

# Real-time probabilistic forecasts of the 2018 – 2019 Ebola outbreak in North-Eastern Democratic Republic of Congo

Candidate number: **J05463**

September 29, 2019

## **Word count:**

Abstract: 406

Report: 9,513

## **Abstract**

**Objectives:** Construct and evaluate probabilistic forecasting models for the 2018–2019 Ebola outbreak in north-eastern DRC on both a national and sub-national level.

**Design:** Using publicly available data on daily numbers of reported Ebola cases from the 2018–2019 Ebola outbreak in DRC, we constructed a set of semi-mechanistic forecasting models based on a branching process with a time-varying reproduction number. This class of models is based on offspring distributions, serial intervals and reproduction numbers, with no additional assumption about the disease transmission process. We included models with both Poisson and negative binomial offspring distributions and two methods for forecasting the reproduction number. These methods used the estimated historic evolution of reproduction number and then either forecasted a constant reproduction number or used a forecast based on a Bayesian state-space

model with a semi-local trend. We fitted the models both to national-level data for the whole outbreak and for each individual health zone with forecasting horizons up to 28 days. We compared the forecasts from the different models using proper scoring rules and checked the calibration of the forecasts using the randomised Probability Integral Transform.

**Results:** On the national level we found that a negative binomial distribution was needed for calibrated one-day-ahead forecasts, but that the model with a Poisson offspring distribution and a semi-local trend for the reproduction number was the best model for longer forecasts and could provide calibrated forecasts for a forecasting horizon of nine days. The semi-local-trend models could also provide calibrated forecasts for some of the health zones, but for most of the health zones with a large number of cases the models did at best provide a few days of calibrated forecasts. At both the national and the sub-national level, we found that the models with constant reproduction number would normally underestimate the uncertainty in the process and that semi-local-trend models are likely to overestimate the uncertainty of the process.

### **Conclusions:**

Flexible semi-mechanistic models based on a branching process can be used to give calibrated short-term forecasts of the Ebola outbreak in north-eastern DRC on the national level and partly on the sub-national level. The forecasts can be provided in real time to support the outbreak response and can provide insight to increase our understanding of the outbreak.

**Keywords:** Ebola, DRC, probabilistic forecasts, mathematical modelling, semi-mechanistic model, calibration, proper scoring rules.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	2018–2019 Ebola outbreak in north-eastern DRC . . . . .	5
2.2	Mathematical modelling of Ebola . . . . .	7
<b>3</b>	<b>Methods</b>	<b>8</b>
3.1	Data Sources . . . . .	8
3.2	Ethics . . . . .	9
3.3	Model . . . . .	9
3.3.1	Offspring distribution . . . . .	10
3.3.2	Reproduction number . . . . .	12
3.3.3	Forecasting incidence . . . . .	14
3.4	Assessing probabilistic forecasts . . . . .	15
3.4.1	Calibration . . . . .	15
3.4.2	Sharpness and bias . . . . .	17
3.4.3	Proper scoring rules . . . . .	17
3.5	Implementation . . . . .	18
<b>4</b>	<b>Results</b>	<b>20</b>
4.1	National-level forecasts . . . . .	21
4.2	Sub-national forecasts . . . . .	24
<b>5</b>	<b>Discussion</b>	<b>27</b>
<b>6</b>	<b>Conclusions</b>	<b>35</b>
<b>Appendices</b>		<b>47</b>

# 1 Introduction

Real-time modelling and forecasting of infectious diseases can play a crucial role during disease outbreaks like the ongoing Ebola outbreak in the north-eastern Democratic Republic of Congo (DRC) [1; 2]. The use of models for Ebola response was pioneered during the previous large Ebola outbreak in West Africa in 2014–2016 [3]. The models developed during and after this outbreak were used both to increase our understanding of the outbreak and to predict the spread of the outbreak. Forecasts were used in the official response planning [4] and to estimate needed bed capacity [5]. Disease modelling has also been successfully implemented and used to inform response plans for many other infectious diseases including Zika [6] and influenza [7]. To make effective use of results from mathematical modelling in an outbreak it is of key importance that modelling outputs are integrated into the outbreak response and decision making [2]. This means that forecasts need to be available in real time and integrated into the outbreak surveillance information flow. In addition, improved communication and embedding of modelling and modellers into the outbreak response is needed.

To provide useful information in an outbreak situation, models should be probabilistic so that they assess and incorporate uncertainty into their predictions [8; 9; 10; 11]. An understanding of the range of possible outcomes from a model and their associated probabilities is essential for an effective use of modelling results in an outbreak situation since there are both large uncertainties and randomness associated with the disease transmission process and in the necessary reduction of this process to a mathematical model. Once a model has been formulated, it needs to be critically evaluated before use. Evaluations of probabilistic forecasts not only requires us to assess the accuracy of point predictions, but to assess if the model correctly assess its own uncertainty [12; 13]. A model that successfully assesses its own uncertainty is said to be calibrated. We should strive to create well calibrated models which can be updated in real time and that are interpretable and useful during the outbreak response.

The aim of this report is to describe a framework for flexible probabilistic forecasts for the ongoing Ebola outbreak in north-eastern DRC and to assess the calibration of the resulting models. We will compare different models using proper scoring rules to choose the model that gives the best forecasts, and use these models to explore the epidemiology of the current outbreak.

## 2 Background

### 2.1 2018–2019 Ebola outbreak in north-eastern DRC

Ebola virus disease (EVD) is an often fatal disease caused by the Ebola virus that is transmitted through close contact with bodily fluids from infected persons [14]. The symptoms start 2 – 21 days after infection and include fever, vomiting, diarrhoea and in some cases internal and external bleeding. Risk factors include close contact with or handling of the infected person or their dead body. This has made outbreak response challenging since health-care workers have had a high risk of infection [15]. Another important source of infections have been burial practices that include direct contact with the dead body [15]. There have been over 20 separate Ebola outbreaks since the first recorded outbreak in 1976, with a basic reproduction number,  $R_0$ , that has normally been between 1.5 – 3 [4; 16] and a case fatality rate that has varied between 10 and 90 percent. Apart from the large outbreak in Western Africa in 2014–2016, which had over 28,000 cases, all of the previous outbreaks have had less than 500 cases [14], making the current outbreak the second largest in history.

The current outbreak in north-eastern DRC started on the 1st of August 2018, when the Ministry of Health in the Democratic Republic of Congo notified the WHO about a new Ebola outbreak in the North Kivu province [17]. North Kivu and the surrounding provinces Ituri and South Kivu constitute a populous region in north-eastern DRC which borders both Rwanda and Uganda. In recent history the people in the region have been involved in both the Second Congo War (1998-2003) and the ongoing “Kivu Conflict”. The Kivu conflict has

seen local rebel forces fight against the armed forces of the government. In addition to local forces, a large UN peacekeeping force has been heavily involved in the region. This makes the area very difficult to operate in due to security concerns and the high levels of distrust from local population towards both the central government and international organisations. By the 27th of September 2018, the outbreak had continued spreading and the WHO assessed that the outbreak constituted a very high national and regional risk [18].

The difficult security situation and the distrust from the local community have made the outbreak response very challenging. This has lead to difficulty in tracing contacts, giving vaccinations and treating patients. There have been many episodes of violence directed towards the Ebola responders, which has led to fatalities and multiple stops in the response activities [19; 20]. Despite these difficulties the outbreak has mainly been contained in the region, with only imported cases and no transmission seen in the large city of Goma or across the borders in Uganda and Rwanda.

During the outbreak, the experimental vaccination rVSV-ZEBOV-GP has been used in a ring vaccination strategy. The vaccine was given under the compassionate use regime and has been evaluated during the outbreak. Preliminary data shows positive signs that this vaccine will provide protection from Ebola [21]. The vaccines have also been given to health-care workers and children. This outbreak has also seen a randomised trial for four proposed medicines to treat Ebola. The trial was terminated on the 12th of August as it was found that two of the four drugs that were included showed a clear effect of reducing mortality from Ebola, especially with early treatment [22].

On the 17th of July the Director General of the WHO declared the Ebola outbreak a Public Health Emergency of International Concern [23] and as of mid-September 2019 the outbreak is still ongoing with recent new cases from 17 health zones [24].

## 2.2 Mathematical modelling of Ebola

A large number of different mathematical models have been proposed for modelling and forecasting of Ebola[3; 25]. The models can be divided into categories based on their structure: mechanistic, semi-mechanistic, phenomenological and hybrid models. Mechanistic models are based on a specified mechanism for disease transmission and include standard compartmental models, while phenomenological models aim to describe empirical relationships and do not include any specified mechanism for disease transmission. For Ebola prediction, an example of a mechanistic model is the SEIR model in [26]. Many different phenomenological models have been developed for Ebola predictions. They include logistic growth models [27] and autoregressive integrated moving average models used for many applications of time-series forecasting [28]. A semi-mechanistic model includes some mechanistic assumptions, together with some parts of the model that are more empirical or phenomenological. The aim of these models is to model some important parts of the disease transmission process, while giving enough flexibility to fit complicated outbreak histories. For Ebola forecasting, examples include SEIR models with a random walk describing the transmission rate or reproduction number [8; 29]. Based on a review of models for the West African Ebola outbreak, it was found that it was important that models actively included the uncertainty in the disease transmission process [3; 30] and that community standards for developing, reporting and comparing models are put in place.

The most recent evaluation of Ebola models was the RAPIDD Ebola Forecasting challenge in 2018 [25]. The challenge presented a range of different scenarios of a simulated Ebola outbreak that was similar to the West African outbreak, and invited different groups to submit forecasts and compared them. Eleven different models with a large range of different modelling structures were submitted to the competition. When the models were compared, it was found that for short-term forecasting the accuracy was not related to model complexity and that “light” non-parametric models achieved similar forecasting accuracy to larger mechanistic models. The more complex mechanistic models would, on the other hand, be

better suited to understand the effects of interventions.

In this thesis we will present a semi-mechanistic modelling framework that includes some mechanistic assumptions about serial intervals and forces of infection combined with flexible time-series methods. The aim is a flexible model with parameters that are easily interpretable and useful during an outbreak response situation. This also requires that the models can be run in real time and integrated with the response surveillance.

## 3 Methods

### 3.1 Data Sources

We will use daily reports of new cases from the Ebola outbreak in north-eastern DRC as reported by the Ministry of Health in DRC and the World Health Organisation. This data has been made publicly available and has been distributed via the Humanitarian Data Exchange [31]. We use data from the beginning of the outbreak until the 1st of September 2019. The data gives the total cumulative number of cases reported every day, which we used to calculate the incident cases both at the national and at the health-zone level. For our models, we will use confirmed cases so that estimates of the reproduction number and other epidemiological parameters correspond only to Ebola and not to Ebola and Ebola-like illnesses. Due to the nature of the outbreak and the nature of the data, some confirmed cases are later found to not be Ebola after all. This can lead to days with negative incidence since the cumulative number of cases used to calculate the incidence has been corrected from an earlier date. We have chosen to set any negative incidence to zero as we do not have any information about which of the earlier cases has been removed. In July and August 2018 there were some days without any reported data. For these days we equally distribute the number of new cases seen at the end of the interval of missing data to each missing day.

## 3.2 Ethics

Since only secondary data fully in the public domain was used, no ethics approval was needed (LSHTM MSc Ethics ref: 16735).

## 3.3 Model

With our aim of constructing a flexible semi-mechanistic framework for the modelling of Ebola we considered models where the daily incidence,  $I_t$ , follows a modified branching process [32]. This class of models has been considered for infectious disease modelling previously in [33; 34; 35] and was used to model Ebola during the 2014–2016 outbreak in Western Africa [4; 36]. The main idea is that each infected person gives rise to  $\nu$  new infections, where  $\nu$  is a random variable distributed with an offspring distribution with an expected value given by the reproduction number  $R$ . This gives rise to the standard interpretation of the reproduction number as the average number of secondary cases caused by one primary case. Using random variables for the number of secondary cases gives a stochastic model that allows us to model the inherent uncertainty in the disease transmission process. To model the time between the initial case and the secondary cases, we will use the serial interval. This gives a process where the expected number of new cases at time  $t$  is given by the force of infection,  $\lambda_t$ , which is a product of previous incidence values weighted by the discretised serial interval,  $w_\tau$ , and the instantaneous reproduction number,  $R_t$ . Here  $R_t$  is reproduction number for the infected people who are transmitting the disease at time  $t$ . We use daily time intervals for all the calculations. The expected value of the incidence at time  $t$  is given by

$$E(I_t) = \lambda_t = R_t \sum_{s=1}^{t-1} I_s w_{t-s}. \quad (1)$$

This equation gives a general and flexible modelling framework where an individual model is specified by the probability distribution for  $I_t$ , the serial interval and the reproduction number as a function of time. Due to the semi-mechanistic nature of this model, it does not

depend on many assumptions about the disease transmission process. The main assumptions include that the force of infection is given by the weighted sum of incidence and serial interval and that the offspring distribution and time-varying reproduction number can adequately describe the outbreak. For all the models in this study we will use a gamma distribution with a mean of 15.3 days and standard deviation of 9.3 days as the serial interval. This was the serial interval observed during the West-African Ebola outbreak [4]. Our general approach will be to use a few different specifications for the offspring probability distribution and for estimating the time-varying reproduction number and then assess which model gives the best fit to the data.

### 3.3.1 Offspring distribution

The offspring distribution gives the distribution of the number of new infections from each infected person. By definition the expected value of this distribution is the reproduction number. The offspring distribution needs to be a discrete probability distribution over the non-negative integers. The simplest such distribution is the Poisson distribution, where the probability of observing  $k$  cases,  $P(X = k)$  is given by

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

where the  $\lambda$  is both the mean and the variance of the distribution. This distribution implies that the number of new infections over a small time interval is described by a constant rate where the chance of a new infection in each small interval is independent.

For a number of diseases it has been shown that the offspring distribution has a larger variance than that given by the Poisson distribution and has so called super-spreaders [34]. One common way of modelling this over-dispersed distribution is by the negative binomial distribution. The negative binomial distribution describes a process where we have a series of independent opportunities to infect a new person, each with probability,  $p$ , where the

process continues until we have had  $r$  infection opportunities without an infection. The distribution is given by

$$P(X = k) = \binom{k + r - 1}{k} (1 - p)^r p^k,$$

with  $E(X) = \frac{pr}{1-p}$  and  $Var(X) = \frac{pr}{(1-p)^2}$ . We will use a parameterisation of the negative binomial distribution where we use the mean,  $\mu$ , and the dispersion parameter  $k$ . In this formulation the variance is given by  $\mu + \frac{\mu^2}{k}$ , and when  $k \rightarrow \infty$  we get back a Poisson distribution. The smaller  $k$  is, the more the outbreak is dominated by a few super-spreaders, and the larger  $k$  is, the more similar the impact of each infected person. For SARS, it has been found that  $k = 0.16$  [34]. For the Ebola outbreak in West Africa, they found  $k = 0.53$  using conservative assumptions and data from the transmission tree of the infections [36].

The different offspring distributions can be seen in Figure 1

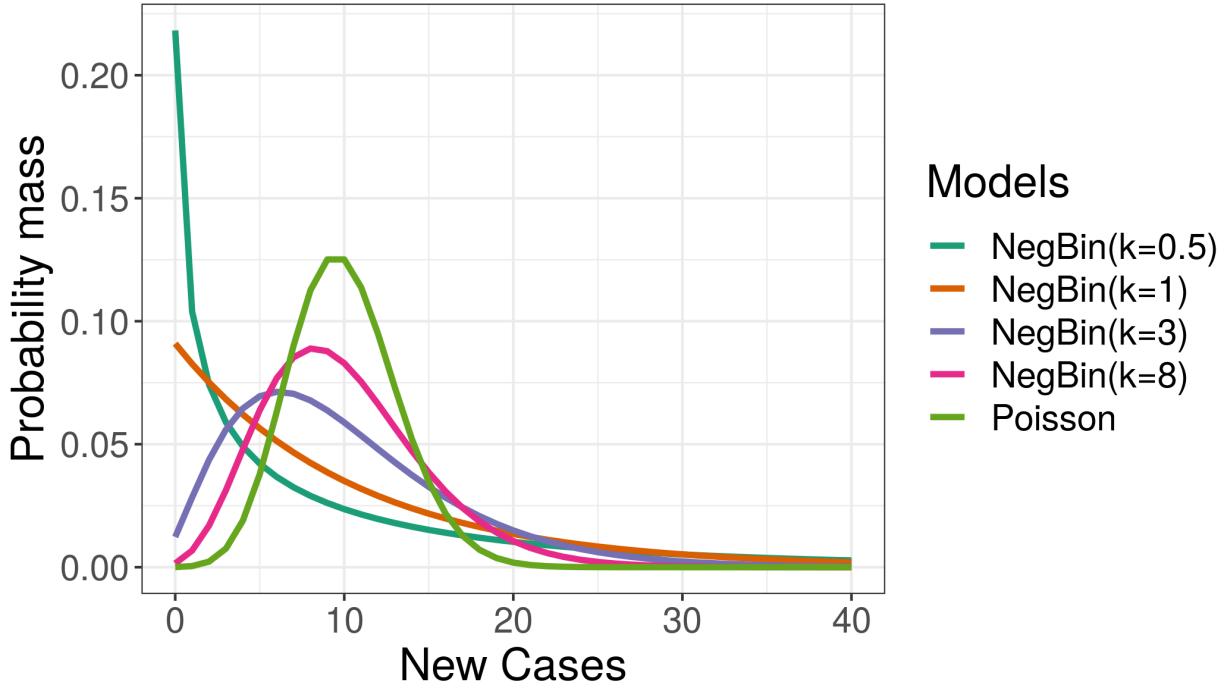


Figure 1: Poisson and negative binomial offspring distributions with a varying dispersion parameter  $k$ . All the distributions have a mean of 10.

The distribution for total incidence is given by the sum of the offspring distributions for

each infected person weighted by the serial distribution. For both the Poisson and negative binomial distribution, the sum of independent realisations of the distributions will give back the same distribution with an expected value given by the sum of the expected values. For the negative binomial distribution the dispersion parameter  $k$  will remain the same if it is equal for all the realisations. Therefore the probability distribution for incidence is also given by either the Poisson or the negative binomial distribution if the individual offspring distribution is given by one of the two distributions. We will investigate models with both of these offspring distributions to assess which models make better predictions during the current outbreak.

### 3.3.2 Reproduction number

The final ingredient needed to completely specify the dynamics of the model is the evolution of the reproduction number with time. The function for  $R_t$  could be modelled as a parametric function where we could fit the parameters to data. Alternatively, we can use non-parametric specification to allow a fairly unconstrained evolution with time, which we choose to do here. This free function for the reproduction number is what gives the model its flexibility to fit a large number of outbreak types. As is common for other non-parametric methods, we will impose conditions on the rate of change of the reproduction number with time to ensure that the evolution of the reproduction number is sensible and not jumping around too much, in order to avoid overfitting.

To forecast  $I_t$ , we first need to forecast  $R_t$  from data prior to  $t$ . In this study we will use an approach where to forecast  $R_t$ , we first estimate the values of the reproduction number prior to  $t$  using the method described below. We then use these estimated values to predict the future evolution of  $R_t$ . To estimate the reproduction number from the historical incidence data, we use the method developed in [33; 37]. This approach used the same formulation as in Equation 1 to estimate the reproduction number from daily incidence. If the reproduction number is estimated daily, it is likely to vary too much from day to day in a manner that is

unrealistic. Therefore the method averages over the last 7 days to get more stable estimates. A Bayesian procedure is used to estimate both the best fit value and the credible region of the estimate. We use the R-package EpiEstim [38] to estimate the reproduction number using a parametric gamma distribution for the serial interval with the mean of 15.3 days and a standard deviation of 9.3 days [4].

Once we have calculated the historical values of  $R_t$ , we can use them to calculate a forecast for  $R_t$  that can be used for predictions of new incidence. We will use two different procedures to forecast the reproduction number. The first method is a simple method where we assume that the reproduction number remains constant from the last historical value. We use the uncertainties given by the method in [33] as a prediction interval. In practice this means that to predict the value of  $R_t$ , we use the method above to find the parameters of the posterior distribution for  $R_{t-1}$ . We then get realisations of  $R_t$  by drawing random numbers from this posterior distribution. When we use a gamma-prior distribution and have no uncertainty in the serial interval parameters, the posterior distribution is also a gamma distribution [33].

The second approach to forecast the reproduction number is based on Bayesian structural time series model fitted to the historical values of  $R_t$ . We use a model with a semi-local linear trend to allow the estimation of a trend in the recent data. To ensure that predicted values of  $R_t$  are between 0 and 15, we fit the time series on a transformed scale

$$r_t^* = \log \left( \frac{R_t}{15 - R_t} \right). \quad (2)$$

We use the R-package BSTS [39] to fit the Bayesian structural time series with the following model:

$$r_{t+1}^* = r_t^* + \delta_t + \epsilon_t, \epsilon_t \sim N(0, \sigma_\mu),$$

$$\delta_{t+1} = D + \phi(\delta_t - D) + \eta_t, \eta_t \sim N(0, \sigma_\delta).$$

$\delta_t$  is the semi-local trend that we model as an auto-regressive process of order one. This gives a trend process that can oscillate around the level  $D$  with smaller random-walk excursions. Inverse gamma-priors are used for the standard deviation parameters  $\sigma_\mu$  and  $\sigma_\delta$ , a Gaussian prior on  $D$  and a normal  $N(0, 0.1)$  prior for  $\phi$  that is truncated at -1 and 1. The  $\phi$  parameter determines how much the trend behaves as a random walk. For  $\phi = 1$  the trend follows a random walk, while for  $\phi = 0$  the trend is just constant with Gaussian noise. We use a regularising prior for small values of  $\phi$  to keep the model from overfitting by e.g having a very rapid growth in the variance of  $r^*$  that would make the model very over-dispersed. A Markow Chain Monte Carlo (MCMC) algorithm is used to estimate the parameters in the model which then allows us to forecast future values of  $r^*$ . After the model is fitted for all the values up to  $r_{t-1}^*$ , we can draw samples of the predictive distribution for  $r_t^*$ . After using the inverse transform of Equation 2 to get predictions for  $R_t$ , we can use these values to forecast incidence.

### 3.3.3 Forecasting incidence

Once we have specified our model by specifying the offspring distribution and the method for forecasting  $R_t$ , we can use the model to generate probabilistic forecasts. As in many other applications, the forecast is not available in a closed form, but we can simulate the model to get a sample of values from the predictive distribution. If we want to generate a forecast for  $I_t$ , we first use all the data up until and including time  $t - 1$  to estimate the history of the reproduction number until time. Our probabilistic forecast will be based on sampling possible outcomes to generate a distribution of outcomes. We therefore first draw a sample for  $R_t$  from the predictive distributions of the models discussed above, and then combine this with the historical incidence data to calculate the total force of infection  $\lambda_t$ . We then sample  $I_t$  from the specified offspring distribution with an expected value given by  $\lambda_t$  (Equation 1). The procedure is repeated many times to build up the predictive distribution. If we want to forecast over multiple time steps, we follow the same procedure by sampling values for

the reproduction number for  $t + 1$ . When calculating  $\lambda_{t+1}$ , we use the sampled value for  $I_t$  together with the historical data prior to  $t$ . This procedure can easily be iterated to allow forecasts for any forecasting horizon.

### 3.4 Assessing probabilistic forecasts

The aim of probabilistic forecasts is not only to predict the correct average value, but to forecast the correct probabilities for all the different possible outcomes. Therefore, when evaluating such forecasts, it is not sufficient to use metrics that only depend on a point estimate like the mean or median, such as the mean absolute error. To assess the forecasts of the models discussed in the previous section, we will follow the paradigm of maximising sharpness of the predictive distribution subject to calibration [12]. In addition we will consider proper scoring rules for comparing probabilistic forecasts. We follow the approach taken in [8], where probabilistic forecasts for the West African Ebola outbreak were assessed using similar methods.

#### 3.4.1 Calibration

A forecasting model is calibrated if the forecasted distribution is equal to the actual distribution of observed values. This means that the model can accurately assess its own uncertainty. For example, a well calibrated model that predicts that it will rain with a 60% chance should see that on average it does rain on 60% of those days. Mathematically, if we assume that the real distribution of outcomes in nature is given by a cumulative density function,  $G_t$ , and our model predicts a cumulative density function,  $F_t$ , we say that the forecast is ideal and therefore perfectly calibrated if  $F_t = G_t$ . To assess calibration, we will use a randomised Probability Integral Transform (PIT) [40]. We consider a forecasted cumulative distribution,  $F_t$ , compared to an observed value  $k_t$ . For each such pair of forecast and observed value we calculate

$$u_t = F_t(k_t) + \nu(F_t(k_t) - F_t(k_t - 1)),$$

where  $\nu$  is a standard uniform random variable. If the prediction is ideal, the values  $u_t$  will be distributed as a standard uniform distribution. Once we have calculated a set of  $u_t$  values, we can then use the Anderson-Darling test of uniformity to assess if the distribution of  $u_t$  is uniform. If the p-value for this test is less than 0.1, we find that there is evidence that the model is not calibrated, following the suggestion in [8]. We implement this using the goftest [41] R-package. It is important to note that uniform PIT values are a necessary but not sufficient condition for an ideal forecast and that since we only have limited data, we might not be able to reject calibration even if the model is not well calibrated. In addition to assessing calibration, a histogram of PIT values can tell us if the forecast is under or over-dispersed [40]. If the forecast is too dispersed (i.e. uncertainty is overestimated), the PIT values are likely to cluster at the centre of the PIT histogram, while if the forecast is under-dispersed (i.e uncertainty is underestimated), they are likely to cluster along the edges of the histogram. We introduce a simple measure of centrality, which is equal to the fraction of  $u_t$  values that are between 0.25 and 0.75 as a way to assess if the forecasts are under or over-dispersed if they are not calibrated

$$\text{centrality} = \frac{\text{Number where } 0.25 < u_t < 0.75}{\text{Total number}} - 0.5.$$

When the centrality score is less than 0, most of the PIT values are outside of the central region, suggesting that the forecasts are underestimating the real uncertainty. If the centrality score is larger than 0, then the PIT scores are mainly in the central region and this indicates that the forecasts are overestimating the amount of uncertainty. Since the PIT values include a random component, we will repeat the calculations of the PIT values 30 times for each set of forecasts and take averages to average out the effect of the randomness in the definition of  $u_t$ .

### 3.4.2 Sharpness and bias

Sharpness is defined as the range of values in the forecast. The sharper a forecast, the more certain the model is of the predicted values. Sharpness depends only on the forecast and not on the observed values. We will use the normalised absolute deviation about the median of  $I_t$  to quantify sharpness

$$S(I_t) = \frac{1}{0.675} \text{median}(|I_t - \text{median}(I_t)|),$$

where the normalisation factor means that  $S$  is equal to the standard deviation if  $F_t$  is normal.

It is also of importance to assess the bias of the forecast which indicates if we are more likely to predict too large or too small values. Following [8], we will quantity bias as

$$B_t(F_t, k_t) = 1 - (F_t(k_t) - F_t(k_t - 1)).$$

If  $B_t = 0$ , half the probability mass is above and half below the observed value, and the forecast is unbiased.  $B_t$  is always between -1 and 1, with both extreme values signifying a completely biased forecast where all the forecasted values are either smaller or larger than the observed value.

### 3.4.3 Proper scoring rules

Proper scoring rules have been developed to compare and rank probabilistic forecasts [42]. To be a proper scoring rule, the scoring rule has to be minimised when the forecast is ideal. These scoring rules combine calibration and sharpness and give a consistent ranking of forecasts. We will use the ranked probability score (RPS) and the Dawid-Sebastiani score (DSS) as implemented in the scoringRules R-package [43] for samples from the predictive

distribution. The RPS score is given by

$$RPS(F_t, k_t) = \sum_{i=0}^{\infty} (F_t(i) - \mathbb{1}(i \geq k_t))^2,$$

and can be viewed as an extension of the mean absolute error for a deterministic forecast. The DSS score only depends on the mean,  $\mu_p$ , and standard deviation,  $\sigma_p$ , of the predictive distribution

$$DSS(F_t, k_t) = \left( \frac{k_t - \mu_p}{\sigma_p} \right)^2 + 2 \log \sigma_p.$$

The DSS allows an intuitive understanding of the proper scoring rules. The first term tells us about how far away the observed value is from the mean of the predictive distribution in units of the standard deviation, and the second term gives information about the sharpness of the predictive distribution. For our models, we will only have samples from the predictive distribution. To calculate the RPS, a kernel density estimate is used to estimate  $F_t$ , and for the DSS we use the mean and standard deviation of the sample.

### 3.5 Implementation

All the models and assessments were implemented in the R programming language [44] and are available as open-source code at [45]. We will consider four different models to assess which model provides the best forecasts. The models are:

1. Model 1 (Poisson Latest): Constant reproduction number and Poisson offspring distribution
2. Model 2 (NegBin Latest): Constant reproduction number and negative binomial offspring distribution
3. Model 3 (Poisson Semi-local): Varying reproduction number with a semi-local trend and Poisson offspring distribution

#### 4. Model 4 (NegBin Semi-local): Varying reproduction number with a semi-local trend and negative binomial offspring distribution

To determine the dispersion parameter  $k$  for the negative binomial models, we find the value that minimises the ranked probability score for the one-day-ahead predictions for the model at the national level using the simple negative binomial model (Model 2). We run the models both on the national level using data from the entire outbreak and on the sub-national level. For sub-national predictions, we run the model on each health zone separately, using only data from that health zone. This corresponds to an assumption that once a case has been imported into a health zone, the outbreak in the health zone is independent of the rest of the outbreak.

To evaluate a model, we will estimate the calibration, sharpness, bias, DSS score and RPS for daily forecasts with forecasting horizons up to 28 days. To do this, we start 17 days<sup>1</sup> after the first case of the outbreak in the given location and calculate the  $d$ -days-ahead prediction for all historically available data. For the calibration, we use all forecasts for the  $d$ -days-ahead forecasts compared to the observed values to assess if PIT values are uniformly distributed, while for all the other metrics we average them over all the available start days for the forecast.

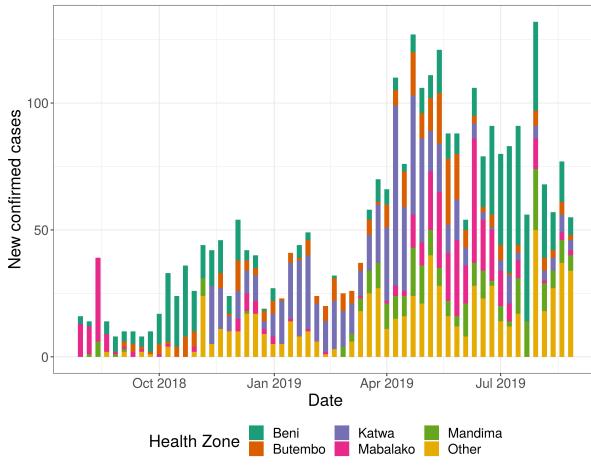
To evaluate the overall quality of the health zone forecasts, we first combine the calibration and scores from each health zone. For the proper scoring rules, sharpness and bias, we average the  $d$ -day-ahead prediction scores over all the health zones, and for the centrality and calibration based on the randomised PIT, we combine all the  $u_t$  values from all the health zones and calculate one combined centrality and calibration score for each  $d$ -day-ahead prediction. In addition we evaluate the health zone forecasts by adding up the predictions from each health zone and compare that to the national-level data.

---

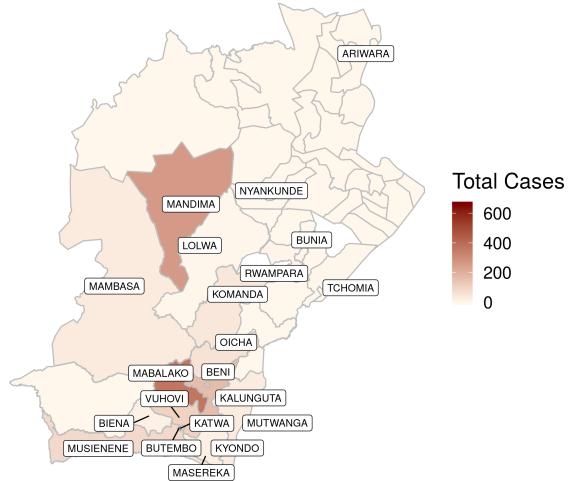
<sup>1</sup>The 17 days are due to the fact that the method for estimating the reproduction number only starts providing estimates of the reproduction number after a number of days that is equal to the mean of the serial interval plus one. Since we used a serial interval with mean 15.3, this gives 17 days.

## 4 Results

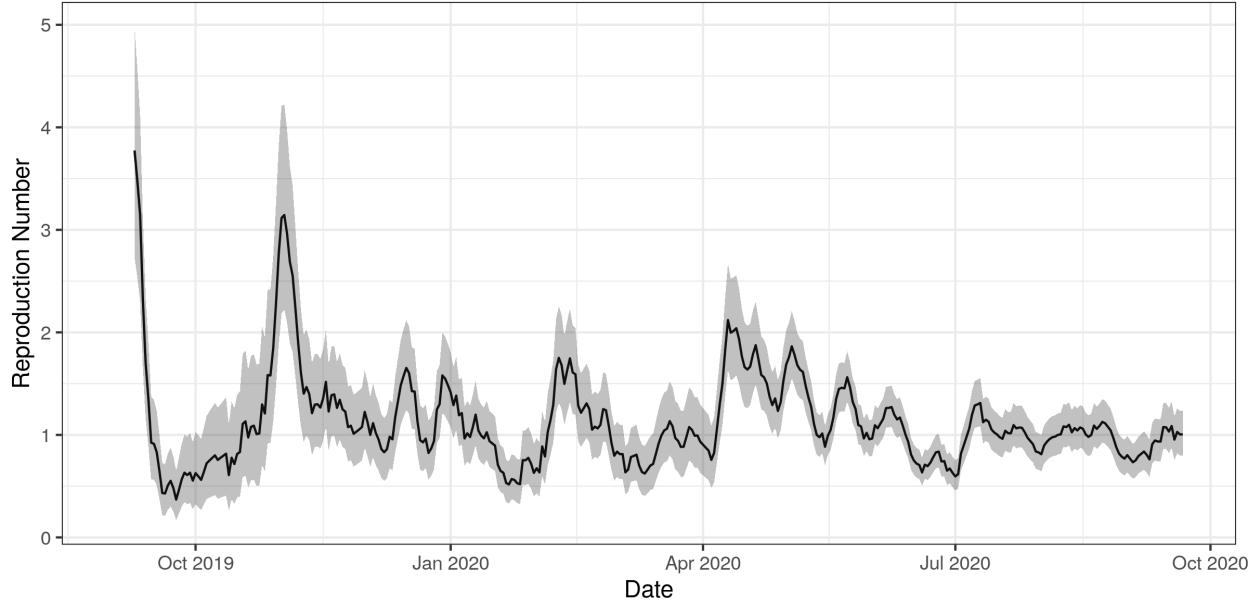
A total of 2,926 confirmed Ebola cases and 1,926 confirmed Ebola deaths have been reported in the ongoing outbreak in DRC as of the 31st of August 2019, giving a case fatality ratio of 66%. Figure 2a shows the weekly number of confirmed cases and the map in Figure 2b shows the total number of cases from each health zone. From Figure 2a and the estimate of the instantaneous reproduction number in Figure 2c, we can see the evolution of the outbreak. After the initial period with varying weekly case numbers, there was an increase in cases from October 2018 with a large reproduction number, followed by a more varied period where the reproduction number stayed around 1 with some waves of larger weekly incidence around December 2018 and February 2019. From late March and early April 2019, the number of cases per week increased significantly as the reproduction number rapidly increased to over 2. The larger number of weekly cases continued in June and July even as the reproduction number decreased. From July 2019 the reproduction has returned to around one, indicating that the outbreak is still not under control, but that the intensity of the outbreak is less than in May and June. There are large regional differences in the number of cases. Beni and Katwa health zones have been hardest hit sharing close to half the cases between them. From the epi-curve there seems to be some connection between the large increases in incidence and introduction or reintroduction of Ebola into specific health zones. The outbreak started mainly in Mabalako, then in October 2018 the peak was due to increased transmission in Beni, while the peak of incidence in February 2019 was mainly dominated by cases in the health zone of Katwa. The large increase of cases starting around March/April 2019 saw an increasing number of health zones reporting new cases. From this it is clear that the hot spot of the disease is moving from health zone to health zone.



(a) Number of new confirmed Ebola cases by week and by health zone.



(b) Total number of confirmed cases by health zone.



(c) Mean estimated reproduction number as a function of time, with the shaded region showing the 95% credible interval.

Figure 2: The three figures show the number of cases and the reproduction number as a function of time and health zone.

## 4.1 National-level forecasts

The first step in modelling the outbreak on the national level was to find the best fitting value of the dispersion parameter for the negative binomial models. We found that the dispersion

parameter that gives the smallest RPS score for one-day-ahead predictions for the negative binomial models was  $k \approx 8$ , but a fairly large range of parameters gave similar scores. We therefore chose to use  $k = 8$  for the negative binomial distributions both for the national and health zone forecasts.

All four of the models were fitted on incidence data from 395 days of national data and then used to provide short-term forecasts. The calibration and proper scoring rules scores for the four models for different forecasting horizons can be seen in Table 1 for weekly forecasting horizons up to four weeks. We also show the evolution of calibration, bias, RPS and the centrality of PIT values as a function of the forecasting horizon in Figure 3.

Model	Horizon (days)	Sharpness	Bias	RPS	DSS	Centrality	Calibration
NegBin Semilocal	1	4.08	-0.03	1.87	3.30	0.04	0.20
NegBin Latest	1	4.14	-0.00	1.82	3.26	0.06	0.16
Poisson Semilocal	1	2.93	0.01	1.87	3.32	-0.07	0.00
Poisson Latest	1	2.97	0.05	1.82	3.24	-0.08	0.00
NegBin Semilocal	7	4.83	-0.07	1.80	3.41	0.16	0.00
NegBin Latest	7	4.33	0.01	2.36	3.94	-0.03	0.00
Poisson Semilocal	7	4.19	-0.02	1.72	3.23	0.09	0.01
Poisson Latest	7	3.08	0.06	2.47	4.31	-0.14	0.00
NegBin Semilocal	14	5.98	-0.10	2.56	4.62	0.13	0.00
NegBin Latest	14	4.62	0.02	2.82	4.74	-0.08	0.00
Poisson Semilocal	14	5.78	-0.06	2.52	4.54	0.10	0.00
Poisson Latest	14	3.28	0.09	2.99	5.32	-0.22	0.00
NegBin Semilocal	21	7.55	-0.13	3.24	5.69	0.15	0.00
NegBin Latest	21	5.12	0.01	4.07	5.57	-0.19	0.00
Poisson Semilocal	21	7.67	-0.10	3.26	5.88	0.15	0.00
Poisson Latest	21	3.53	0.06	4.54	6.85	-0.26	0.00
NegBin Semilocal	28	9.14	-0.16	4.77	7.43	0.17	0.00
NegBin Latest	28	5.87	0.03	5.30	7.38	-0.20	0.00
Poisson Semilocal	28	9.64	-0.13	4.89	7.12	0.16	0.00
Poisson Latest	28	3.88	0.08	6.00	9.08	-0.28	0.00

Table 1: Model evaluations for predictions with different forecasting horizons for the national level.

Together they show that for next-day forecasts, the negative binomial offspring distribution is needed for a calibrated forecast as only the two negative binomial models are

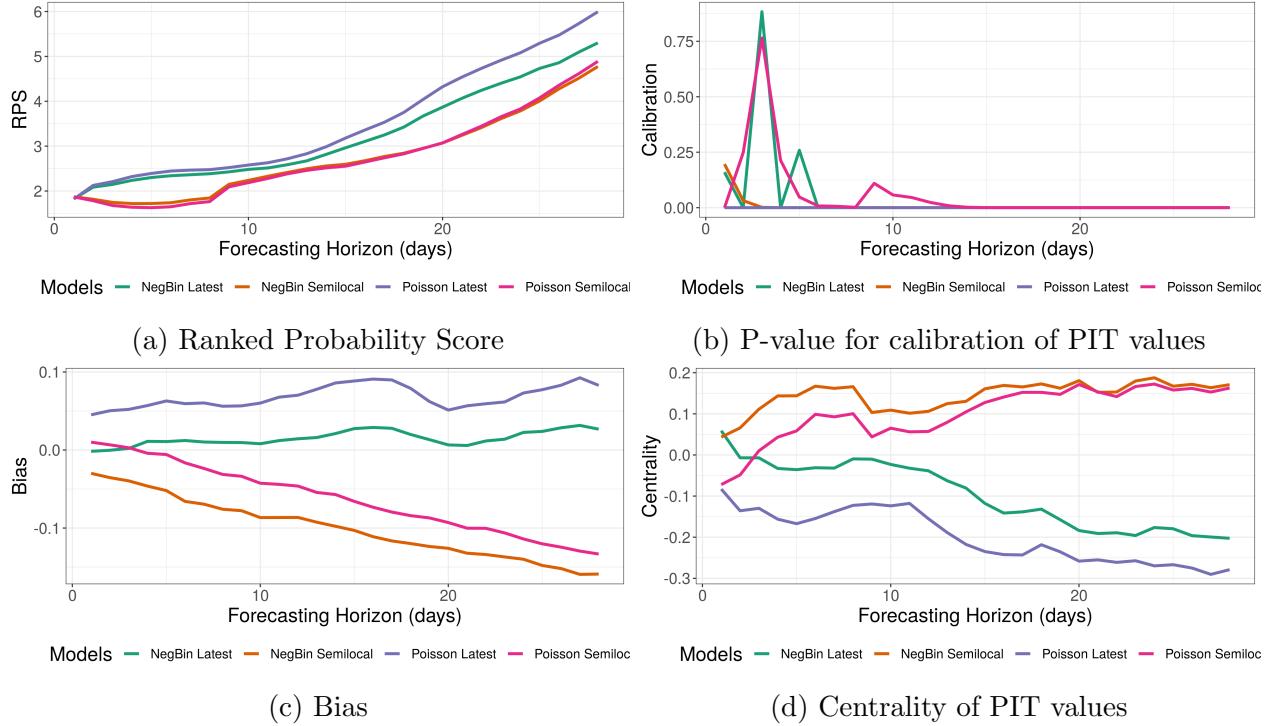


Figure 3: Evaluation of the four models as a function of forecasting horizon at the national level.

calibrated with a p-value larger than 0.1. We can also see that there are no big differences for one-day-ahead forecasts based on the method of forecasting the reproduction number. For longer forecasting horizons, the Poisson distribution with the semi-local time series prediction for  $R_t$  is the best model and is calibrated up until a maximum of nine days even if it is not well calibrated for all the days before. This is the only model that is calibrated for longer forecasting horizons and it also has the lowest RPS and DSS scores. The Poisson model with a constant reproduction number could not provide any calibrated forecasts, the negative binomial model with a constant reproduction number was calibrated during the first 5 days, and the negative binomial model with a semi-local trend in the reproduction number was only calibrated for the first day.

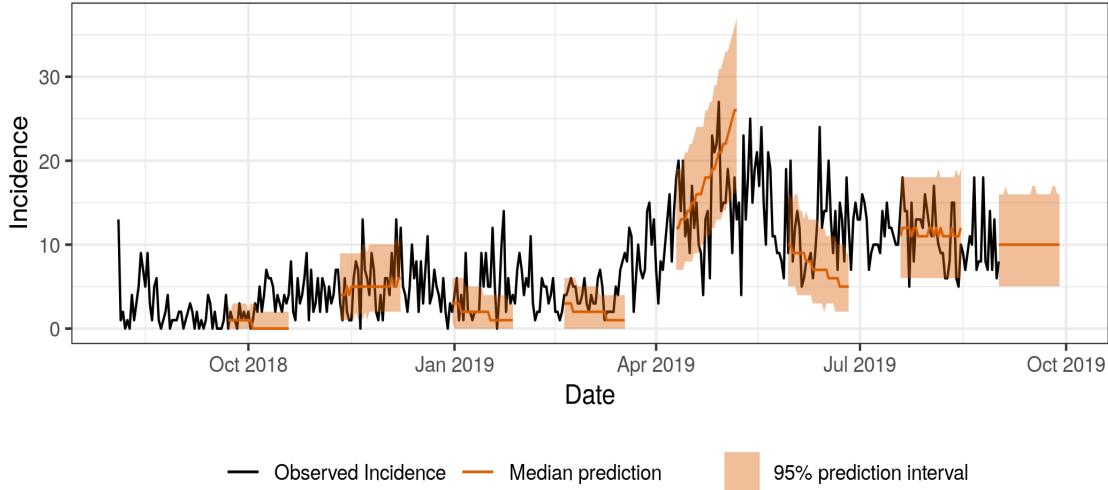
The centrality of the PIT scores in Figure 3d gives us extra information about why the models become uncalibrated. A negative centrality means that the model has too little uncertainty, and we can see that this is the case for the two models based on using the

latest reproduction number. The two models using the semi-local trend for the reproduction number are closer to zero centrality, but they have too much uncertainty since their centrality score is above zero. For the same method of forecasting the reproduction number, the negative binomial models always have a larger uncertainty, as we can also see from the sharpness figures. This is why the Poisson model is preferred at larger forecasting horizons. After about one week, the sharpness of the Poisson semi-local model has grown to be the same as the negative binomial model with a constant reproduction number. We can see how the semi-local models quickly become less sharp, while the sharpness of the constant models does not change as much with the forecasting horizon. The two semi-local models have a small negative bias indicating that on average they underestimate the incidence.

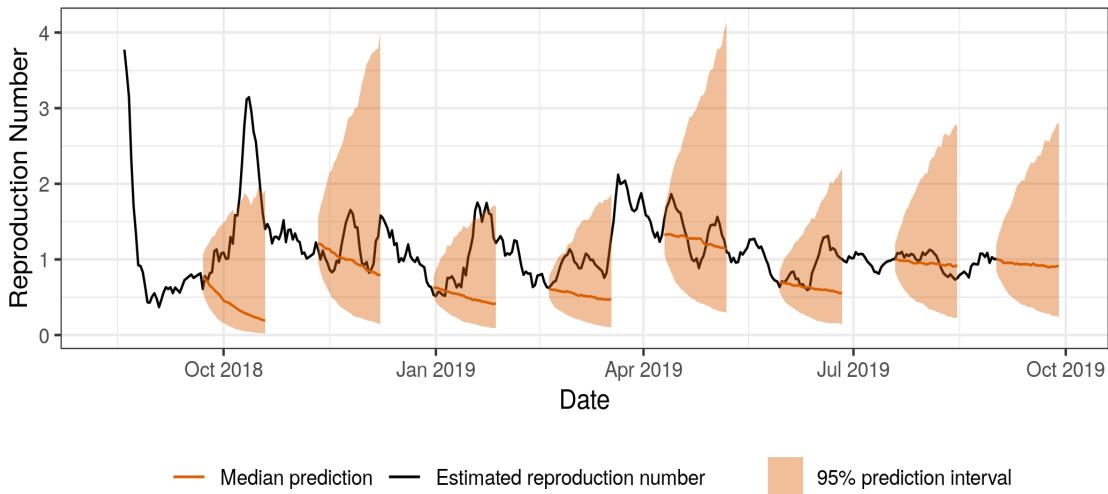
In Figure 4 we show 28-day forecasts every 50 days for the semi-local Poisson model and the corresponding forecasts for the reproduction number. These results show the same pattern as the more formal evaluations. In the short term the model works quite well, but the amount of uncertainty often seems a bit too large for the reproduction number forecasts. We also forecast the incidence for the 28 days following the 1st of September 2019 in Figure 4. Here the model predicts a slightly decreasing median reproduction number and a fairly steady incidence through the month of September. We expect between 10–15 cases per day over the next month.

## 4.2 Sub-national forecasts

We fitted all four models separately for each health zone that had more than one case of Ebola. The performance of the models varied significantly from health zone to health zone. Table 2 shows the best model(s) for each health zone, how large a forecasting horizon we can have without ruling out calibration, and the total number of confirmed cases in the health zone. There is a large difference ranging from health zones where the forecasts are calibrated for 28 days to health zones where not even the one-day-ahead predictions are calibrated. Even if the model is calibrated at 28 days, it does not mean that it is calibrated for all



(a) Forecasted and predicted incidence for the semi-local Poisson model



(b) Forecasted and predicted reproduction numbers for the semi-local Poisson model

Figure 4: Median forecast with 95% prediction intervals and observed values for incidence and reproduction number for the semi-local Poisson model.

forecasting horizons less than 28 days.

For all health zones apart from two, we find that the best model is a model with a semi-local trend in the reproduction number. Sometimes the Poisson models are better and sometimes the negative binomial models are better. The health zones where we get very well calibrated models at 28 days are health zones with small bursts of Ebola activity followed by long periods of no cases. Tchomia health zone is a good example of this and the forecasts

Location	Largest Horizon	Best model	Cases
Kayna	28	NegBin Semilocal	22
Nyiragongo	28	NegBin Semilocal	3
Tchomia	28	NegBin Semilocal	2
Lolwa	26	All models	3
Mwenga	18	All models	6
Bunia	17	NegBin Semilocal	4
Alimbongo	12	NegBin Semilocal	5
national	9	Poisson Semilocal	2942
Kayina	7	Poisson Semilocal	10
Rwampara	6	NegBin Latest	8
Lubero	4	NegBin Semilocal	31
Musienene	4	Poisson Semilocal	84
Kyondo	3	Poisson Semilocal	22
Vuhovi	3	Poisson Semilocal	103
Beni	2	Poisson Semilocal	661
Biena	2	Poisson Latest	16
Butembo	2	NegBin Semilocal	279
Mabalako	2	NegBin Semilocal	371
Kalunguta	NA	No calibrated model	164
Katwa	NA	No calibrated model	647
Mangurujipa	NA	No calibrated model	20
Masereka	NA	No calibrated model	50
Mutwanga	NA	No calibrated model	31
Oicha	NA	No calibrated model	55
Komanda	NA	No calibrated model	43
Mambasa	NA	No calibrated model	32
Mandima	NA	No calibrated model	264

Table 2: For each health zone we show the maximal forecasting horizon where we cannot exclude calibration at the  $p=0.1$  level. If multiple models are equally calibrated, we choose the one with smallest RPS. For some health zones there were no calibrated forecasts.

can be seen in the Appendix. The models provide calibrated forecasts for some of the larger health zones where there is more sustained transmission, but not for as long forecasting horizons as on the national level. There are also some health zones with a large number of cases where the models do not work well. We show the predictions and the centrality measures for the Beni and Katwa health zones in Figure 5. The forecasts and evaluations for a selected number of other health zones can be seen in the Appendix. For both Beni and Katwa, the main reason the models are not well calibrated is that they overestimated

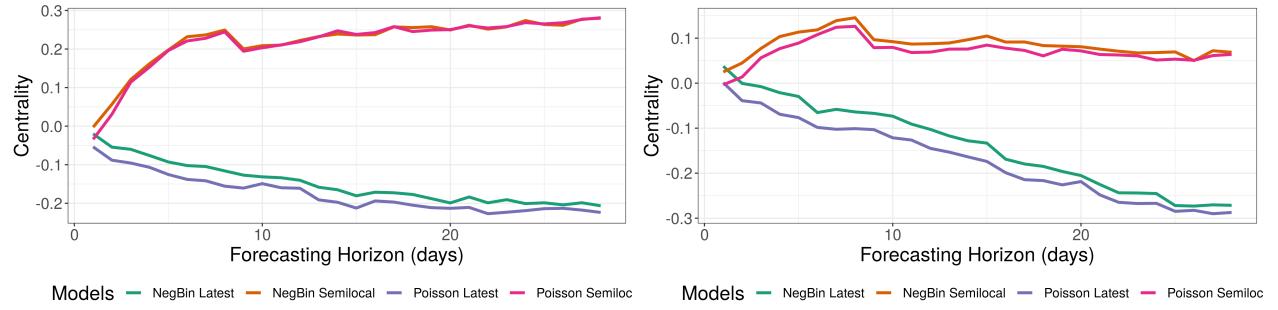
their own uncertainty. This is especially the case in Katwa and explains why even the one-day-ahead prediction is not well calibrated.

We used two methods to assess if the health zone model overall provided good forecasts. The first method was to combine the scores and PIT values, and we found that this gave clear evidence that the overall health zone model was uncalibrated at all forecasting horizons with p-values less than 0.001. The Ranked Probability Score and centrality from this combination of the scores in Figure 6 shows that on the combined level the models with semi-local trend produce significantly better forecasts, but overestimate the uncertainty as they did on the national level. The combined bias also showed a positive bias indicating that on average the models forecast too large incidence.

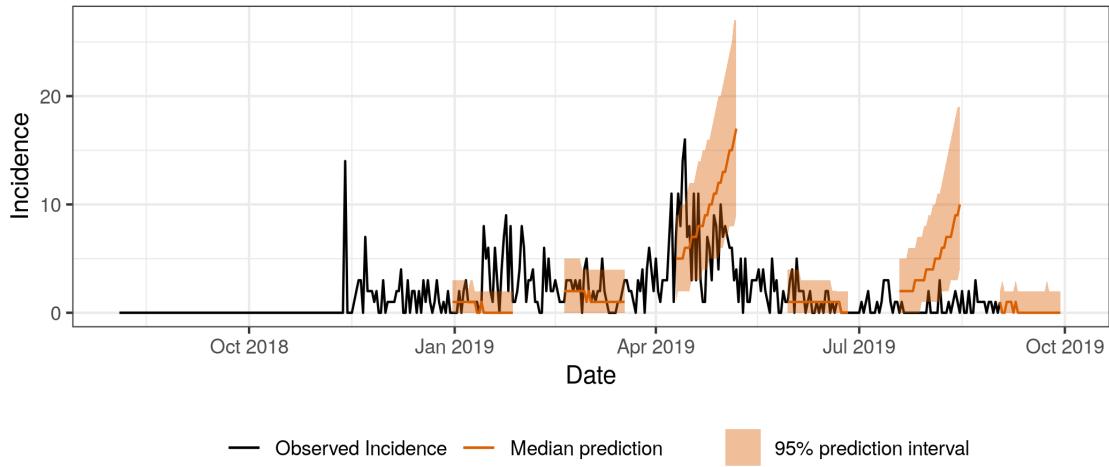
The second method to check the overall performance of the health zone model was by aggregating the forecasts from each health zone and comparing that to the national forecast model. Figure 7 shows the comparison of the one-day-ahead predictions of the national-level model with the one-day-ahead predictions obtained by aggregating over all health zones when we used the semi-local Poisson model for all the locations. The predictions are fairly close, but when looking at the calibration we find that the combined model is not calibrated and is also biased to predict too large incidence. The centrality scores for the combined models show that it actually underestimates the uncertainty at large forecasting horizons.

## 5 Discussion

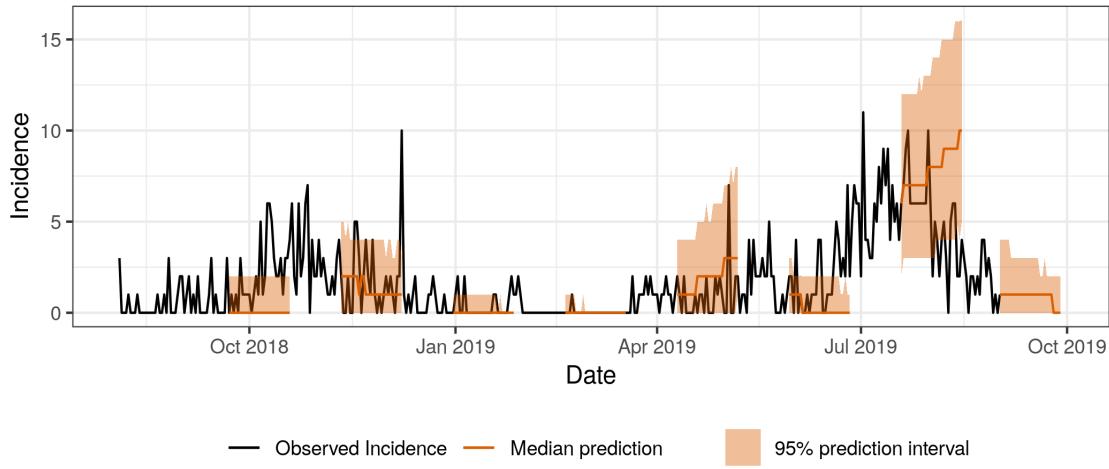
The ongoing Ebola outbreak in north-eastern DRC is the second largest Ebola outbreak in history and is of significant regional and international concern. As of the 1st of September 2019, the outbreak is not under control, as we can see both from the fact the reproduction number is not below one and that there are still plenty of new cases not on contact lists and multiple community deaths [46]. In addition to the large number of confirmed Ebola cases, this outbreak has seen a high CFR of 66% compared to 51% in the West African outbreak



(a) Centrality of PIT values for the models in Katwa health zone (b) Centrality of PIT values for the models in Beni health zone



(c) Median forecast with 95% prediction intervals and observed values for incidence for the Poisson semi-local model in Katwa



(d) Median forecast with 95% prediction intervals and observed values for incidence for the Poisson semi-local model in Beni

Figure 5: Evaluations and prediction for Katwa and Beni.

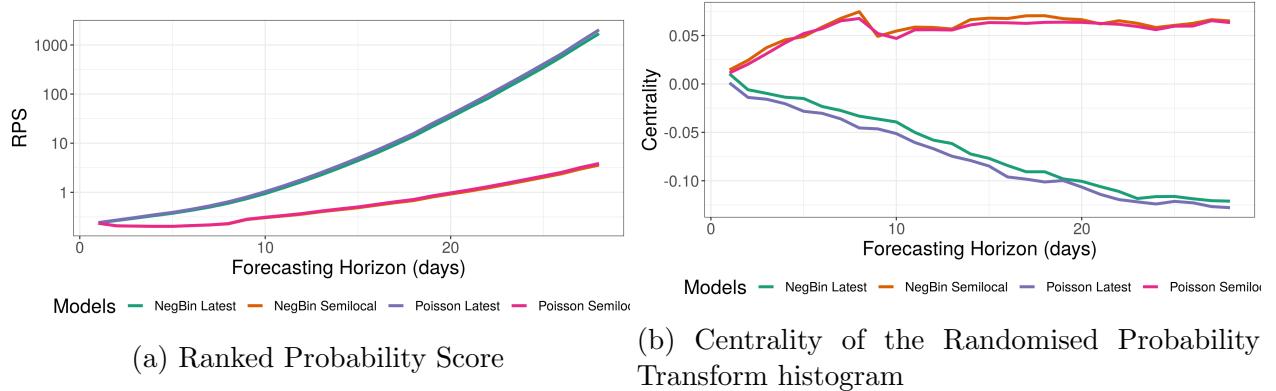


Figure 6: Ranked Probability Scores and centrality scores for the combined health-zone-level forecasts. RPS calculated as an average of scores from individual health zones and the centrality by combining all the randomised PIT values into one histogram.

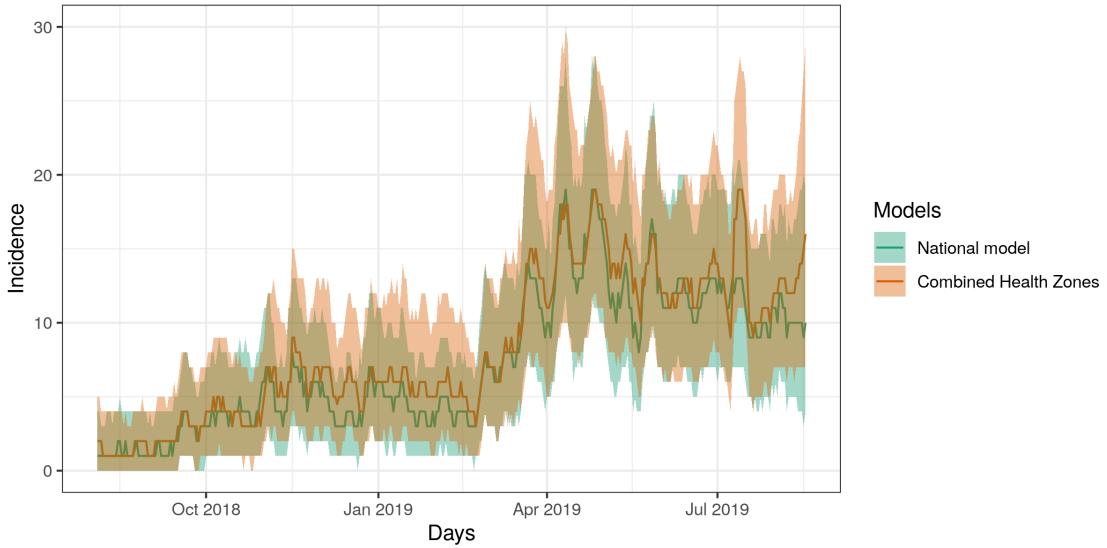


Figure 7: Median and 95% prediction intervals for one-day-ahead predictions from the Poisson semi-local model for national level compared with the same model when summing up predictions from all health zones.

[47]. Both the long duration, difficulty of control and high CFR are likely due to the difficult context for the outbreak response. The affected area of DRC has a history of violent conflicts and mistrust of the central government and international organisations. This makes all parts of the response from finding cases and tracking contacts to ensuring early treatment and safe burials much more difficult. Based on the models investigated in this paper, it seems that the large increase of cases seen from March to June 2019 is over and there is hope that the

outbreak can be controlled. However, the large uncertainty associated with the outbreak which is also reflected in the large uncertainty in the models makes long-term predictions very difficult if not impossible.

Compared to the estimates of the reproduction number in the West African Ebola outbreak [48], the reproduction number in the DRC outbreak has seen larger variations on short timescales and less clear patterns over longer timescales. In the previous outbreak, both Guinea and Liberia saw that the reproduction number peaked around two after a few months, followed by a long and slow decline. The current outbreak has been characterised by multiple short-term peaks or waves of higher reproduction number, with the peaks reaching over 3 at certain times. It is possible that the different patterns from the different outbreaks can be explained by the characteristics of the outbreak response. The West African response was slower to get properly started, but once it did, the response was fairly uninterrupted. In the current outbreak, the response had a faster start, but has been interrupted due to the security situation which could explain the short term-peaks in reproduction number. The range of reproductive values seen is also similar to other previous Ebola outbreaks [16].

We find that on the national level the models with a semi-local trend for the reproduction number were calibrated out to a forecasting horizon of 9 days and significantly outperformed the simpler models based on a non-changing reproduction number. The length of the forecasting horizon is similar, but a bit shorter than the 2 weeks found in [8], with a semi-mechanistic model based on a SEIR model that included reporting delays. While direct comparisons are not possible between different outbreaks, it would be of interest if our model with much less deterministic structure can give fairly similar calibrated forecasting horizons. Other models have been used to forecast the current Ebola outbreak, e.g. [49; 50]. They use very different modelling techniques, and an appropriate comparison between the models would require comparing proper scoring rules scores using the same data. All the models presented in this paper can easily be run daily on a normal laptop to provide real-time predictions in an outbreak setting.

The best model for the national outbreak is uncalibrated when forecasting more than 9 days forward because it overestimates the amount of uncertainty in the incidence. Depending on how the forecasts are used, it is possible that uncalibrated but over-dispersed forecasts can still be useful for planning purposes since they are conservative. An upper limit for an over-dispersed forecasts will give a conservative upper limit for future incidence. Forecasts based on under-dispersed models require a lot more care and should be treated with caution. Models that only provide point predictions or do not actively try to quantify the uncertainty are by definition under-dispersed. The main reason for the over-dispersion in the semi-local models is likely the forecasting of the reproduction number. This model allows a large variation in the reproduction number over a fairly short time, but does not have any information about what range of reproduction numbers is actually feasible. In this outbreak we see fairly fast changes of the reproduction number, but the reproduction number is almost never larger than three. A model for the reproduction number than allows this fast variation and short peaks, but keeps the reproduction number within a more narrow band, might provide better forecasts.

To compare forecasts and evaluate models for probabilistic forecasts, tests for calibration and the use of proper scoring rules is a crucial. The proper scoring rules are mainly useful for model comparissons, while the calibration gives an absolute measure for how well the forecast captures the uncertainty seen in the process. In this study, we found these tools crucial both for model selection and for the final interpretation of our model. Even a well calibrated model can make large forecasting errors if the underlying process changes in a way not seen during the time period used to fit and evaluate the model. Therefore it is always crucial to evaluate critically both the models and the metrics used to evaluate the models. We find that both the RPS and DSS scores give similar conclusions, but that calibration using the PIT values was especially valuable as an absolute measure of calibration. The introduction of the centrality measure of the PIT histograms gives additional valuable information and allows us to understand better how our models are failing if they are not well calibrated. The

centrality is a fairly crude measure of this, and measures that also include other departures from uniformity of the PIT values could give additional information.

We find that for one-day-ahead forecasts the negative binomial distribution is needed for a calibrated forecast. This indicates that a Poisson offspring distribution where everyone is always equally likely to transmit the disease in every small time step does not adequately describe the transmission process. We found a relatively large value for the  $k$  parameter which gives a fairly modest increase in the variance, or degree of superspreading. This increase in variance is much smaller than what was found when investigating transmission chains in the West African Ebola outbreak [36]. It is possible that the variance is smaller in this outbreak, but this study should be taken as fairly weak evidence and a proper investigation of transmission chains would be needed.

We could provide forecasts for all health zones with more than one case, but the quality was varying. For some health zones where there were very few cases and then a longer period with no transmission, the model managed fairly well, but for most health zones with a large number of cases the models were again too over-dispersed to provide calibrated predictions. Once our model has seen no new infections in a few weeks, it will predict close to zero new cases with essentially zero sharpness, irrespective of the current reproduction number. This behaviour works well when there are no new imported cases, as we can see in some of these health zones. Some of the well calibrated health zone forecasts only saw their first Ebola case quite late in the outbreak, giving many fewer data points to test the model. This can make it hard to reject calibration even if the model is uncalibrated. For almost all health zones, the models with semi-local trend were better than the models with a constant reproduction number. This was also the case when we combined all the forecasts from all the health zones where the combined RPS showed that the semi-local-trend models performed significantly better. Overall, the health-zone-level forecasts were not well calibrated when we combined the PIT values or when we aggregated the forecasts to compare against the national level incidence. It is clear that our model works much better on the national than the sub-national

level.

The models and forecasts in this report have multiple limitations. First, using a data source that only includes daily updated cumulative case counts has many disadvantages compared to a line list of all cases. A line list would give a much more accurate data on of the date of the onset of disease, and it would remove the problem with negative incidence rates after corrections and any problems with having to interpolate incidence rates. It is a strength of the modelling approach that it can be used on less than ideal data and still give reasonable short-term forecasts. The stringent tests of calibration can give us confidence in the forecasts even without perfect data. This is of key importance in outbreak situations where having up-to-date line lists of cases can be difficult or impossible. In these situations, we could still hope that the approaches in this report could give reasonable forecasts.

The models in their current form also have several limitations. As previously discussed, the semi-local-trend models seem to predict too much variance in the reproduction number which leads to over-dispersed predictions. There are also clear limitations when the models are used on the health-zone level since the spread between health zones is not modelled. Each health zone is modelled independently as if it were the whole outbreak. We only model the spread of the disease after it has been introduced to a health zone. It is possible that this is partly the reason for the uncalibrated forecasts on the health zone level. At the national level we do not have this limitation, but it makes it very difficult to use the current model to study the spatial spread of the disease. Due to the flexibility of the model structure, it would be fairly straightforward to improve the models by incorporating an additional force of infection term that gives the force of infection from outside the health zone. There are multiple options for how to model the spatial dependence that would all easily fit into our model structure. This includes spread from adjacent health zones, using a gravity model where the amount of spread between health zones is based on the populations [51] or, if available, data on travel between health zones could be used. Without this spread term, the model cannot be used to forecast probabilities for the spread of Ebola to new health zones.

In addition, the current way of estimating the time-varying reproduction number requires at least 17 days of data, which means that we need at least 17 days of data in a health zone to be able to forecast.

In addition to improvements related to the spatial spread, it would be beneficial to combine the separate parts of the model into one generative model. This means that the estimation of the reproduction number could be included into the modelling of the incidence. One promising way of doing this is a full Bayesian generative model. One of the main benefits of a generative model is that it models the whole data generation process and would allow us to simulate whole outbreaks. For future work, it would also be very interesting to investigate different models for the time evolution of the reproduction number. A full Bayesian treatment would also allow us to model uncertainty in the parameters of the offspring distribution and in the serial interval. It could even include imputation for missing data using the general modelling framework presented here.

In this report we used a semi-mechanistic model with fairly few assumptions about the exact mechanism of disease transmission. The flexibility of the reproduction number in this model allows us to fit complicated incidence histories that could not have been easily fit by traditional compartmental models. The benefit of the structure we do have in the model is that we ensure that the model correctly models certain basic facts about the outbreak like the serial interval and that all the parameters are interpretable. More structure is likely to make the model better if the assumptions in the model are true, but will often make it impossible to get a good fit when the assumptions are wrong. The basic structure included in our models is also likely to describe a large number of other infectious diseases. Therefore, they could easily be adopted to other diseases and other contexts without any real changes in the model structure. We would need a different serial interval, but the rest of the model could be used as it is. Another benefit of adding more structure to a model, for example a SEIR structure is that it is easier to model the results of interventions like vaccination<sup>2</sup>. One key aspect

---

<sup>2</sup>Using the fact that  $R_t = R_o * s(t)$  in a compartmental model of the SEIR type, it would likely be possible to also model some simple interventions in the models considered in this report.

to consider when choosing models for an outbreak situation is that perhaps more than one model should be used during the outbreak. At the start of the outbreak when the important mechanisms of disease transmission and population behaviours are less well known, a model with less mechanistic structure is perhaps the most useful. Once we learn more about the outbreak, we can add more structure to the model. In many forecasting challenges it has been shown that an ensemble of models performs better than individual models [52; 53]. This was also found in the RAPPID modelling competition for Ebola [25]. Therefore, combining multiple models with different strengths could give a better overall forecast. For example, a combination of the negative binomial model with a constant reproduction number, which did well for early forecasts, with the Poisson semi-local model, which did better for slightly longer forecasts, could give an even better model.

Theoretically it should be possible to forecast infectious disease outbreaks on the timescale of individual outbreaks [54]. The models in this report and other disease forecasting models still have a long way to go to achieve this goal. For the current Ebola outbreak there are many additional sources of data that are not currently used in our models. These includes deaths, community deaths versus deaths in treatment centres, contacts followed up, vaccinations, data on response activities etc. It is possible that combining these additional data sources together with improved modelling techniques could significantly increase forecasting horizon where we can provide calibrated predictions.

## 6 Conclusions

We presented a probabilistic disease modelling and forecasting framework based on a modified branching process with a time-varying reproduction number. When we modelled the reproduction number with a semi-local trend, the models could give well calibrated forecasts for the 2018–2019 Ebola outbreak in DRC for forecasting horizons of nine days. The models provide easy-to-interpret forecasts of incidence and reproduction number. They could also

provide good incidence forecasts for some health zones at the sub-national level. At both the national and the sub-national level, forecasts based on a constant reproduction under-estimated the uncertainty and forecasts based on a semi-local trend for the reproduction number overestimated the uncertainty at longer forecasting horizons.

## References

- [1] Hans Heesterbeek, Roy M. Anderson, Viggo Andreasen, Shweta Bansal, Daniela De Angelis, Chris Dye, Ken T. D. Eames, W. John Edmunds, Simon D. W. Frost, Sebastian Funk, T. Deirdre Hollingsworth, Thomas House, Valerie Isham, Petra Klepac, Justin Lessler, James O. Lloyd-Smith, C. Jessica E. Metcalf, Denis Mollison, Lorenzo Pellis, Juliet R. C. Pulliam, Mick G. Roberts, Cecile Viboud, and Isaac Newton Institute IDD Collaboration. Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227):aaa4339, March 2015. URL: <https://science.sciencemag.org/content/347/6227/aaa4339>, doi:10.1126/science.aaa4339.
- [2] Caitlin Rivers, Jean-Paul Chretien, Steven Riley, Julie A. Pavlin, Alexandra Woodward, David Brett-Major, Irina Maljkovic Berry, Lindsay Morton, Richard G. Jarman, Matthew Biggerstaff, Michael A. Johansson, Nicholas G. Reich, Diane Meyer, Michael R. Snyder, and Simon Pollett. Using outbreak science to strengthen the use of models during epidemics. *Nature Communications*, 10(1):1–3, July 2019. URL: <https://www.nature.com/articles/s41467-019-11067-2>, doi:10.1038/s41467-019-11067-2.
- [3] Jean-Paul Chretien, Steven Riley, and Dylan B George. Mathematical modeling of the West Africa Ebola epidemic. *eLife*, 4. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769159/>, doi:10.7554/eLife.09186.
- [4] WHO Ebola Response Team. Ebola Virus Disease in West Africa The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine*, 371(16):1481–1495, October 2014. URL: <https://doi.org/10.1056/NEJMoa1411100>, doi:10.1056/NEJMoa1411100.
- [5] Anton Camacho, Adam Kucharski, Yvonne Aki-Sawyerr, Mark A. White, Stefan Flasche, Marc Baguelin, Timothy Pollington, Julia R. Carney, Rebecca Glover, Eliz-

- abeth Smout, Amanda Tiffany, W. John Edmunds, and Sebastian Funk. Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study. *PLoS Currents*, 7, February 2015. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4339317/>, doi:10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2.
- [6] P.-Y. Kobres, J. P. Chretien, M. A. Johansson, J. Morgan, P.-Y. Whung, H. Mukundan, SY Del Valle, B. M. Forshey, T. M. Quandelacy, M. Biggerstaff, C. Viboud, and S. Pollett. A systematic review and evaluation of Zika virus forecasting and prediction research during a public health emergency of international concern. *bioRxiv*, page 634832, May 2019. URL: <https://www.biorxiv.org/content/10.1101/634832v1>, doi:10.1101/634832.
- [7] Jean-Paul Chretien, Dylan George, Jeffrey Shaman, Rohit A. Chitale, and F. Ellis McKenzie. Influenza Forecasting in Human Populations: A Scoping Review. *PLOS ONE*, 9(4):e94130, April 2014. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094130>.
- [8] Sebastian Funk, Anton Camacho, Adam J. Kucharski, Rachel Lowe, Rosalind M. Eggo, and W. John Edmunds. Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15. *PLOS Computational Biology*, 15(2):e1006785, February 2019. URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006785>, doi:10.1371/journal.pcbi.1006785.
- [9] Wei Wei and Leonhard Held. Calibration tests for count data. *TEST*, 23(4):787–805, December 2014. URL: <https://doi.org/10.1007/s11749-014-0380-8>, doi:10.1007/s11749-014-0380-8.
- [10] Leonhard Held, Sebastian Meyer, and Johannes Bracher. Probabilistic forecasting in

- infectious disease epidemiology: the 13th Armitage lecture. *Statistics in Medicine*, 36(22):3443–3460, September 2017. URL: <https://onlinelibrary.wiley.com/doi/10.1002/sim.7363>, doi:10.1002/sim.7363.
- [11] Tilmann Gneiting. Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2):319–321, 2008. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2007.00522.x>, doi:10.1111/j.1467-985X.2007.00522.x.
- [12] Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2007.00587.x>, doi:10.1111/j.1467-9868.2007.00587.x.
- [13] Claudia Czado, Tilmann Gneiting, and Leonhard Held. Predictive Model Assessment for Count Data. *Biometrics*, 65(4):1254–1261, 2009. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2009.01191.x>, doi:10.1111/j.1541-0420.2009.01191.x.
- [14] World Health Organisation. Ebola virus disease fact sheet. URL: <https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease>.
- [15] Julii Brainard, Lee Hooper, Katherine Pond, Kelly Edmunds, and Paul R. Hunter. Risk factors for transmission of Ebola or Marburg virus disease: a systematic review and meta-analysis. *International Journal of Epidemiology*, 45(1):102–116, February 2016. URL: <https://academic.oup.com/ije/article/45/1/102/2363618>, doi:10.1093/ije/dyv307.
- [16] J. Legrand, R. F. Grais, P. Y. Boelle, A. J. Valleron, and A. Flahault. Understanding the dynamics of Ebola epidemics. *Epidemiology & Infection*, 135(4):610–621, May 2007.

URL: <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/understanding-the-dynamics-of-ebola-epidemics/BAECC049DE2893FA58DD8F38E19F62D8>, doi:10.1017/S0950268806007217.

- [17] World Health Organization. Ebola Outbreak in DRC 01: 7 August 2018. August 2018.
- [18] World Health Organisation. Ebola Outbreak in DRC 09: 04 October 2018. October 2018.
- [19] World Health Organization. Ebola Outbreak in DRC 16: 21 November 2018. November 2018.
- [20] World Health Organization. Ebola Outbreak in DRC 38: 24 April 2019. April 2019.
- [21] World Health Organization. Preliminary results on the efficacy of rVSV-ZEBOV-GP Ebola vaccine using the ring vaccination strategy in the control of an Ebola outbreak in the Democratic Republic of the Congo: an example of integration of research into epidemic response. Technical report, April 2019. URL: <https://www.who.int/csr/resources/publications/ebola/ebola-ring-vaccination-results-12-april-2019.pdf>.
- [22] National Institute of Allergy and Infectious Diseases. Independent Monitoring Board Recommends Early Termination of Ebola Therapeutics Trial in DRC Because of Favorable Results with Two of Four Candidates. August 2019.
- [23] World Health Organization. Ebola Outbreak in DRC 51: 23 July 2019. July 2019.
- [24] World Health Organization. Ebola Outbreak in DRC 58: 10 September 2019. July 2019.
- [25] Ccile Viboud, Kaiyuan Sun, Robert Gaffey, Marco Ajelli, Laura Fumanelli, Stefano Merler, Qian Zhang, Gerardo Chowell, Lone Simonsen, and Alessandro Vespignani. The RAPIDD Ebola Forecasting Challenge: Synthesis and Lessons Learnt.

- Epidemics*, 22:13–21, March 2018. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5927600/>, doi:10.1016/j.epidem.2017.08.002.
- [26] Robert H. Gaffey and Ccile Viboud. Application of the CDC EbolaResponse Modeling tool to disease predictions. *Epidemics*, 22:22–28, March 2018. URL: <http://www.sciencedirect.com/science/article/pii/S1755436517300555>, doi:10.1016/j.epidem.2017.03.001.
- [27] Bruce Pell, Yang Kuang, Cecile Viboud, and Gerardo Chowell. Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics*, 22:62–70, March 2018. URL: <http://www.sciencedirect.com/science/article/pii/S1755436516300433>, doi:10.1016/j.epidem.2016.11.002.
- [28] Manikandan M, Velavan A, Zile Singh, Anil J. Purty, Joy Bazroy, and Senthamarai Kannan. Forecasting the trend in cases of Ebola virus disease in West African countries using auto regressive integrated moving average models. *International Journal Of Community Medicine And Public Health*, 3(3):615–618, February 2017. URL: <https://www.ijcmph.com/index.php/ijcmph/article/view/764>, doi:10.18203/2394-6040.ijcmph20160620.
- [29] Jason Asher. Forecasting Ebola with a regression transmission model. *Epidemics*, 22:50–55, March 2018. URL: <http://www.sciencedirect.com/science/article/pii/S175543651730021X>, doi:10.1016/j.epidem.2017.02.009.
- [30] Aaron A. King, Matthieu Domenech de Cells, Felicia M. G. Magpantay, and Pejman Rohani. Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806):20150347, May 2015. URL: <https://royalsocietypublishing.org/doi/10.1098/rspb.2015.0347>, doi:10.1098/rspb.2015.0347.
- [31] Hummanitarian Data Exchange. Ebola Cases and Deaths in the North

Kivu Ebola Outbreak in the Democratic Republic of the Congo (DRC) - Humanitarian Data Exchange. URL: <https://data.humdata.org/dataset/ebola-cases-and-deaths-drc-north-kivu>.

- [32] Christine Jacob. Branching Processes: Their Role in Epidemiology. *International Journal of Environmental Research and Public Health*, 7(3):1186–1204, March 2010. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872325/>, doi:10.3390/ijerph7031204.
- [33] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013. doi: 10.1093/aje/kwt133.
- [34] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355, November 2005. URL: <https://www.nature.com/articles/nature04153>, doi:10.1038/nature04153.
- [35] Pierre Nouvellet, Anne Cori, Tini Garske, Isobel M. Blake, Ilaria Dorigatti, Wes Hinsley, Thibaut Jombart, Harriet L. Mills, Gemma Nedjati-Gilani, Maria D. Van Kerkhove, Christophe Fraser, Christl A. Donnelly, Neil M. Ferguson, and Steven Riley. A simple approach to measure transmissibility and forecast incidence. *Epidemics*, 22:29–35, March 2018. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5871640/>, doi:10.1016/j.epidem.2017.02.012.
- [36] International Ebola Response Team, Junerlyn Agua-Agum, Archchun Ariyarajah, Bruce Aylward, Luke Bawo, Pepe Bilivogui, Isobel M. Blake, Richard J. Brennan, Amy Cawthorne, Eilish Cleary, Peter Clement, Roland Conteh, Anne Cori, Foday Dafae, Benjamin Dahl, Jean-Marie Dangou, Boubacar Diallo, Christl A. Donnelly,

Ilaria Dorigatti, Christopher Dye, Tim Eckmanns, Mosoka Fallah, Neil M. Ferguson, Lena Fiebig, Christophe Fraser, Tini Garske, Lice Gonzalez, Esther Hamblion, Nuha Hamid, Sara Hersey, Wes Hinsley, Amara Jambei, Thibaut Jombart, David Kargbo, Sakoba Keita, Michael Kinzer, Fred Kuti George, Beatrice Godefroy, Giovanna Gutierrez, Niluka Kannangarage, Harriet L. Mills, Thomas Moller, Sascha Meijers, Yasmine Mohamed, Oliver Morgan, Gemma Nedjati-Gilani, Emily Newton, Pierre Nouvellet, Tolbert Nyenswah, William Perea, Devin Perkins, Steven Riley, Guenael Rodier, Marc Rondy, Maria Sagrado, Camelia Savulescu, Ilana J. Schafer, Dirk Schumacher, Thomas Seyler, Anita Shah, Maria D. Van Kerkhove, C. Samford Wesseh, and Zabulon Yoti. Exposure Patterns Driving Ebola Transmission in West Africa: A Retrospective Observational Study. *PLoS medicine*, 13(11):e1002170, November 2016. doi:10.1371/journal.pmed.1002170.

- [37] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P. A. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, and A. Cori. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, page 100356, August 2019. URL: <http://www.sciencedirect.com/science/article/pii/S1755436519300350>, doi:10.1016/j.epidem.2019.100356.
- [38] Anne Cori. *EpiEstim: EpiEstim: a package to estimate time varying reproduction numbers from epidemic curves*. 2013. URL: <https://CRAN.R-project.org/package=EpiEstim>.
- [39] Steven L. Scott. *bsts: Bayesian Structural Time Series*. 2019. URL: <https://CRAN.R-project.org/package=bsts>.
- [40] Claudia Czado, Tilmann Gneiting, and Leonhard Held. Predictive Model Assessment for Count Data. *Biometrics*, 65(4):1254–1261, 2009. URL: <https://>

[onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2009.01191.x](http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2009.01191.x), doi:10.1111/j.1541-0420.2009.01191.x.

- [41] Julian Faraway, George Marsaglia, John Marsaglia, and Adrian Baddeley. *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions*. 2017. URL: <https://CRAN.R-project.org/package=goftest>.
- [42] Tilmann Gneiting and Adrian E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, March 2007. URL: <https://doi.org/10.1198/016214506000001437>, doi:10.1198/016214506000001437.
- [43] Alexander Jordan, Fabian Krueger, and Sebastian Lerch. Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software*, 2018.
- [44] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL: <https://www.R-project.org/>.
- [45] Link to source code for Ebola modelling MSC project. Url removed as it could identify the student.
- [46] World Health Organisation. Disease Outbreak News 12th of September : Ebola virus disease Democratic Republic of the Congo. URL: <http://www.who.int/csr/don/12-september-2019-ebola-drc/en/>.
- [47] Amanda M. Rojek, Alex Salam, Robert J. Ragotte, Emily Liddiard, Ahmed Elhussain, Anna Carlqvist, Michael Butler, Nzelle Kayem, Lyndsey Castle, LangO Odondi, Kasia Stepniewska, and Peter W. Horby. A systematic review and meta-analysis of patient data from the west Africa (2013–16) Ebola virus disease epidemic. *Clinical Microbiology and Infection*, 0(0),

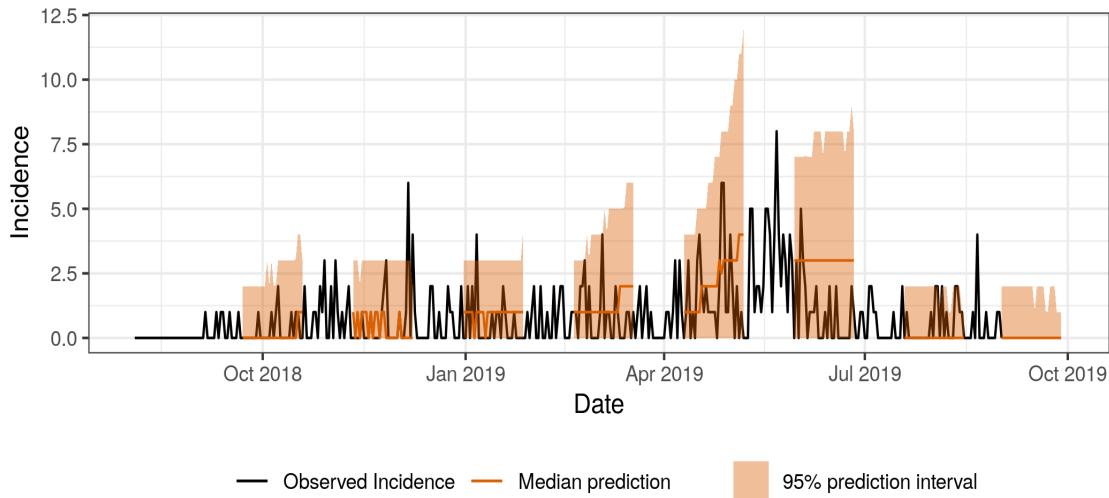
- July 2019. URL: [https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X\(19\)30387-8/abstract](https://www.clinicalmicrobiologyandinfection.com/article/S1198-743X(19)30387-8/abstract), doi:10.1016/j.cmi.2019.06.032.
- [48] West African Ebola Epidemic after One Year Slowing but Not Yet under Control. *New England Journal of Medicine*, 372(6):584–587, February 2015. URL: <https://doi.org/10.1056/NEJMc1414992>, doi:10.1056/NEJMc1414992.
- [49] J. Daniel Kelly, Junhyung Park, Ryan J. Harrigan, Nicole A. Hoff, Sarita D. Lee, Rae Wannier, Bernice Selo, Mathias Mossoko, Bathe Njoloko, Emile Okitolonda-Wemakoy, Placide Mbala-Kingebeni, George W. Rutherford, Thomas B. Smith, Steve Ahuka-Mundeke, Jean Jacques Muyembe-Tamfum, Anne W. Rimoin, and Frederic Paik Schoenberg. Real-time predictions of the 2018-2019 Ebola virus disease outbreak in the Democratic Republic of the Congo using Hawkes point process models. *Epidemics*, page 100354, July 2019. URL: <http://www.sciencedirect.com/science/article/pii/S1755436519300258>, doi:10.1016/j.epidem.2019.100354.
- [50] Andrei R. Akhmetzhanov, Hyojung Lee, Sung-mok Jung, Taishi Kayano, Baoyin Yuan, and Hiroshi Nishiura. Analyzing and forecasting the Ebola incidence in North Kivu, the Democratic Republic of the Congo from 201819 in real time. *Epidemics*, 27:123–131, June 2019. URL: <http://www.sciencedirect.com/science/article/pii/S1755436519300234>, doi:10.1016/j.epidem.2019.05.002.
- [51] Kingsley E. Haynes and A. Stewart Fotheringham. Gravity and Spatial Interaction Models. In *Book Chapters*, pages 1–48. Regional Research Institute, West Virginia University, 1985. URL: <https://ideas.repec.org/h/rri/bkchap/07.html>.
- [52] Thomas G. Dietterich. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*, Lecture Notes in Computer Science, pages 1–15. Springer Berlin Heidelberg, 2000.
- [53] Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Using Stacking to Aver-

- age Bayesian Predictive Distributions (with Discussion). *Bayesian Analysis*, 13(3):917–1007, September 2018. URL: <https://projecteuclid.org/euclid.ba/1516093227>, doi:10.1214/17-BA1091.
- [54] Samuel V. Scarpino and Giovanni Petri. On the predictability of infectious disease outbreaks. *Nature Communications*, 10(1):1–8, February 2019. URL: <https://www.nature.com/articles/s41467-019-08616-0>, doi:10.1038/s41467-019-08616-0.

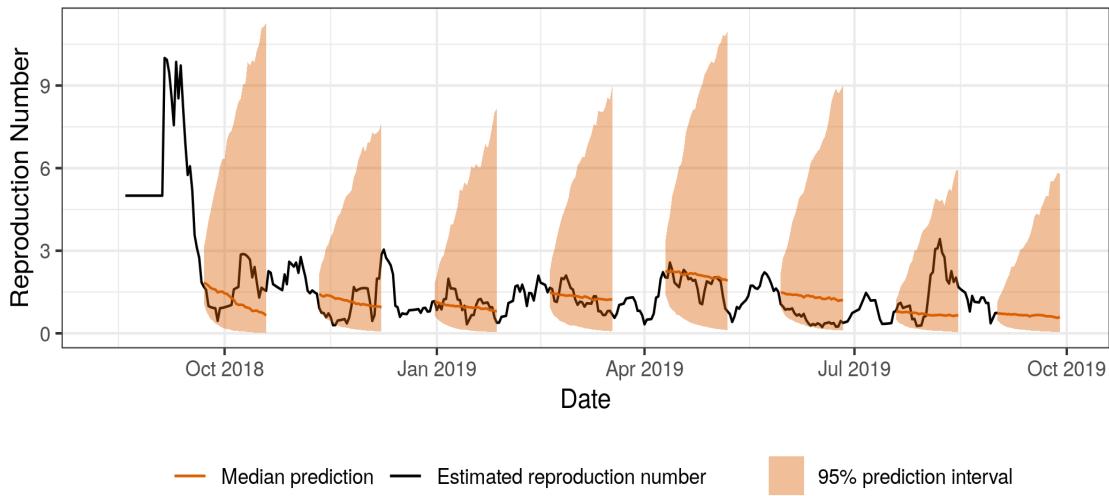
# Appendices

This appendix shows the best fitting model and the evaluation scores for a selection of health zones.

## A Butembo



(a) Forecasted and predicted incidence for the best fitting model



(b) Forecasted and predicted reproduction numbers for the best fitting model

Figure 8: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the best fitting model for Butembo.

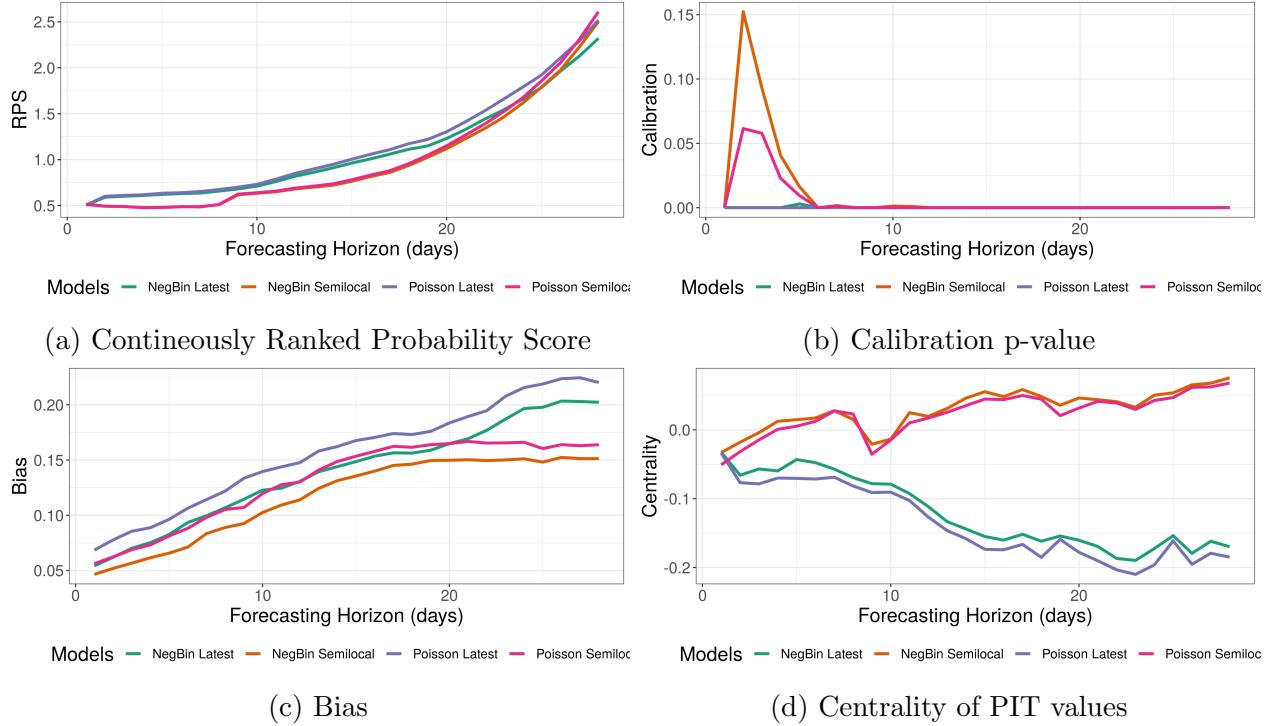
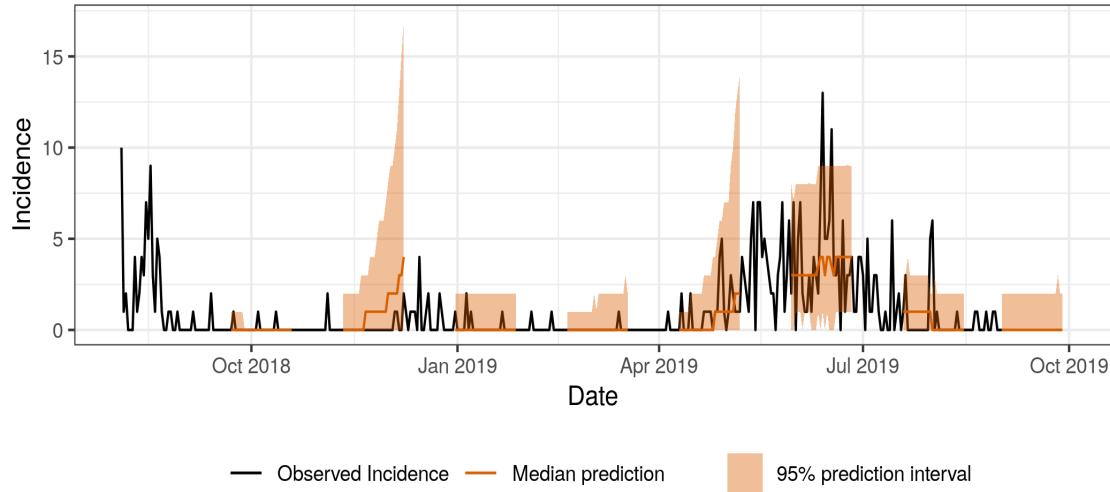
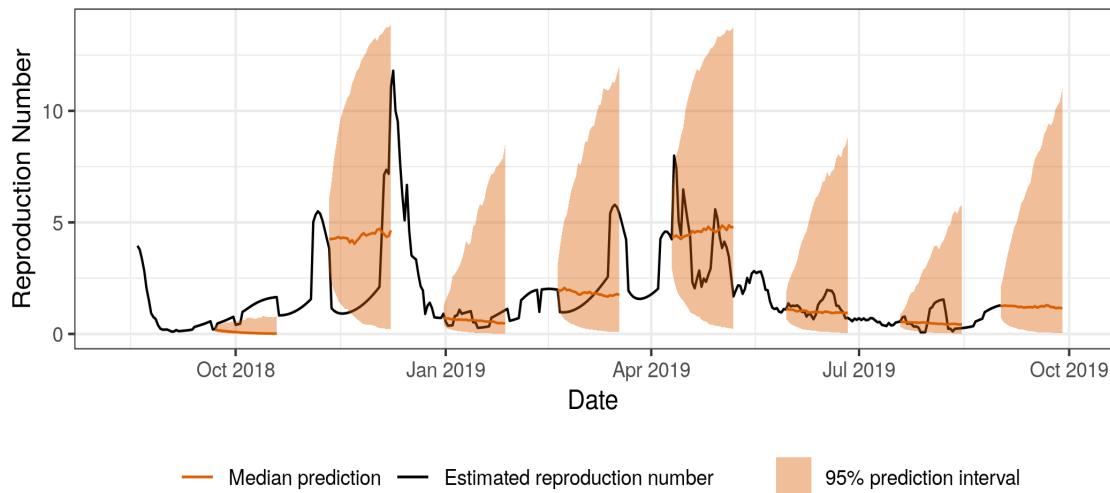


Figure 9: Scores for Butembo as a function of the forecasting horizon.

## B Mabalako



(a) Forecasted and predicted incidence for the best fitting model



(b) Forecasted and predicted reproduction numbers for the best fitting model

Figure 10: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the best fitting model for Mabalako.

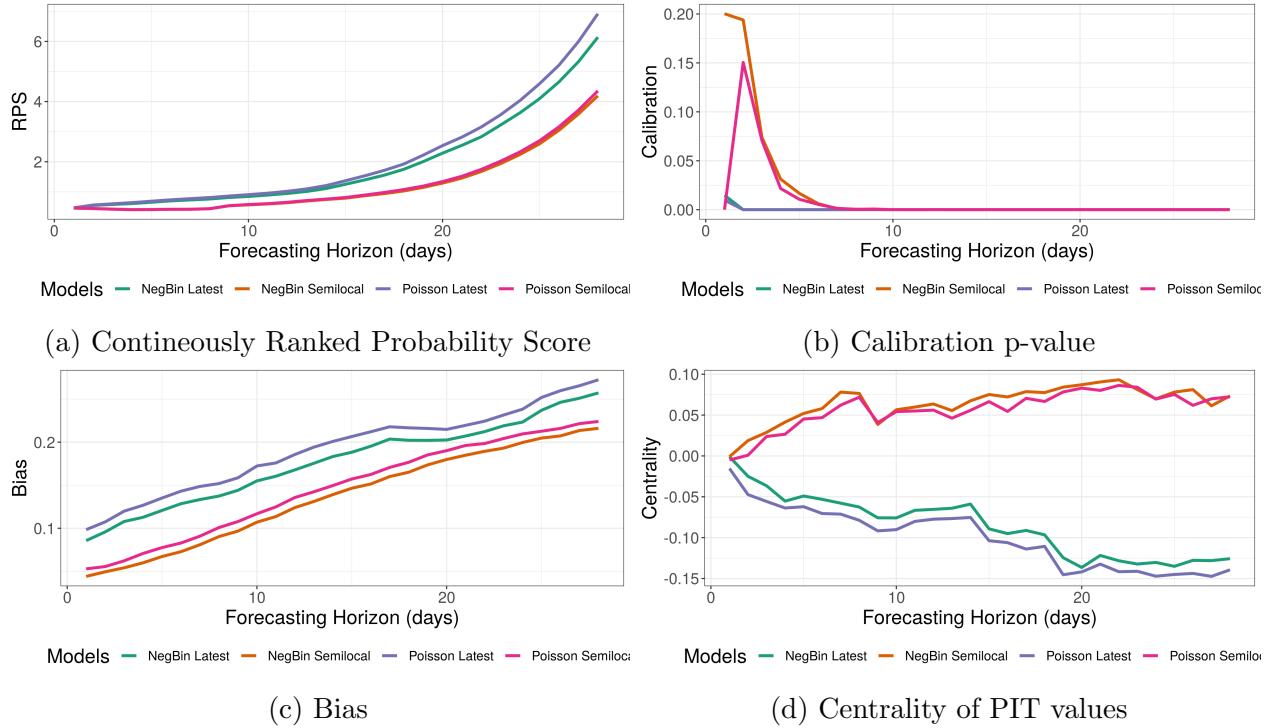
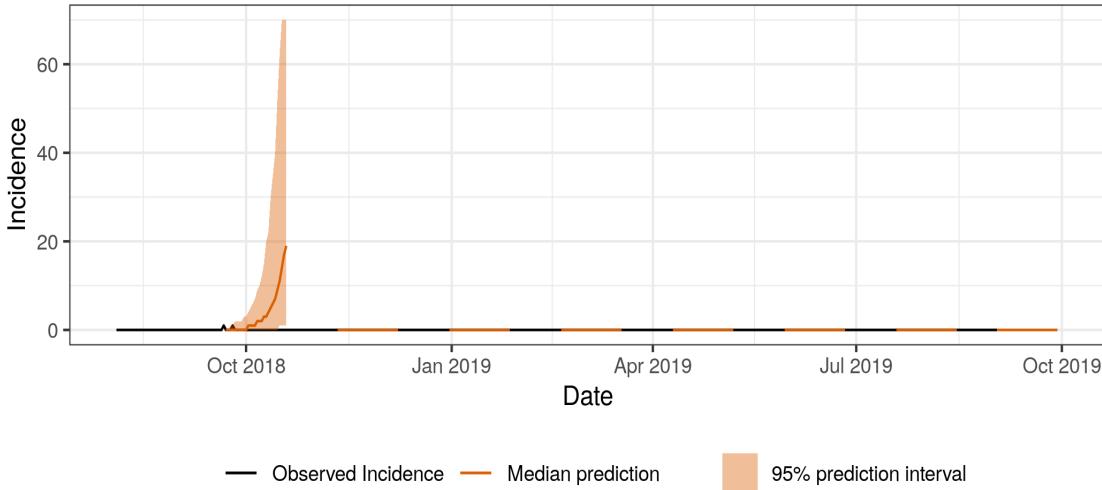
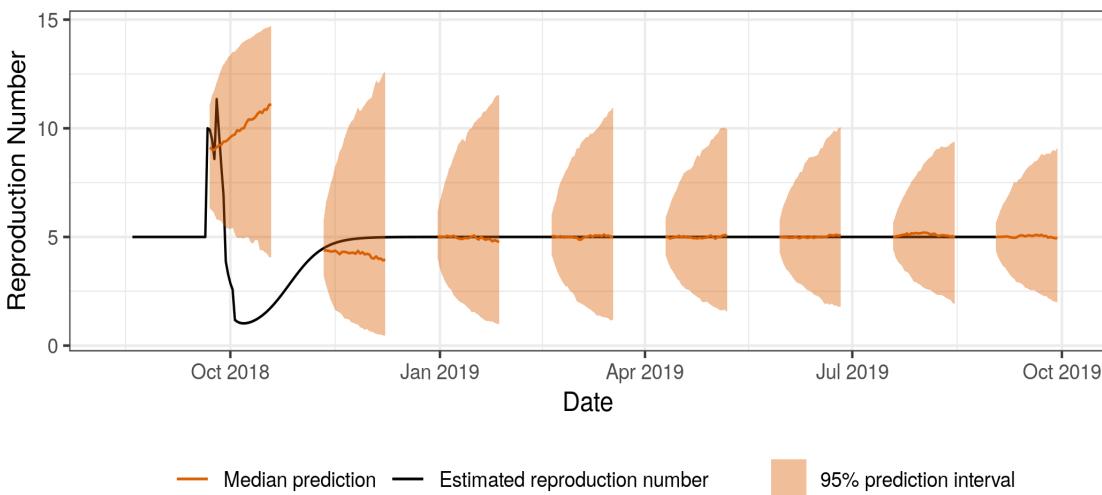


Figure 11: Scores for Mabalako as a function of the forecasting horizon.

## C Tchomia



(a) Forecasted and predicted incidence for the best fitting model



(b) Forecasted and predicted reproduction numbers for the best fitting model

Figure 12: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the best fitting model for Tchomia.

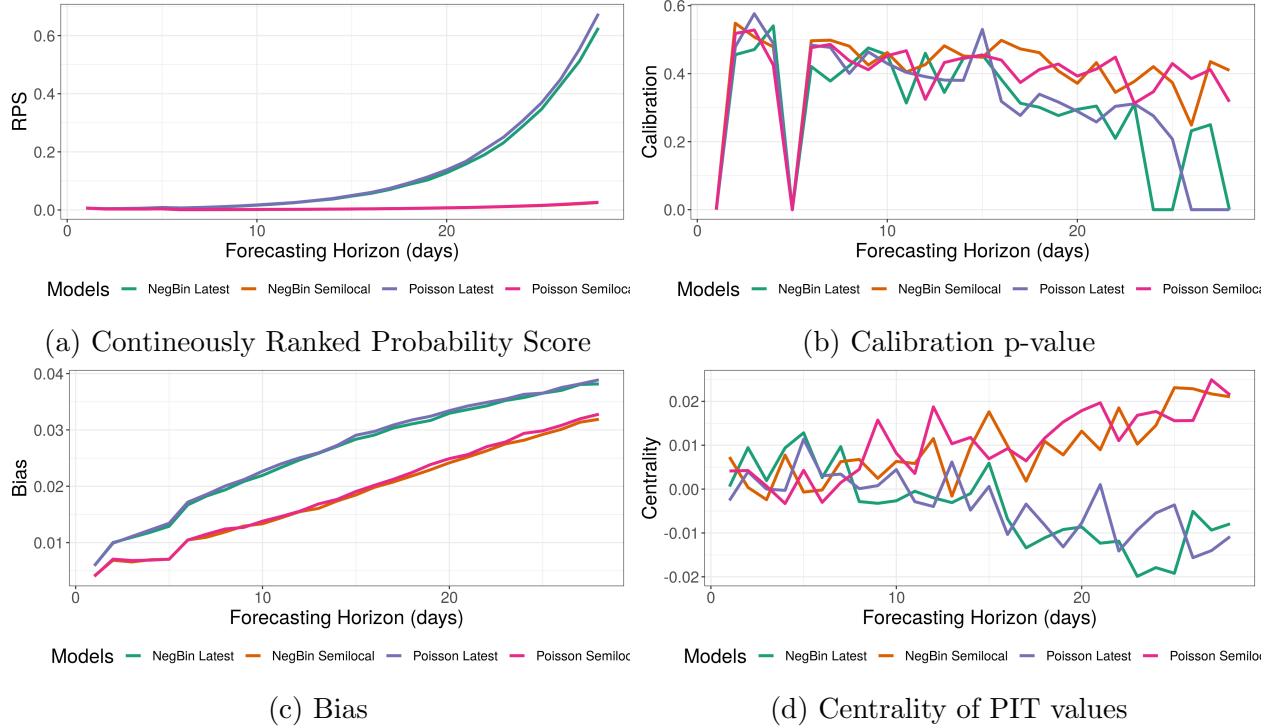


Figure 13: Scores for Tchomia as a function of the forecasting horizon.