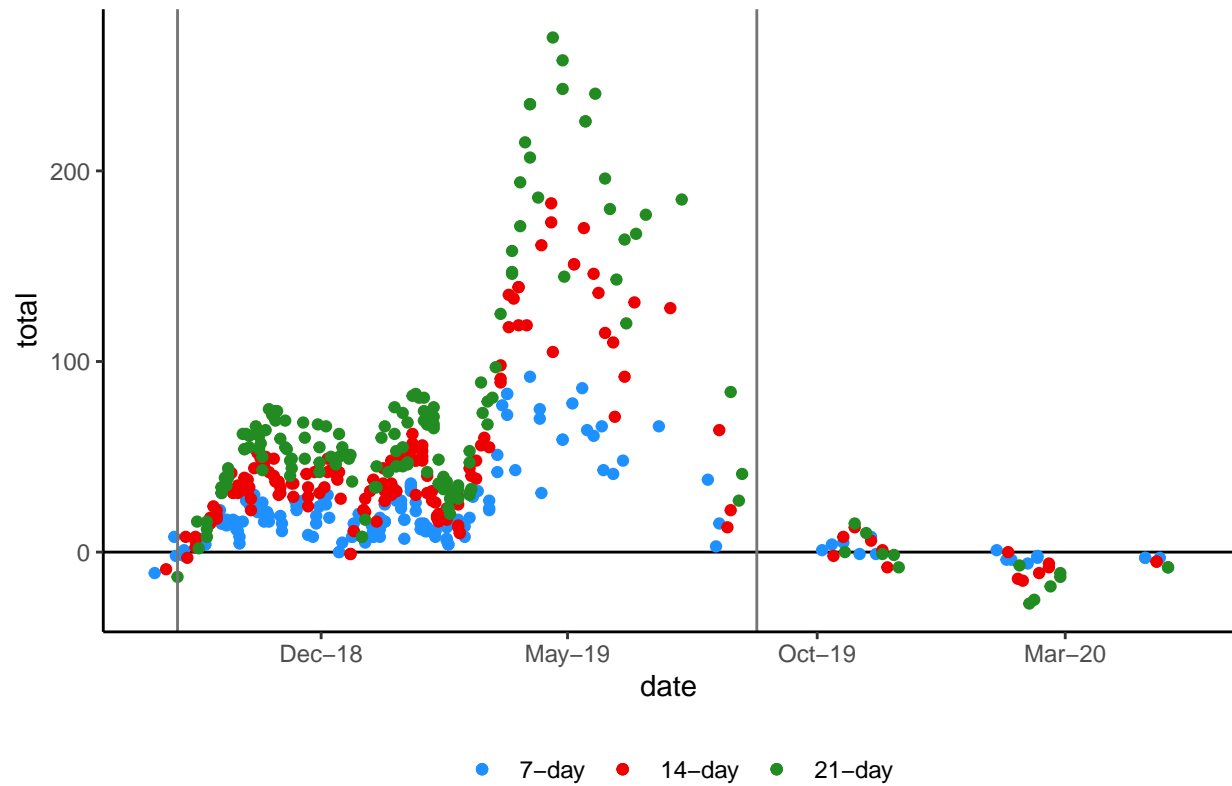
## 5 Hawkes vs Recursive 21-Day Residual Plot

## Warning: Removed 2 rows containing missing values (geom\_point).



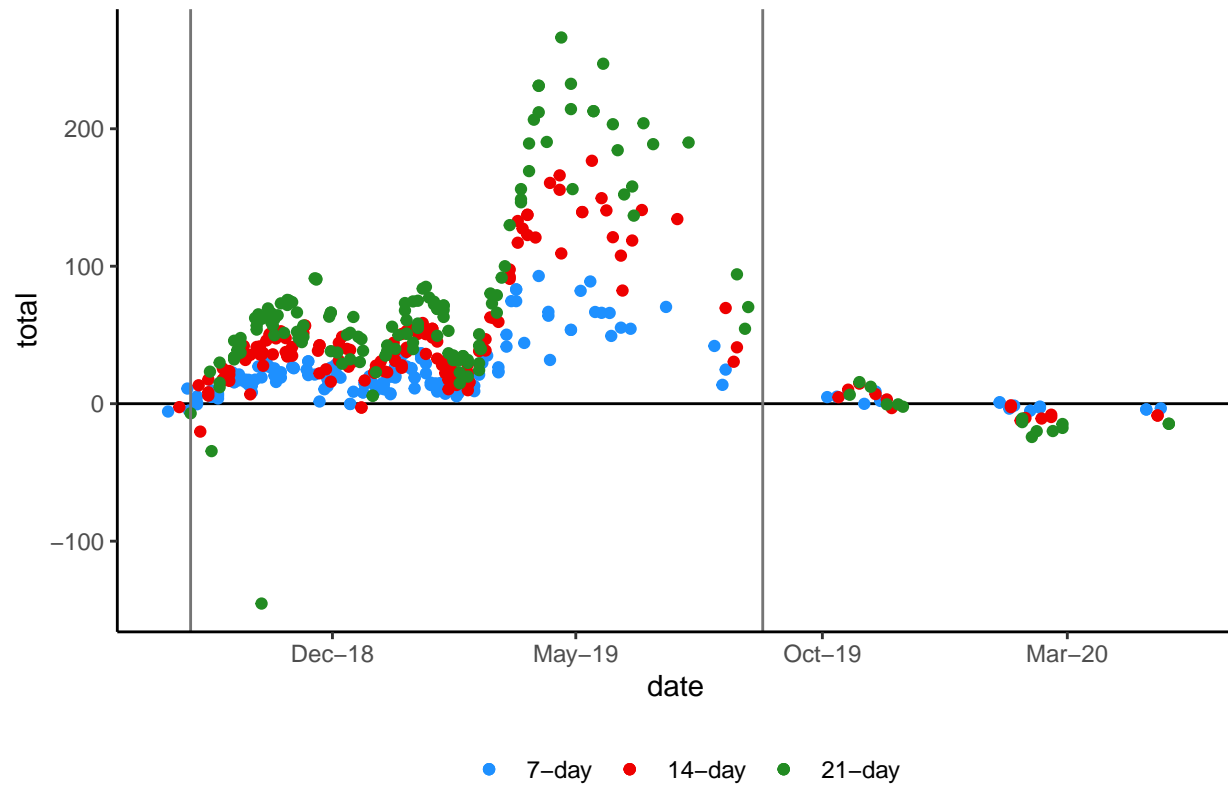
Residual plot of Hawkes and Recursive 21-day models

## 6 Hawkes-Only Residual Plot



Residual plot of Hawkes 7-, 14-, and 21-day models

## 7 Recursive-Only Residual Plot



Residual plot of Recursive 7-, 14-, and 21-day models

## 8 RMSE Calculation During Outbreak

### 8.1 RMSE Calculation

The table below shows the RMSE values from both the Hawkes and Recursive models during the beginning, middle, and end of the outbreak. The dates for each section are defined in section 1.

Table 1: 7-Day Hawkes and Recursive RMSE values during the outbreak

	Hawkes	Recursive
Beginning	5.701	5.879
Middle	30.749	31.760
End	3.804	4.225

Table 2: 14-Day Hawkes and Recursive RMSE values during the outbreak

	Hawkes	Recursive
Beginning	6.586	12.983
Middle	63.869	64.703
End	8.577	8.555

Table 3: 21-Day Hawkes and Recursive RMSE values during the outbreak

	Hawkes	Recursive
Beginning	12.200	22.551
Middle	96.862	98.363
End	13.462	14.326

It is clear that both models are most accurate in the beginning and end of the pandemic as opposed to the middle portion.